THE UNIVERSITY OF CHICAGO

INVESTIGATING THE RELATIONSHIP OF DELUSIONS IN PSYCHOSIS WITH THE FUNCTIONAL INTEGRATION OF PREDICTION ERROR NEURAL NETWORKS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

COMPUTATIONAL NEUROSCIENCE

BY

VICTORIA TEMITOPE OKUNEYE

CHICAGO, ILLINOIS

DECEMBER 2020

This dissertation is dedicated to my parents. Thank you for unconditional love and support. It’s the greatest gift I’ve ever known.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am extremely grateful to my advisor, Dr. Sarah Keedy, for welcoming me into her lab. Thank you for modeling how to be a mental health clinical researcher, for teaching me neuroimaging analysis, for giving me the space to grow as an independent investigator even when I had early doubts, for providing me new ways to think and enabling my many scientific curiosities. I appreciate your infinite patience and your endless support throughout my dissertation. I will always feel blessed to have such a wonderful and kind PhD mentor.

I would also like to thank everyone in the CENSL and CNPRU labs along with the greater BSNIP consortium. This research would not have at all been possible without you support. I am standing on the shoulders of giants. With you all I have grown both as a trainee and as a teacher. Special thanks to - Haley Church, Tasha Rhoads, Emma Herms, Jiaxu Han, Victor Julio, Helen Strnad, Katelyn Kulesa, Jamie Wren-Jarvis, Helen Strnad and Barret Kern. Thank you all for teaching me about team science and how to smile while working towards a common vision.

I would like to thank my committee members, Drs. Harriet de Wit, David Gallo, and Monica Rosenberg, for their insightful comments and contributions to my scientific career. In addition, I thank all my unwitting mentors including those from the UChicago MRIRC, RCC, and numerous computational workshops and scientific conferences around the world. Thanks to the members of the Committee on Computational Neuroscience and Medical Scientist Training Program communities for their camaraderie and assistance.

Thank you to my research participants who remind me why I am here and who I truly do this work for. Thank you to my international dance community for keeping me healthy and sane. Thank you to the black diaspora for serving as my roots and my inspiration.

Finally, I thank Joy Ekuta, Alyson Yee, Daniel Pique, Mariam Alaka, my extended family and most of all my Mom, my Dad for being my support system throughout this entire journey. I am indebted to you all because from the beginning you have believed in me so much that you make it impossible for me to doubt myself. I love you all and I hope I have made you proud.

.

Abstract

Delusions, or false beliefs that are held with high conviction, are signature symptoms of several highly distressing psychotic disorders. Because there are few meaningful biomarkers and limited treatment options for psychotic disorders, understanding this specific symptom may offer a useful transdiagnostic target to move the field forward. As psychosis is commonly understood as a break with reality, it stands that learning more about the neurobiology of the systems that underlie perception of reality, may provide greater insight into the pathology of psychotic delusions. A promising but sparse line of research on delusions is centered on ***prediction error (PE),*** neural signals that register the difference between our expectations and the outcomes that actually occur**.** Problems with this process may account for several cognitive functions implicated in delusion formation and maintenance such as impaired salience detection, increased uncertainty, and reduced precision of belief. While prediction error abnormality is posited as a key mechanism of delusional beliefs, the empirical evidence supporting this is limited. In this project, resting state and task-based fMRI data obtained from patients diagnosed with primary psychotic disorders along with healthy controls were used to investigate the network connectivity of brain regions associated with prediction error and delusions. Though abnormalities in the prediction error activation and functional connectivity were found in psychosis, these changes were not transdiagnostically associated with delusions. However, through use of novel computational modeling approaches, effective connectivity alterations within the intrinsic connectivity of prediction error circuits were shown to be associated with both psychosis and delusions. These results offer new insight into the pathophysiology of delusions and may help to guide more directed investigation of neural circuitry in future biological and clinical studies.

Chapter I: Introduction

Delusions, a Signature Feature of Psychotic Disorders

Delusions are false beliefs, tenaciously maintained even in the face of disconfirming evidence. Clinically, delusions can significantly interrupt adaptive behavior, cause social alienation, and other functional problems. Delusions are observed in several disorders, including schizophrenia, major depression with psychotic features, Parkinson's disorder, dementia and drug induced psychosis. However, within the primary psychotic disorders, i.e. schizophrenia, bipolar and schizoaffective disorder, delusions are considered a prominent and distinguishing feature of theses illnesses.

Psychotic disorders are debilitating mental health illnesses, with lifetime prevalence estimated at between 1-4% within the general population (Rössler et al., 2005; Perälä et al., 2007; Van Os et al. 2009). Psychotic disorders are one of the costliest forms of psychiatric illnesses for societies to treat, with $155 billion in estimated yearly costs to the U.S. (Rössler et al., 2005 Cloutier et al., 2016). The majority of those with psychotic disorders are unable to hold employment (Cook et al., 2006) and one large study found 71% of schizophrenic patients received long-term disability benefits (Rosenheck et al., 2006), suggesting high impairment globally and a significant need for novel approaches to reduce clinical prevalence and impact. No significant treatment development has occurred in decades to address what is often the most distressing and prominent features of psychosis - the positive symptoms, which includes delusions.

Current available treatments, antipsychotic medication and cognitive behavior therapy (CBT), are only moderately effective at best. An estimated 10-40% of patients are resistant to

these treatments (Conley et al., 1997; Schennach et al., 2015) and treatment "success" is rarely defined in clinical studies as total symptom remission. Even given potential gains from treatment, non-adherence is extremely high, due in part to antipsychotic medication's high rate of adverse side effects. Hence, novel treatments with reduced side effect profiles are a high priority. In the future, neurostimulation approaches such as transmagnetic stimulation (TMS), transcranial direct current stimulation (tDCS), or deep-brain stimulation (DBS) may offer such benefits, but among barriers to development of such treatments are our lack of knowledge of the specific neural system dysfunction to target. To this extent, it is necessary to identify and characterize the circuits critical to mediating delusions in order to guide future translational research.

Conceptual Features of Delusions

The emergence and maintenance of delusions in psychotic disorders are poorly understood phenomena. Delusions are a common symptom seen across psychotic disorders. Delusions are classified as a type of positive symptom, a group of psychotic symptoms which are distinguished as abnormal experiences that are "additions" to the range of normal experiences, as opposed to the negative psychosis symptoms that are defined as the absence of some normal behaviors or experiences. Aside from delusions, other positive symptoms are hallucinations, thought disorder, and bizarre behavior. Positive symptoms often co-occur and directly implicate reality monitoring neurocognitive systems in psychosis. While there has been extensive investigation into the clinical and biological features of positive symptoms as a group due to their tendency to covary together more than with other psychosis symptoms (Buchanan and Carpenter, 1994; Serretti and Ogiati, 2004), there has been minimal investigation into how delusions as specific and unique symptoms present across psychotic disorders.

One unique aspect about delusions relative to other positive symptoms is that they are explicitly concerned with firmly held irrational beliefs. False or irrational beliefs are not the singular characteristic of psychotic delusions, as seemingly unfounded beliefs or conspiracy theories are regularly held by individuals and collective groups in our society. Accordingly, to help clarify this, the Diagnostic and Statistical Manual of Mental Disorders defines a delusion as: "A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everybody else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (4th ed., text rev.; DSM–IV–TR; American Psychiatric Association, 2000)."

Several features of delusional beliefs are considered clinically diagnostic of schizophrenia or other psychotic disorders. Additional aspects that distinguish delusions in psychotic disorders are the prevalence, range, the conviction for which they are held, and the distress associated with the false belief (Lees-Grossmann, 2010). Delusions in psychotic disorders are often highly maladaptive, as they can cause both emotional and social distress and ultimately interfere with capacity of patients to engage in normal life activities. The content of delusions can vary; they can reflect both plausible scenarios and impossible scenarios. Several themes are frequently seen in psychosis: paranoid delusions such as "*strangers are plotting against me";* control delusions such as *"my neighbors are using microwave radiation to control my actions"*, thought insertion such as *"the government is implanting thoughts into my mind"*, and delusions of grandeur such as "*I am special advisor to the President".*

Challenges in Investigating Delusional Processes

Biological investigations of delusions have been limited by the inability to create realistic animal models of delusions. In addition, studies attempting to directly capture delusion symptoms in patients are hindered by the poor insight inherent during psychotic states. Furthermore, no task is known to reliably invoke delusions (Arjmand et al., 2020). Challenges in conceptually defining delusions, in addition to determining consistent and objective boundaries for delusional beliefs make their study more indirect. Investigations are made even more complex if taking into account the large variability in delusional experiences. Due to these obstacles delusions have been primarily investigated circuitously, most frequently by standardized patient interviews conducted by trained clinicians. Along such lines, several instruments have been developed to assess delusions using a variety of approaches including phenomenological, dimensional, quantitative and categorical.

Presently, the Positive and Negative Syndrome Scale (PANSS) is the most common approach to investigate psychosis symptoms, and in turn delusion symptoms (Kay et al., 1987). The PANSS is a clinician administered questionnaire that assesses the severity of symptoms in psychosis. Symptoms related to delusions are primarily captured by a single delusion severity score rated 1-7, but to a lesser extent, delusion symptoms may also be incorporated in the grandiosity, persecution, and "unusual thought content" items of the PANSS. All of the PANSS items are grouped into three domains of psychosis symptoms: positive, negative, and general. Psychosis research most often investigates these individual domains or the collective total symptom severity, obtained by summing all PANSS items.

The PANSS is commonly used in research due to its breadth of symptom coverage, and many studies are not aiming for single symptom understanding but rather aim for whole-illness understanding. Indeed, this is what that PANSS was invented for – to support broad illness

severity characterization. However, due to its near ubiquitous use, the PANSS has been used for single symptom investigations including for delusions. The benefit of this is that a common metric is present across many studies. However, there are downsides. One is that the multidimensional nature of delusions is not well captured by the PANSS. A broad range of delusions have been observed clinically and they have been categorized into numerous subtypes, i.e., Persecutory, Grandiose, Control, Reference, Religious, Somatic Delusions, and more. There are also properties of the symptom that may be relevant including level of conviction, level of associated distress, or other such constructs that contribute to the symptom's "severity." An alternative to the PANSS, and perhaps the second most commonly used psychosis symptom severity instrument, is the Scale for the Assessment of Positive Symptoms (and companion Scale for the Assessment of Negative Symptoms), which has separate items to evaluate different types of delusions in addition to the overall delusion severity (Andreasen et al., 1984).

Another limitation of the PANSS, and the SAPS for that matter, is that it only captures a small snapshot of patient symptomology: the symptoms experienced within the last week. However, most psychosis patients during their illness course will have some history of delusions, with varying degrees of severity and duration, a factor often ignored in studies. In answer to this, another instrument exists that captures the history of a several psychosis and mood symptoms, called the Lifetime Dimensions of Psychosis Scale (LDPS; Levinson et al., 2002). Within the clinician-rated LDPS, the severity and duration of delusional symptoms during the patient's lifetime are separately assessed on a scale of 1-4. However, it is rare for this history of symptoms vs. present symptom severity perspective to be acknowledged in studies. Instead, some may simply divide subjects in a binary manner, such as those with a history (at any point) vs. no

lifetime experience with delusions. One drawback to such an approach is that it fails to provide any quantitative estimate of severity.

Another aspect of delusions are their spectrum of severity, from the most impaired and behaviorally-disrupting, to non-clinically-significant/non-impairing subtle odd beliefs. These latter beliefs can be included among "schizotypal traits," observable within the general population. Schizotypy is a theoretical concept which refers to the spectrum of experiences and personality features such as strange perceptions and odd beliefs (Kwapil and Barrantes-Vidal, 2015). In addition, there has been a long-term recognition that psychosis traits often exist at reduced or sub-clinal levels during the earliest or prodromal phases of the disorder (Yung and McGorry, 1996; Mishara, 2010). To this end delusional beliefs have also been reported in the family members of those with psychosis (Schürhoff et al., 2003). Studies endeavoring to address delusions, therefore, may benefit from investigating this full spectrum. One approach to doing so is use of the self-report assessment, the Peters Delusion Inventory (PDI) (Peters et al., 2004). Use of the PDI has shown that delusional ideation is common amongst individuals, e.g., schizotypy, that this subpopulation of non-help-seeking individuals endorse delusional beliefs at the same rate as individuals with psychotic disorders, though of lower severity (Peters et al., 2004). In addition to quantifying a spectrum of delusional belief in healthy subjects and patients, an advantage of the PDI is it more fully assesses the multidimensionality of delusions. It captures delusional subtypes (e.g. paranoid, grandiose, religious, etc.), the total number of delusion-like beliefs endorsed, and within each separate belief, yields a quantified level of distress, preoccupation, and conviction. Whether such properties may have separable etiology or treatment responsiveness may be important to track.

Investigating Delusions with a Transdiagnostic Lens

Much remains unknown regarding the nature of delusions in psychotic disorders, but comparison of subtypes of delusions across disorders has not yielded any clear direction. The full range of delusional themes are experienced across the psychotic disorders (Picardi et al., 2018). Still, there is some evidence that the prevalence of delusional subtypes differs slightly. Grandiose and persecutory delusions are some of the most common subtypes experienced by psychotic patients and have been the most extensively studied. Grandiose delusions are more common during mania, which can occur in bipolar or schizoaffective disorder. Paranoia is the most prominent type among those diagnosed with schizophrenia. It may be that delusional subtypes are supported by distinct mechanisms, or that share common underlying causes, but context, emotion, past traumas or additional patient abnormalities determine their content. One such possibility is that features such distress or conviction may be critical aspects of clinical delusions. In this regard, a shared biology may exist for both paranoid and grandiose delusions, supporting certain core features of the delusional experience such as conviction without relevance to content or form.

It is also of interest how delusions compare among the psychotic disorders and within similar experiences in healthy populations. Overall, phenomenologically the evidence suggests that delusions exist on a continuum that extends from the healthy population to severe psychosis (Freeman, 2006). However, the prevalence and functional impact of delusions are significantly reduced in healthy subjects. It may also be that healthy individuals are more prone to monothematic vs polythematic delusions. It remains unclear if the cognitive and biological features associated with sub-clinical delusions in healthy versus psychotic individuals are the same or also exist on a continuum. Bipolar patients may have delusional profiles more similar to healthy individuals, as they have higher rates of social and clinical functioning compared to

schizophrenic patients, particularly between mania episodes, when they may gain insight into their delusions (Kurtz and Gerraty, 2009; Mann-Wrobel et al., 2011; Nenadic et al., 2015) . It may be that a more similar neurocognitive pathway is shared among healthy/schizotypal individuals and bipolar disorder, than with schizophrenia.

Ultimately, the question that emerges is whether delusion pathology is transdiagnostic or disorder-specific. This question is important to clarify, as it will guide the generalizability of psychological and pharmacological interventions for delusions. Schizophrenia, bipolar and schizoaffective disorders are diagnosed by a range of overlapping subjective criteria that have undetermined biological validity in and of themselves (Dutta et al., 2007; Clementz et al., 2015). Accordingly, similarities have been observed amongst the primary psychotic disorders, both in terms of clinical symptoms beyond delusions, and their biological and genetic etiology (Appelbaum et al., 1999; Badner and Gershon, 2002; Purcell et al., 2009; Keshavan et al., 2011). The shared treatment approaches currently used across psychotic disorders provides further evidence in support of a shared biological mechanism. These observations suggest that a common neural mechanism for delusions may exist within psychotic disorders.

Lessons from the Biology of Psychosis

Several major hypotheses have been put forth around the biological cause of psychotic disorders which may help guide investigation into the neurobiology of delusions. Though still debated the most prominent theory is the dopamine hypothesis of schizophrenia (Van Rossum, 1966). Though originally centered on schizophrenia, the dopamine hypothesis has since been extended to psychosis spectrum disorders due to the overlap of clinical and biological evidence among these disorders. The dopamine hypothesis, very simply, implicates disruptions in dopamine mediated neurotransmission (Howes and Kapur, 2009; Tost et al., 2010; Gründer and

Cumming, 2016). Various points of biological evidence have been provided support for this claim.

The initial evidence for the dopamine hypothesis came from the fortuitous discovery of the first antipsychotic, chlorpromazine. This led to further development of antipsychotic medications which targeted the dopamine system. Molecular imaging research has provided further support of dopamine dysfunction in psychosis, with meta-analyses showing increased pre-synaptic dopamine in the subcortical basal ganglia. Ultimately, the greatest evidence for the dopamine hypothesis come from the discovery that the efficacy of antipsychotics correlated with dopamine D2 receptor occupancy (Farde et al, 1992; Kapur et al., 1999).

Additional evidence exists supporting a role for dopamine in delusions. Dopamine mediated hyperactivity of striatum has also been found to correlate with positive symptoms (Howes et al., 2012). Delusions have been induced in healthy individuals using 1) amphetamines which increase intracellular dopamine concentrations and 2) through dopamine agonists such as L-Dopa which is well known to cause psychosis symptoms in patient's being treated for Parkinson's disorder (Friedman and Sienkiewicz, 1991, Bramness et al., 2012). Furthermore, these drugs have also been shown to worsen symptoms when given to schizophrenic patients (Snyder et al., 1974). In addition, anti-psychotics are commonly used to treat delusions, not only within the psychotic disorders but also withing other illnesses suffering delusional experiences.

An important caveat is that although blockade of striatal dopamine helps reduce positive symptoms in some patients, it does not necessitate that delusions are caused by dopamine. Reduced salience or alterations in other dopamine mediated processes in brain may be a ubiquitous mechanism that helps address reality monitoring deficits caused by other primary

abnormalities in psychosis. Thus, although the dopamine hypothesis remains the most prominent theory of psychosis, it is still debated as there is not yet conclusive evidence that dopamine pathophysiology is the cause of psychotic symptoms. In this regard, several other popular theories have been put forward to explain psychosis such as the disconnection hypothesis (Friston and Frith, 1995) and glutamate hypothesis (Coyle, 1995). However, it has also been suggested that the primary cause of psychosis may likely be heterogenous with several potential impairments (e.g. genetic, neurodevelopmental, environmental, etc.), with dopamine dysfunction a possible final common pathway (Howes and Kapur, 2009).

Cognitive Theories of Delusions

Numerous cognitive models have been proposed to explain delusions. Several of the more commonly studied models are briefly reviewed below (Blackwood et al., 2001; Bell et al., 2006; Garety and Freeman, 2013; Poletti and Smbataro, 2013). First amongst the cognitive explanations is the aberrant salience theory, which suggests that delusions arise due to misdirected attention given to stimuli that should not normally engage salience detection systems (Kapur et al., 2003). Multiple variations of this model exist to explain how delusional beliefs are generated. One perspective of this account is that delusions may form as a top down explanation of anomalous beliefs or perceptions experienced during psychosis. Another perspective is based on irregular salience tagging, which leads to irrelevant stimuli being coded as important or informative.

Another cognitive theory is disrupted attribution style, which proposes that individuals with delusions have difficulty assigning proper self vs. other attributions to beliefs. This can manifest explicitly as deficits during theory of mind tasks (Frith, 1994; Ventura et al., 2011). The attribution theory can also extend to poor meta-cognition or insight into self-generated stimuli.

This has been hypothesized to be due to increased self-confirmation bias. This may be especially relevant for delusions of reference and control. A related cognitive account for delusions centers on probabilistic reasoning. Research suggests that false beliefs arise due to the propensity of delusional patients to jump to conclusions and establish firm beliefs earlier based on less evidence (Fine et al., 2007, So et al., 2012).

A separate framework for delusions attempts to address the emotional aspect common to delusions such as those that occur in paranoia or mania. This theory of delusions suggests that certain false beliefs are due to abnormal emotional regulation and in some cases may even be a result of defense mechanisms (Bentall et al., 1994). Various behavioral studies have investigating these cognitive models, which have provided some evidence to support that these cognitive processes are abnormal in psychosis. However, the biological evidence linking most of these models to delusions, such as through functional imaging capturing both the cognitive operation and brain function, remains sparse or inconsistent.

Neurocognitive Models of Delusions

A variety of neuroimaging techniques have been used to investigate if abnormalities are present in the neural systems subserving the cognitive domains hypothesized as disrupted in delusions. These methods include positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG). These tools have been used to investigate the neurobiology of the cognitive domains thought to underlie delusions and also directly identify changes in the neural circuitry of delusional subjects (Knobel et al., 2008; Broyd det al., 2017; Arjmand et al., 2020). Among the most common biological changes explored are alterations in dopaminergic neurotransmission; in the brain networks known to support self-reflection and theory of mind; in the neural systems supporting belief evaluation; episodic

memory; emotional processing; and in the neural circuits supporting reward and salience detection (e.g. dysregulation of the basal ganglia through either increased bottom-up processing or disruptions in top-down regulatory control, along with other possible dysconnectivity within the salience related brain networks). As a result of the accumulated findings, multiple neurocognitive models of delusions have been advanced, aiming to integrate both cognitive theories and biological evidence to explain the formation and maintenance of delusions (Broyd et al., 2017). Currently, the aberrant salience, two-factor, and predictive coding models are among the most discussed neurocognitive theories.

Aberrant Salience

First proposed by Shitij Kapur (2003), the aberrant salience framework explains the development of delusions by combining the evidence from the pharmacology of anti-psychotics, neurobiological evidence of dopaminergic dysfunction in psychosis, and growing insight into the role of dopamine in the brain's normal reward processing. This theory was originally developed from the incentive salience hypothesis as applied to research in addiction and reinforcement learning. Within normal cognition, dopamine mediates the attribution of salience to conditions that predict reward. This salience then helps motivate decisions and guide behaviors. Specifically, the aberrant salience hypothesis proposes dysfunctional transmission of dopamine in psychosis, resulting in stimulus-independent release of dopamine. This undermines the normal role of dopamine to mediate contextually relevant salience. Based on the empirical evidence of increased dopamine within the midbrain and striatum for psychosis patients, it was argued that psychosis is a result of hyperdopaminergic activity within the basal ganglia that lead patients to assign salience to irrelevant cues. Similarly, it was proposed that inappropriate salience may give heightened awareness to internal stimuli which are normally suppressed. Under the original

framework, delusions are conceptualized as a result of top-down explanation for abnormally salient experiences. This theory provides a straightforward interpretation of D2-receptor antagonism by anti-psychotics. These medications help reduce but not necessarily eliminate delusional beliefs by dampening the salience provided to the beliefs and reducing the potential for aberrant salience to trigger new delusions.

Two-Factor Model:

The two-factor model of delusions was initially proposed by Max Coltheart based on his investigation of monothematic delusions that occur following stroke and other neurological disorders, but it has also been extended to cover polythematic delusions experienced within psychotic disorders (Coltheart et al., 2007, Coltheart et al., 2011). According to the two-factor theory, two abnormalities must be present for delusions to occur. First, a perceptual or inferential anomaly must exist that initially prompts and determines the content of the belief. Second, a more generalized anomaly must be present which impedes rejection of the belief. It has been hypothesized that the first factor is due to underlying but non-specific damage to various perceptual systems. The second factor is proposed to be due to damage more consistently occurring in the right frontal lobe, an area argued as critical for belief evaluation. Lesion studies of patients with delusions of misidentification such as Capgras delusions (the belief others have been replaced by identical-appearing imposters) offer some evidence supporting this, at least for monothematic delusions. However strong conclusions are limited as the incidence of these clinical cases is exceeding rare (Darby and Prasad, 2016). A recent network mapping of the lesions from delusion misidentification cases provides evidence of how changes in functional connectivity among them may impact networks associated with the two factors, going beyond the original source lesion explanation (Darby et al., 2016). They found significant overlap in the

connectivity of the lesions with the left retrosplenial cortex, a region associated with familiarity in meta-analysis and implicated as part of factor one. They also noted connectivity of the lesions to the right ventral frontal cortex, a region associated with expectation violation in meta-analysis and implicated for factor two. These results suggest that analysis of the network disruptions, rather than solely within brain region disruptions, may be an important aspect in understanding the emergence of delusions.

Within psychosis it has been proposed that this same two-factor model may exist, but manifests differently to cause polythematic delusions. Observations of right prefrontal cortex dysfunction in schizophrenia has been argued to support evidence of Coltheart's model, or at the least suggest a shared pathway for disrupted belief evaluation in delusions (Coltheart et at., 2007). Poletti and Sambataro (2013) proposed a modified two-factor model that transdiagnostically explains the development of delusions. The first factor is aberrant affect generation regarding emotions and rewards. The second factor is predicated on impaired construction of internal models and theory of mind in patients. However, it has been argued by others that a two-factor account of psychosis is unnecessary as the perception and belief are served by the same underlying neural substrate (Fletcher and Firth, 2009; Krishnan et al., 2011).

Predictive Coding

According to the predictive coding or prediction error account of psychosis, positive symptoms such as delusions arise due to a core neural deficit in inferential processing (Fletcher and Frith, 2009).Under the prediction coding framework, the brain operates in a Bayesian fashion (Friston et al., 2006; Friston, 2010). The Bayesian brain model argues that our perceptions and beliefs about the world are determined by matching our cognitive *priors* with sensory *posteriors.* Thus, the brain works to maximize efficiency and minimize surprise by

integrating prior beliefs or expectations with the new sensory information it receives. It is proposed that minimization of prediction error is the central mechanism by which the brain achieves this process.

Prediction error is a psychological process defined as the discrepancy between what an individual expects to experience and what is actually experienced. Prediction errors are argued to be crucial for belief updating, by providing the drive or awareness to update prior predictions (Fig 1-1). In this regard, by providing feedback that guides the deduction of reward relationships or causal associations, prediction error operates as important signal for both learning and reality monitoring (Schultz and Dickenson, 2000).

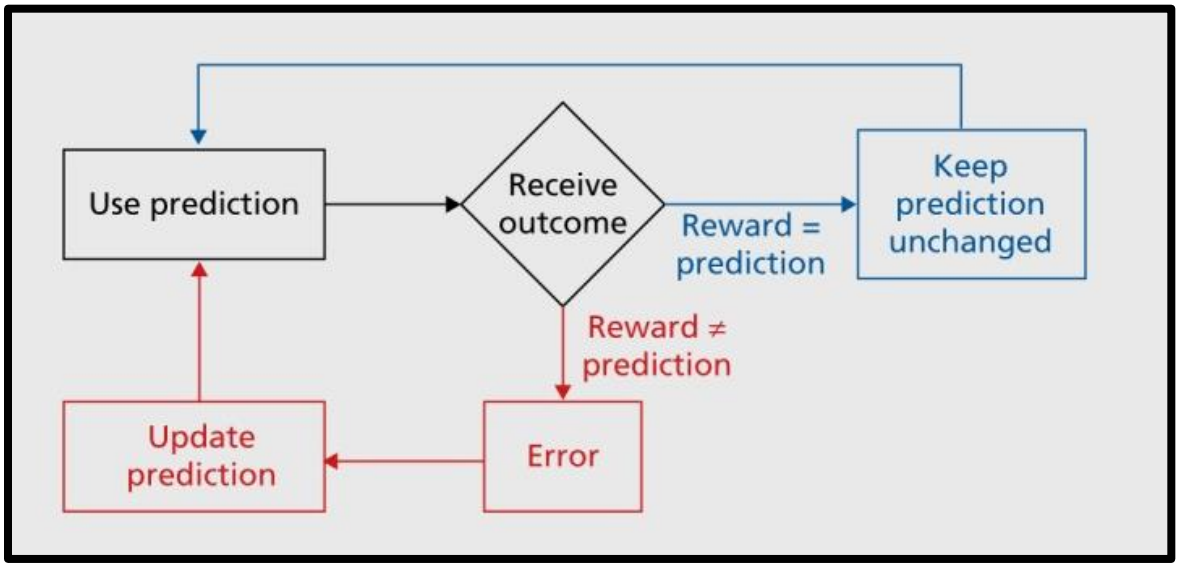

Figure 1-1. A Mechanism for Belief Updating Though Prediction Error

Demonstrated in the context of reward learning, a prediction error occurs when a prior expectation for a reward does not match the received award. The perceived prediction error (red) is then used to update prediction or belief, which can then be used to guide behavior. Alternatively, no prediction error is generated if the prior expectation matches the outcome (blue). This results in a prediction that remains unchanged.
Reprinted from 'Dopamine reward prediction error coding' by W. Schultz W, 2016, Dialogues in Clinical Neuroscience)

The impairments in belief updating due to prediction error deficits are hypothesized to lead to the development and maintenance of delusions. Abnormal prediction error is proposed to disrupt many of the cognitive process hypothesized as important for delusions: altered salience detection, increased uncertainty, and reduced precision of belief (Corlett et al., 2010; Bortolotti and Miyazano, 2015). Thus, the prediction error hypothesis is argued to reduce the two-factor account to a single mechanism which can account for the aberrant salience associated with delusions in psychosis.

Both the aberrant salience and predictive coding frameworks are conceptualized to explain more than delusional symptoms. They may more broadly explain the disruptions in reality monitoring that can also predispose individuals with psychotic disorders to other positive symptoms such as hallucinations and disordered thoughts (Kapur et al. 2003; Rossier et al., 2009; Adams et al., 2014). Given the heterogeneity of psychotic illness and delusion experiences, it may be that multiple cognitive mechanisms among these are applicable to subsets of individuals, with the commonality being disrupted prediction error signaling. Conclusive evidence supporting one account over another is still limited. However, support for the prediction error hypothesis has been greatly enhanced in recent years due to cellular evidence of a direct association of prediction error with dopamine signaling and neuroimaging evidence showing disrupted prediction error activity is associated with delusions.

Prediction Error

In the reinforcement learning literature prediction error has been assessed in multiple ways (O'Doherty et al., 2003; D'Astolfo and Rief, 2017). This includes correlating neural activity with surprise and expectation violation, which are often considered cognitive proxies of prediction error, and also correlating neural activity with formal computational models of

prediction errors such as those described by temporal difference and Rescola Wagner learning models. In addition, two forms of prediction error have been investigated: signed prediction errors (based on the valence of reward or aversive outcomes) and unsigned prediction errors (based on absolute expectation violation).

Prediction error in many of the early animal studies was identified as conditions of expectation violation to reward expectancies during reinforcement learning paradigms. Electrophysiological studies of laboratory animals (both within rodents and non-human primates) have identified that phasic firing of the dopaminergic neurons in the midbrain and striatum corresponds to prediction error. (Schultz, 2000, Waelti et al., 2001, Bayer and Glimcher, 2005, Schultz, 2006). These studies showed that at a neuronal level, prediction error is linked to dopamine transmission.

Similar findings have been shown when neuroimaging is used to identify the neurocorrelates of prediction error within humans. During reinforcement learning,prediction error activation is found most repeatably in dopamine-rich regions such as the midbrain, striatum (including the caudate, putamen, nucleus accumbens) and within the lateral prefrontal cortex. However, meta-analyses have shown a much larger common network of brain regions to be associated with prediction error (Garrison et al., 2013; Chase et al. 2015; Astolfo and Rief, 2017). Extended prediction error activation is seen in the insula, thalamus, amygdala-hippocampal complex, cingulate cortex, and medial prefrontal cortex.

This pattern of activations from the reinforcement learning paradigms suggest that regulation of prediction error may depend on basal ganglia – corticothalamic loops (sometimes also referred to as cortico-striatal-thalmo-cortical loops). The basal ganglia are composed of

several subcortical structures which include the striatum, limbic, and midbrain regions. The basal ganglia are connected to the thalamus and to cortex though several parallel circuits (Alexander et al., 1986; Haber and Knutson, 2010, den Ouden et al., 2012). Animal models have verified that neurotransmission through these loops is important for mediating reinforcement learning. The culminative evidence suggests that dopaminergic neurons in the midbrain and striatum are critical for mediating salience while the prefrontal cortex activity is likely to mediate holding beliefs online, switching between beliefs, and inhibiting irrelevant information (Bromberg-Martin et al., 2010; Hampshire et al., 2010; Niendam et al., 2012).

Prediction Error and Delusions

Several neuroimaging studies in psychosis have examined whether neural alterations exist within the prediction error system. As expected from prior neuropharmacological imaging of dopamine systems in psychosis, many studies report abnormal prediction error activation in psychosis subjects relative to healthy controls (Kerns et al., 2005; Murray et al., 2008; Gradin et al., 2011; Krawitz et al., 2011; Schlagenhauf et al., 2013; Ermakova et al., 2018).

The striatum, midbrain, and lateral prefrontal cortex have been most strongly implicated within the limited neuroimaging studies that specifically investigate delusions. Corlett et al. (2007) reported that first episode psychosis patients with delusions exhibited abnormal neural signaling of prediction error within the dorsolateral prefrontal cortex and striatum during feedback of unexpected events (i.e., expectation violations), and decreased activity of the lateral prefrontal cortex was associated with increased delusion severity. Another study of patients with schizophrenia and schizoaffective disorder used temporal difference modeling to measure prediction error during a Pavlovian conditioning paradigm (Romaniuk et al., 2010). They reported evidence of altered prediction error in the midbrain and striatum but did not find a direct

association of prediction error with delusion severity. However, they did report that increased midbrain activity to neutral cues compared to aversive cues was associated with delusions in schizophrenia. This finding suggests prediction error abnormalities do exist in psychosis, but aberrant salience may be more relevant for delusions. However, in another study investigating delusion spectrum beliefs within healthy subjects by Corlett and Fletcher (2012), bilateral caudate activity was found to be negatively associated with odd beliefs, while decreased prediction error activation in the prefrontal cortex, striatum, and midbrain were all associated with distress related with odd beliefs.

Coupled with the cellular evidence on the mechanism of prediction error in these regions, these findings provide some neural evidence supporting the predictive coding account of delusions. However, some recent studies in these regions have failed to find a similar association of prediction error with positive symptom (Ermakova et al., 2018; Katthegan et al., 2018). One issue may be that the relationship with delusions is obscured amongst the other positive symptoms. While many prior studies have considered positive symptoms as one factor – lumping hallucinations, delusions, thought disorder and affective symptoms all together - evidence suggests the positive symptoms are multidimensional and may hold divergent neurobiological pathophysiologies (Steel et al., 2007). Nevertheless, another important limitation is the sample size of the earlier studies. With patient cohorts of fewer than 25 subjects, the power of the studies to confidently identify true effects is reduced (Turner et al., 2018). While prior research offers some promising empirical evidence in support of the predictive coding account of delusions, it is clear the findings require replication in larger studies.

In order to clarify the specific etiology of psychotic delusions, future work must also expand beyond neural activation to study the dynamics (e.g. interactions) of the prediction error

network. This would involve determining the normal brain dynamics of prediction error during tasks and at rest. Such investigations would reveal how the prediction error circuit functions internally and how it is regulated indirectly by the influence of other brain networks, such as those involved in attentional, self-referential and emotional processing. Dysregulation of prediction signaling may be due solely to context specific interactions among the key brain regions thought to mediate prediction error, or alternatively, dysregulation could be a consequence of already-disrupted intrinsic communication more ubiquitously in the brain, affecting prediction error brain regions as well as others. Understanding the neural circuits involved in prediction error dysregulation may enable greater understanding of the mechanisms that predispose patients to delusions, and ultimately more precise therapeutic targets for treating delusions. Furthermore, in recent decades the use of computational models to investigate normal and abnormal neurocognition has been significantly advanced. Examining the brain through a computational framework can provide increased insight into the neural abnormalities which may underlay the relationship of delusions and prediction error. Further, such work can lead to the development of explicit models that can be tested and validated using empirical data.

With this motivation, the set of studies within this project assessed the association of prediction error response across the psychosis spectrum using neuroimaging data from both patients and healthy subjects. By identifying which specific neural correlates are disrupted in psychotic prediction error response, I subsequently interrogated previously unexamined functional and effective connectivity of prediction error regions within these groups. Further examination of subjects' resting state brain activity can clarify the pervasiveness of functional network changes, beyond just that observed during prediction error tasks. The project's integration of neuroimaging methods provides an innovative opportunity to understand how

dysregulation of the cognitive systems subserving prediction error may contribute to specific psychosis symptomology and can illuminate the therapeutic path for tractable biological targets.

Central Questions for Dissertation

Within this project I investigate the following central questions. (1) Is the prediction error biomarker a transdiagnostic predictor of delusions? In chapter 2 I address this question by recruiting and administering a prediction error task within a new psychosis cohort. Through this study I aim to investigate if the association of delusions with prediction error activation is replicated within a larger transdiagnostic sample. (2) Within the prediction error processes, is task-dependent connectivity in the delusion associated prediction error (D-PE) circuit altered in psychosis and/or in association with delusions? This latter question is investigated is chapter 2 with generalized psychophysiological interaction (gPPI), a task basked connectivity analysis method that examines how connectivity in the brain changes under different contexts. Changes in brain connectivity during periods of expectation violation for the prediction error task are investigated within the sample. (3) Outside the context of the prediction error task, is intrinsic connectivity of the prediction error network altered in psychosis generally and more specifically with delusions? In chapter 3, the whole brain functional connectivity of D-PE circuit is evaluated in a large transdiagnostic resting state dataset. (4) Does empirical evidence support a specific network model of prediction error dysfunction in delusions? The effective connectivity of the D-PE circuit is tested for an association with psychosis diagnosis and delusion severity using the same transdiagnostic resting state dataset. Bayesian evidence is used to determine the optimal neural circuit model that explains intrinsic activity with the D-PE circuit.

Chapter II: Prediction Error Task Activation and Functional Connectivity in Psychosis and its Association with Delusion Symptoms

Introduction

Although numerous genetic and pathophysiological changes have been reported in psychotic disorders in recent years, the specific link between these changes and patients' experience of psychotic symptoms is poorly understood. There remains little insight into the biological mechanisms behind the emergence of positive psychotic symptoms such as delusions, hallucinations, and thought disorder, nor how antipsychotic medication and psychotherapies may help resolve them for certain patients. Recent neuroimaging studies have reported an association of the neural substrates of prediction error – brain signals that encode expectation violations - with not only pathological changes in psychosis but also delusion symptoms. These results suggest a biological pathway may exist for psychotic delusions, linking the symptom to disruptions in association learning and reality monitoring (Corlett and Fletcher, 2015) that in turn may emerge as a result of abnormal predictive coding caused by dysregulated dopaminergic signaling (Fletcher and Frith, 2009). However, the reproducibility of these findings must be verified, and numerous questions remain such as how the normal prediction error response is regulated, how it becomes disrupted in psychosis, and why this may be relevant for delusions.

Human neuroimaging studies of reinforcement learning have identified evidence of brain regions that encode prediction error – regions that activate in response to expectation violations during learning. This prediction error neural signature has been observed in several regions of

the brain, with results found most repeatably in dopamine-rich regions such as the midbrain, striatum (including caudate) and prefrontal cortex (Garrison et al., 2013; Chase et al. 2015; Astolfo and Rief, 2017). This is consistent with electrophysiological studies in rodents and non-human primates that associate phasic dopamine release to prediction error (Schultz et al., 1997, Schultz, 2000). Within these same circuits, and connected regions, abnormal prediction error signaling has been observed in psychosis during reinforcement learning and related tasks. Such studies report reduced prediction error-related activation in the striatum (which includes the caudate), midbrain, insula, amygdala-hippocampal complex, cingulate cortex, medial prefrontal cortex and lateral prefrontal cortex (Kerns et al., 2005; Murray et al., 2008; Gradin et al., 2011; Krawitz et al., 2011; Schlagenhauf et al., 2013; Ermakova et al., 2018).

While many of the abnormal prediction error response studies in individuals with psychosis address the problem at a general illness level, showing how there is general difference from health with the prediction error process, a more specific relationship to delusions has become of interest given the ecological validity of the connection. It may be that holding on to a belief tenaciously despite contrary evidence, the definition of a delusion, is directly related to problems in error detection and updating knowledge. Promisingly, a few small studies found evidence that suggested abnormal prediction error activation is specifically associated with delusional beliefs (Corlett et al., 2007, Romaniuk et al., 2010; Corlett and Fletcher, 2012). In an analysis of 12 patients with first episode psychosis and 12 matched healthy subjects who were scanned while performing a causal learning paradigm, Corlett et al. (2007) reported that patients exhibited decreased activation within the right ventrolateral prefrontal cortex and caudate (dorsal striatum) during feedback of unexpected events (i.e., expectation violations). Further, this activity in patients negatively correlated with severity of an approximation of delusion symptom

severity, as measured by the "unusual thought content" item from the Brief Psychiatric Rating Scale (no more direct delusion assessment was reported).

Using a similar causal learning fMRI design, Corlett and colleagues conducted a subsequent study evaluating prediction error in relation to schizotypal traits in 18 healthy participants. They used the Peters Delusion Inventory (PDI) and Chapman scale, both instruments that can measure schizotypal traits. These traits include the spectrum of strange perceptual experiences and odd beliefs or that exist in the general healthy population and appear to be a mild end of a continuum with psychosis symptoms (Kwapil and Barrantes-Vidal, 2015). The Chapman scale was used to measure "magical ideation," a schizotypal trait similar to delusions. The PDI was used to assess the range of odd beliefs endorsed by subjects in addition to subjective dimensions associated with the beliefs: distress, conviction, and preoccupation. The study found that magical ideation was negatively correlated with the magnitude of prediction error responses in the caudate, while distress related to odd beliefs measured by the PDI was negatively associated with prediction error response in the caudate, midbrain, and right dorsolateral prefrontal cortex (Corlett and Fletcher, 2012).

Another research group (Romaniuk et al. 2010) evaluated 20 patients diagnosed with schizophrenia or schizoaffective disorder and 20 healthy participants using a Pavlovian conditioning (learning) paradigm. They found patients had heightened midbrain activation to neutral cues and that this was associated with their current delusion severity measured by the Positive and Negative Syndrome Scale (PANSS). They also found abnormal prediction error response in the midbrain and ventral striatum, but that this was not linearly associated with delusion severity.

Taken together, these neuroimaging studies provide preliminary evidence supporting abnormal prediction error signaling in psychosis relating to delusions. However, the small number of studies, inconsistencies in methods and results, and small sample sizes suggests further work is needed. In addition to studying larger samples that have psychotic delusions, attention to additional neural system features beyond strength of evoked activation may be important. Such features include connectivity, synchrony, and directionality among brain regions that mediate prediction error. These characteristics are referred to as neural dynamics. Studies that can assess these additional dimensions may offer richer information to further our understanding of prediction error in psychosis and its relation to delusions. Among the issues, there is also a particular concern regarding the assessment of delusion severity, which was inconsistent and indeed very broadly speaking, does not received scrutiny in the literature. Whether the assessments used are reasonable representations of "delusion severity" in a manner that may meaningfully reflect gradations of illness severity/more abnormal biology is unknown. Furthermore, it is unclear if prediction error circuity is associated with delusions generally or is specific to schizophrenia in which it has been more extensively studied.

Study Aims

The current study aims to investigate prediction error neural activity abnormality in psychotic patients and ascertain the relation to delusion symptoms. The first objective is to test if neural activation during prediction error activity is altered in a transdiagnostic psychosis sample, and to determine whether that alteration is associated with delusion symptoms. Through this process the replicability of the previous studies will also be tested. Then, looking beyond neural activation alone, the study aims to compare groups on the functional connectivity among key brain regions mediating prediction error, as well as assess any relationship of connectivity to

delusion severity. Delusion severity will be characterized in different ways and each analyzed separately in relation to neural activation. If differential findings emerge among the severity instruments, strength of association will be compared between them to determine which delusion severity instrument may be optimal.

To accomplish these aims, the following hypotheses will be tested: 1) Individuals with psychosis will have lower activation than healthy controls during periods of expectation violation (i.e. prediction error) in the midbrain, prefrontal cortex and striatum measured during an associative learning task. 2) In the patients, prediction error activation in the striatum and prefrontal cortex will be associated with clinician assessed and self-reported current delusion severity and with lifetime delusion severity; 3) Functional connectivity of prediction error regions found to be associated delusions will be altered in patients relative to controls during expectation violation. These hypotheses are replications and extensions of the studies showing delusion associated prediction error neurocorrelates. The extensions include use of not only current delusion severity, as all prior work uses, but also includes assessment of whether lifetime delusion susceptibility, or self-reported delusion severity, associate with altered prediction error neural activity. Another extension is the investigation occurring in a robustly sized transdiagnostic sample, to capture a breadth of major psychotic disorder cases with delusional symptom histories. Lastly, the plan to analyze prediction error related connectivity expands beyond task-evoked neural activation, potentially offering a more complete picture of relevant neural system alterations in psychosis patients and delusion symptom severity.

Methods

Participants

54 psychotic patients and 20 healthy controls were enrolled in the prediction error study. Prediction error task participants were recruited from among subjects consented to the ongoing Bipolar & Schizophrenia Network for Intermediate Phenotypes 2 (B-SNIP2) study at the University of Chicago. The study was approved by University of Chicago Institutional Review Board. Psychotic subjects were diagnosed with bipolar disorder with psychosis (n=16), schizoaffective disorder (n=20), or schizophrenia (n=13). Recruited healthy subjects were matched to the patients on age, sex, race and education. Subjects met standard B-SNIP2 inclusion and exclusion criteria: able to provide written informed consent, age 18-60, IQ > 60, no current substance abuse disorders or major neurological/cognitive/cerebrovascular-affecting disorders, no head trauma history. Healthy controls had additional criteria of no personal history of any psychiatric disorder or history of schizophrenia, schizoaffective disorder, or bipolar disorder in first-degree relatives.

Clinical Assessments

Subjects were given a SCID-I/P (Tamminga et al., 2014; First et al., 2002a; First et al., 2002b) by a trained clinical rater to confirm diagnosis (or lack of one in healthy subjects). For patients, a rater also then assessed the severity of a range of psychotic symptoms with the Positive and Negative Syndrome Scale (PANSS), capturing positive, negative, and general psychopathology symptoms (Kay et al., 1987). Delusional severity measures include the "Delusions" item of the PANSS (P1), a single item with a score range 1-7 and is the primary Delusion severity assessment of PANSS. Additional PANSS items are also relevant for exploratory analyses, including "Unusual Thought Content," (as used in Corlett et al., 2007), "Grandiosity," and "Persecution," all items which encompass distinct types of delusional beliefs and can overlap with the "Delusions" item. Self-assessment of delusion symptoms was also

captured via responses subjects gave on the Peters Delusion Inventory (PDI). This is a 21-item self-report measure that captures delusional ideation in both normal and patient populations (Peters et al., 2004). In addition to obtaining a spectrum of delusional belief in healthy subjects, the PDI assesses the multidimensionality of delusions by capturing subtypes (paranoid, grandiose, religious, etc.) and yielding separate measures of number of delusion types endorsed, level of distress, level of preoccupation, and level of conviction; when distress, preoccupation, and conviction subscales are summed, this is the PDI total severity score. A third instrument, the Lifetime Dimensions of Psychosis Scale (LDPS) (Levinson et al., 2002) provided clinician-rated scores of a variety of psychosis and mood symptoms over patients' lifetimes, from which the delusional severity item was selected (score range 1-4), supplementing the PANSS which reflects present delusional severity. In sum, PANSS is a clinician-rated current severity instrument, PDI is a self-rated lifetime severity instrument, and LDPS is clinician-rated lifetime history of severity.

Neuroimaging Acquisition Parameters

All subjects underwent a one-hour MR scanning session with the Phillips Achieva Quasar Dual 16 channel, 3 Tesla MRI scanner at the Magnetic Resonance Imaging Research Center at the University of Chicago. *Prediction Error Task (fMRI) acquisition parameters:* T2* weighted echo planar imaging with a gradient-echo pulse sequence was run with TR=1500ms. To allow full brain coverage, FOV= 200 mm^2, Matrix=64x64. Slices were 27 axial slices, 4mm thick with 1mm gap, orthogonal. *Structural MRI (MPRAGE):* 3D high resolution isotropic T1-weighted volume acquired per ADNI (Alzheimer's Disease Neuroimaging Initiative) protocol (same as for B-SNIP study).

Prediction Error fMRI Task

The task is a modified version of the associative learning task created by Corlett and Fletcher (2012). Subjects are asked to act as an allergist for a hypothetical patient, taking on the difficult challenge of narrowing down the patient's specific food allergies by making predictions on the potential allergic reaction to the patient's meals. Across stages, each trial consists first of a 3 second presentation of a meal stimulus comprised of one food item (Stage 1 and 3) or a pair of food items (Stage 2). Subjects are asked to respond yes or no via button press if they believe the meal presented will cause an allergic reaction. Then, there is a 2 second outcome presentation indicating whether an allergic reaction did or did not occur (Fig 2-1). Trials are run consecutively with a 3-5 second jittered inter-trial interval with a fixation cross. Subjects complete a total of 252 trials and the total task length is ~40 min.

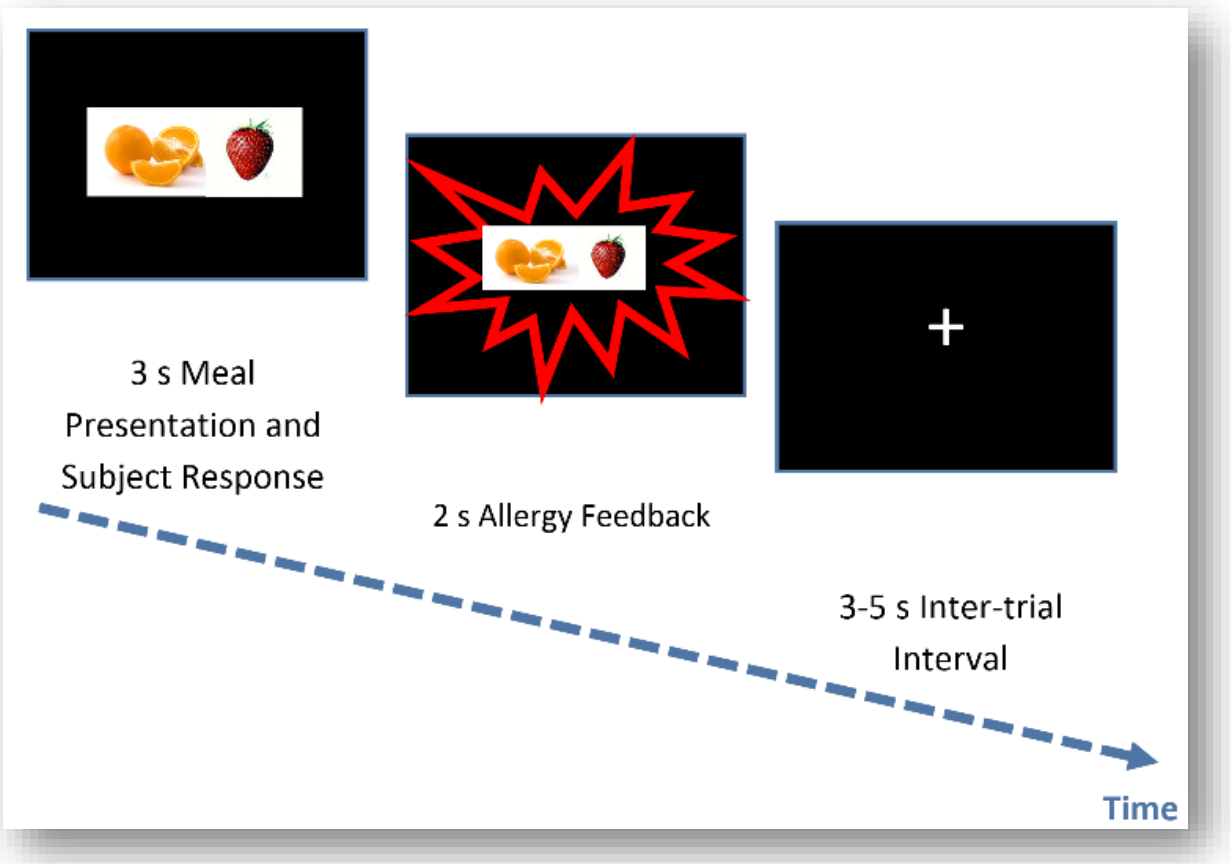

Figure 2-1. Trial Design

Participants are shown meal stimuli and respond Yes or No to predicted allergy outcome within 3 sec. Allergy outcome is then shown for 2 sec (red explosion = allergic reaction; green rectangle = no reaction), and then fixation. A Stage 2 trial is depicted. Stage 1 or 3 show a single food item.

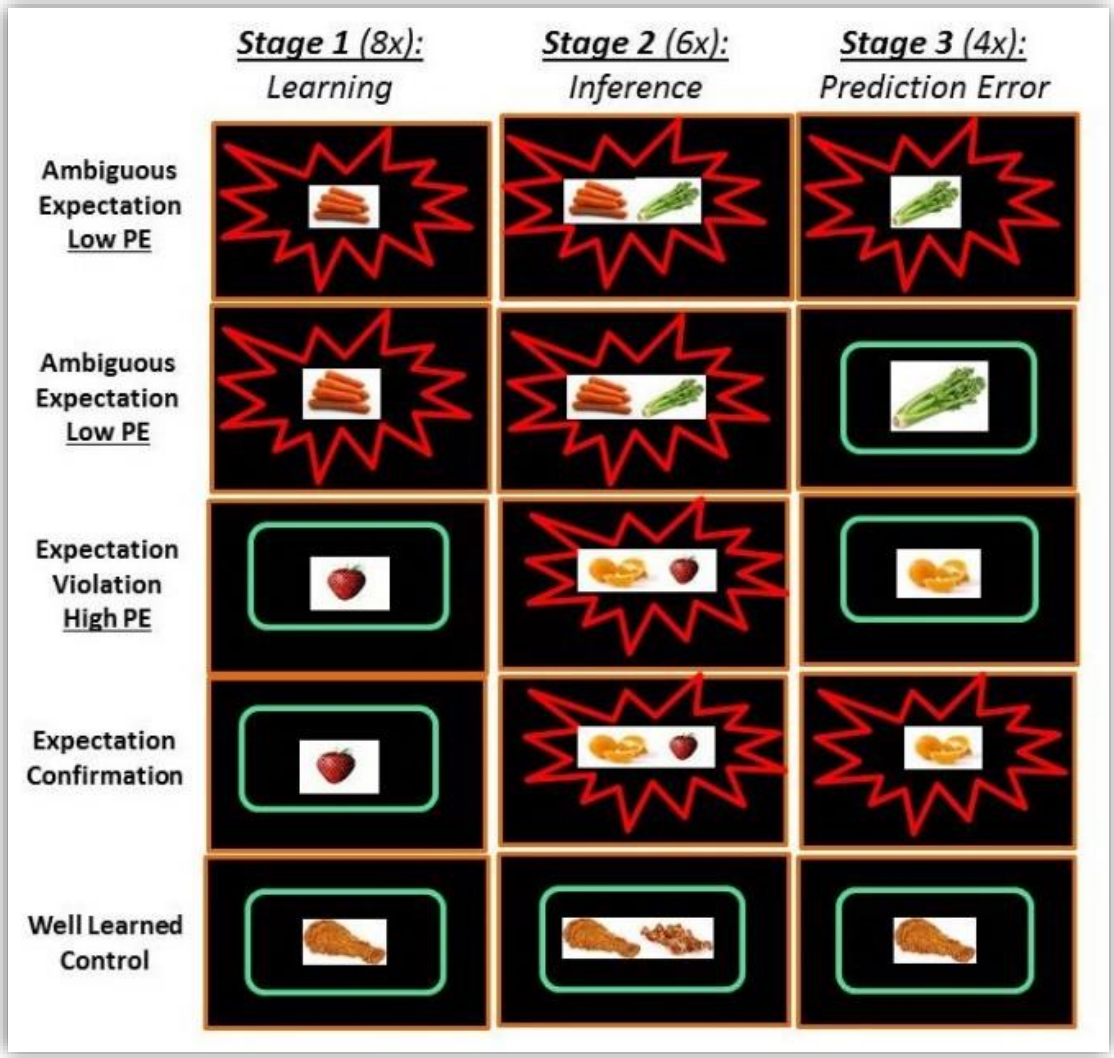

Figure 2-2. Stages of *Allergy* Prediction Error Task.

Stages of Allergy Task: Allergy associations to pairs of food are learned in Stage 1. Participants learn associations for single food items and generate expectancies for the other food item in Stage 2. Participants experience expectation violation or confirmation in Stage 3.

Trials proceed in 3 stages (Fig 2-2), designed to allow associative learning of 14 unique pairs of food, and then for violation of that learning to assess prediction error. In Stage 1, single foods are presented, and subjects learn which yield an allergic reaction. Each food item and its consistent allergy outcome is presented 8 times (e.g., Strawberry - No Allergy). During Stage 2, single items presented in Stage 1 are presented with a paired food item and consistent allergy outcome (e.g., Strawberry and Orange – Allergy). Each pair is presented 6 times. As with all trials, the meal, in this case a food pair, is presented for 3 seconds. Subjects respond yes or no to predict an allergic reaction during the 3 seconds. Outcome information is presented for 2 seconds. A goal of Stage 2 trials is to manipulate expectations that are then tested at Stage 3. Within Stage 2 trials, some food items taught during Stage 1 as not causing allergic reactions are presented in a pair with another item which does lead to an allergic reaction. This should generate an expectation that the newly paired food item is the cause of the reaction. For example, after Stage 1 learning, Strawberry is known to result in No Allergy. Then, seeing it paired with Oranges in Stage 2 teaches that Oranges are the allergy-causing food. During Stage 3, expectation violation trials occur when the new food item (e.g., Orange) is presented as NOT associated with allergy. Within "ambiguous expectation" trials, there is not enough information available for the subject to make a prediction about the new food item in a pair. This is because the single food item from Stage 1 had allergy associated with it and is now paired with a new food item not shown before where allergic reaction continues to be confirmed. These were included to maintain subject interest and not included in current analysis. Expectation confirmation trials are those in which, during Stage 2, allergy reaction is learned for a food pair containing a Stage 1 item not originally associated with allergy and Stage 3 feedback confirms this for subjects. Subjects are also presented with "well learned control" trials where food

associations remain consistent throughout the task (and are a type of expectation confirmation trial though no inference was required). Each food item is presented 4 times in Stage 3. A balanced number of allergy and no-allergy trials was presented. A graded prediction error response is expected for the various trial types. Expectation violation trials at Stage 3 are anticipated to generate the greatest prediction error, ambiguous expectation trials generating relatively less, and expectation confirmation/well learned control trials generating no prediction error response. An important innovation compared to previous studies was that the number of conditions were increase to the number of expectation violations event subject experienced. This was done to increase the statistical sensitivity of the prediction error analyses; To reduced scanner fatigue, Stage 1 (~20min) was done at a computer outside of the scanner. Immediately following it, subjects were transferred to the scanner to complete Stages 2 and 3 (~20min), followed by the structural scan.

Region of Interest

Significant task associations with prediction error at the group level were evaluated in a restricted portion of the brain, using a region of interest mask that was defined by the brain regions most frequently reported to show prediction error activity in meta-analyses (Garrison et al., 2013; Astolfo and Rief, 2017). The Harvard Oxford Probabilistic Atlas was used to generate the prediction error region of interest mask (Fig 2-3A) by combining bilateral individual masks of the following regions, thresholded at 25% probability: Anterior Cingulate Gyrus, Posterior Cingulate Gyrus, Paracingulate Gyrus, Inferior Frontal Gyrus, Frontal Pole, Superior Frontal Gyrus, Middle Frontal Gyrus, Insular Cortex, Nucleus Accumbens, Amygdala, Hippocampus, Caudate, Putamen, Pallidum, Thalamus, Brainstem. The 25% probability threshold means the mask encompasses voxels estimated to be 25% or higher likelihood of belonging to the region.

The 25% threshold is on the side of inclusivity across varying individual anatomy of the subjects (Mazziotta et al., 1995). As the a priori regions of interests analyzed by the previous studies was much more restrictive than the current study, additional exploratory post hoc analyses were conducted using five delusion associated prediction error (D-PE) regions in order to verify that any potential inconsistency with prior findings was not due to increased type I error risk. The D-PE ROIs were based on neural coordinates from the three published task-based fMRI studies (Corlett et al., 2007, Romaniuk et al., 2010; Corlettt and Fletcher). The MNI space coordinates were (x=34, y=34, z=26) for the right Dorsolateral Prefrontal Cortex [r DLPFC]; (x=54, y=18, z=24) for the right Ventrolateral Prefrontal Cortex [r VLPFC]; (x=15, y=15, z=4) for the r Caudate; (x=-15, y=15, z=4) for the l Caudate; and (x=-11, y=-23, z=-9) for the Midbrain. Masks for each seed were generated by creating a 7mm radius sphere centered on the MNI coordinate (Fig 2.3b).

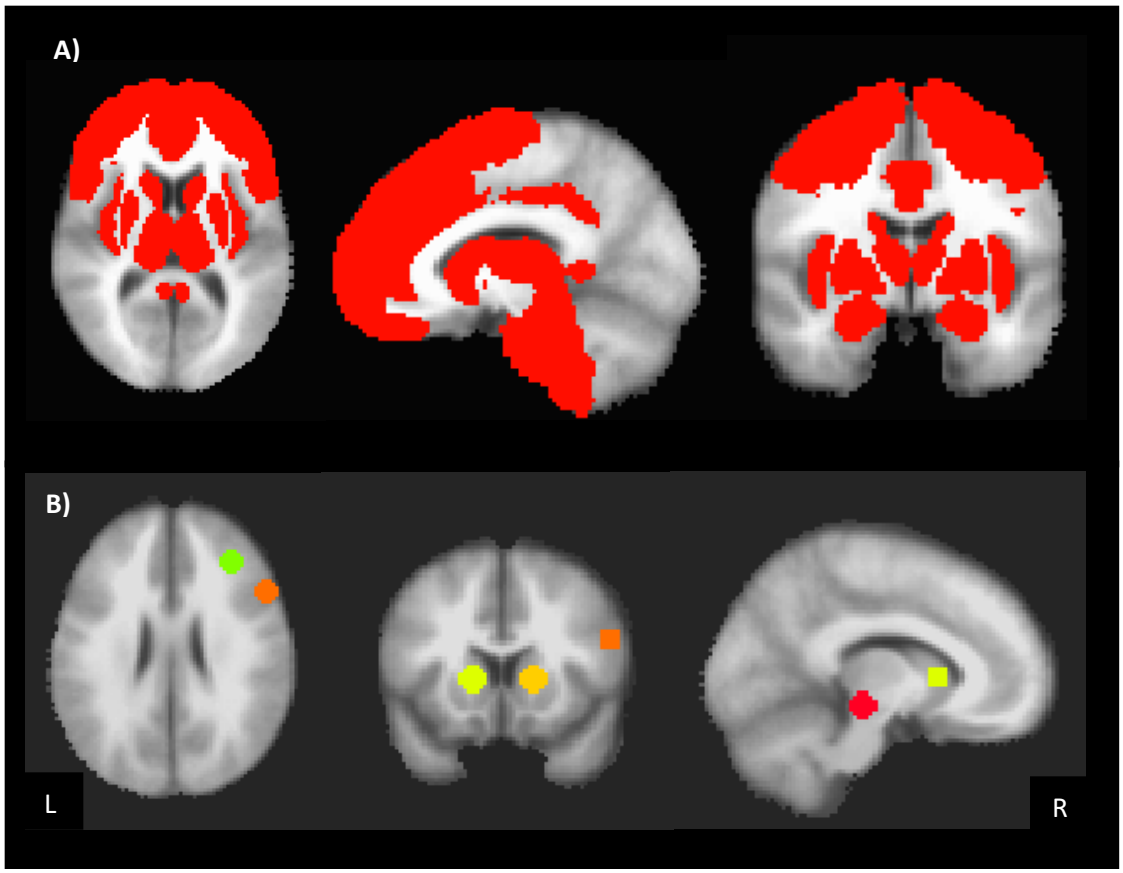

Figure 2-3. Region of Interests Masks for Prediction Task Activation and Connectivity Analyses

A) Red indicates the restricted search area for results of prediction error activation and correlation with delusion severity and is based on meta-analyses of prediction error fMRI tasks in healthy groups.– Included regions were bilateral Anterior Cingulate Gyrus, Posterior Cingulate Gyrus, Paracingulate Gyrus, Inferior Frontal Gyrus, Frontal Pole, Superior Frontal Gyrus, Middle Frontal Gyrus, Insular Cortex, Accumbens, Amygdala, Hippocampus, Caudate, Putamen, Pallidum, Thalamus, Brainstem. B). Delusion associated Prediction Error (D-PE) Masks used in post hoc exploratory analyses: Spherical ROIs based on peak neurocoordinates identified from task based prediction error association with delusion symptoms. Red = midbrain. Orange = ventrolateral prefrontal cortex. Green = dorsolateral prefrontal cortex. Yellow and Green: left and right caudate. Both masks are shown overlaid upon an MNI-152 T1 template.

Neural Activation Analysis of Prediction Error Task

Functional scans from the prediction error task were preprocessed for each subject using SPM12. The scanner acquired and discarded four volumes prior to starting the task. Scans in the

time series were slice time corrected, realigned to the mean image, and co-registered to the participant's anatomical MRI image, spatially normalized to the MNI template and finally spatially smoothed using a 6 mm FWHM Gaussian kernel. Subjects with visually identified scanner artifacts or poor learning as measured by outlier performance accuracy were excluded from subsequent analyses. Primary events of interest for the neuroimaging analysis were the feedback portions of Stage 3 trials and all analyses were restricted to trials with behavioral responses concordant with the expected manipulation of expectation, e.g., trials were only included in analyses if the following criteria were met: "incorrect" feedback for planned expectation violation trials and "correct" feedback was given for planned expectation confirmation (including well learned trials). For each subject, prediction error was measured by creating a BOLD activation contrast map of expectation violation trials > expectation confirmation trials. This contrast should yield information about where activation in the brain is higher for expectation violation relative to expectation confirmation, and this is the conventional contrast employed in the prior studies which this task is based. The commonalities of residual activation from button pressing just before the feedback, as well as viewing similar stimuli on the screen for both conditions should be controlled for in this contrast, leaving activation more strongly associated with just the surprise of the expectation violation. These individual subjects' contrasts (first level analyses) were entered into all subsequent group analyses (second level analyses).

First, a one sample t-test was used to assess prediction error activation within healthy subjects as a manipulation check, confirming that the contrast of expectation violation>expectation confirmation effectively yielded activation in prediction error circuitry. A similar 1-sample t test analysis was conducted on patients for exploratory and descriptive

purposes. Next, to test Hypothesis 1, differences in prediction error activation between psychosis patients and healthy subjects were assessed using a 2-sample t-test. To test Hypothesis 2, within patients, the magnitude of prediction error BOLD response (expectation violation>expectation confirmation contrast maps) was correlated with the severity of delusions as measured from the PANSS Delusion item (P1), LDPS, and PDI scales (three separate correlations were conducted). Age and sex were included as covariates in all analyses. All results were thresholded at $p < .05$, familywise error corrected.

Functional Connectivity Analysis of Prediction Error Task Related Activation

To address Hypothesis 3, the task-related functional connectivity of the prediction error brain regions and any that associate with delusion severity in psychosis patients was characterized and compared between groups using generalized psychophysiological interactions (gPPI). This analysis, implemented with the CONN generalized PPI toolbox in MATLAB (Whitfield-Gabrieli and Nieto-Castanon, 2012), tests how the expectation violation task condition modulates the amount of functional connectivity between the prediction error regions relative to connectivity at baseline. This gPPI analysis contrasts with functional connectivity analyses more widely reported that are conducted on resting state data. In the latter, the entire time series is analyzed for connectivity between brain regions. With gPPI, the data is assessed to determine whether connectivity to brain in a region of interest during the expectation violation trials is altered relative to baseline brain connectivity during rest and unmodeled periods of the task. Preprocessed fMRI data was first denoised in CONN to remove main effects of task, motion, scanner drift and physiological effects via aCompCor (Whitfield-Gabrieli and Nieto-Castanon, 2012). Connectivity maps for each regions of interest (to be determined from results of tests of Hypothesis 1 and 2) were created from each subject's denoised task data. Task-ROI

neural interaction terms were then created through a dot product multiplication of the mean centered time series of each trial type with spatially averaged time series extracted from each ROI. These interaction terms were entered into the general linear model to identify, at the group level, brain regions whose functional connectivity to the ROI is modulated by the task condition of interest (the prediction error trials). Following, the association of gPPI connectivity with delusion symptoms in psychotic subjects, and differences in gPPI connectivity between psychotic and healthy subjects was tested. Age and sex were included as covariates. The primary analysis aimed to evaluate task-based connectivity using the significant delusion associated prediction error clusters identified within the task activation analyses. Post hoc, exploratory analysis was conducted using the literature-based D-PE ROIs. Significance was set at $p < 0.05$, familywise corrected. For the D-PE analysis, familywise error correction was obtained by using a voxel threshold of $p \leq 0.01$ and cluster pFWE$< .01$, accounting for multiple comparisons in the five separate D-PE ROI analyses.

Results

Demographics

In total 47 psychosis patients and 15 healthy controls were included, while 4 patients and 4 healthy controls were excluded due to poor task learning, or excessive scanner motion or image artifacts (3 patients, 1 healthy control). There was a significant difference in the severity of acute positive symptoms for excluded patients compared to included patients, but no significant difference in delusion symptoms or other demographic measures (Appendix A - Table 1). Included psychosis subjects spanned a range of current delusion severity, though no subject had the highest severity rating, consistent with the more stable community sample

targeted for recruitment. A range of lifetime delusion severity was present on the LDPS (Fig 2-4). Full demographic information for included participants presented in Table 2-1.

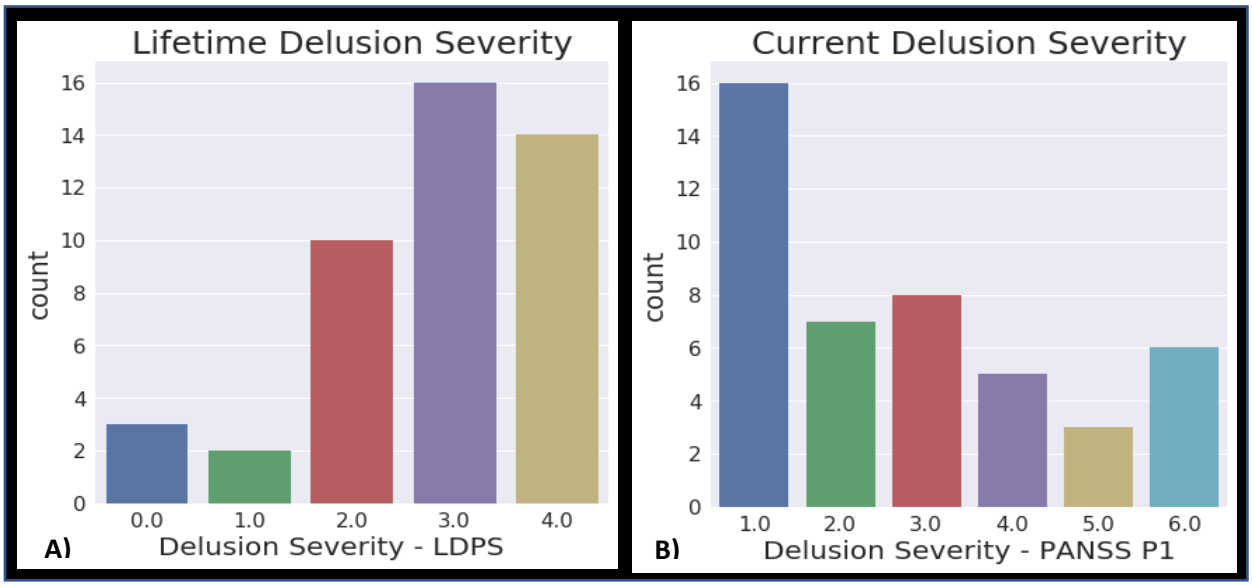

Figure 2-4. Lifetime and Current Delusions Severity in Psychosis Sample

A) Lifetime delusion severity as measured by the LDPS. B). Acute delusion as measured by the PANSS P1 item. For both scales, higher scores indicate worse severity.

Task Performance

As expected, and reported in the Corlett et al studies, a similar task performance was seen across groups. Neither the accuracy rate for expectation confirmation trials nor for expectation violation trials was significantly different between patients and healthy controls (Appendix A - Fig 1). The eight subjects – 4 patients and 4 healthy controls - excluded due to poor learning were identified due to being outliers for accuracy scores for stage 3 expectation confirmation trials and were deemed unlikely to reliably experience expectation violations (Appendix A - Fig 2).

PE Activation in Patients and Healthy Controls

Results of 1-sample t tests of the expectation violation>expectation confirmation contrast maps for each subject showed that prediction error was significantly associated with BOLD activation in several regions for healthy controls, and similar results were found for patients' 1-sample t test (Table 2-2, Fig 2-5). Prediction error was associated with activity in the striatum, midbrain, anterior insula, inferior frontal gyrus, medial superior frontal gyrus, cingulate cortex and medial prefrontal cortex. Within the prediction error meta-analysis based mask, patients had significantly greater prediction error related BOLD response than healthy controls within the anterior cingulate cortex (pFWE = 0.043, Table 2-3, Fig 2-6).

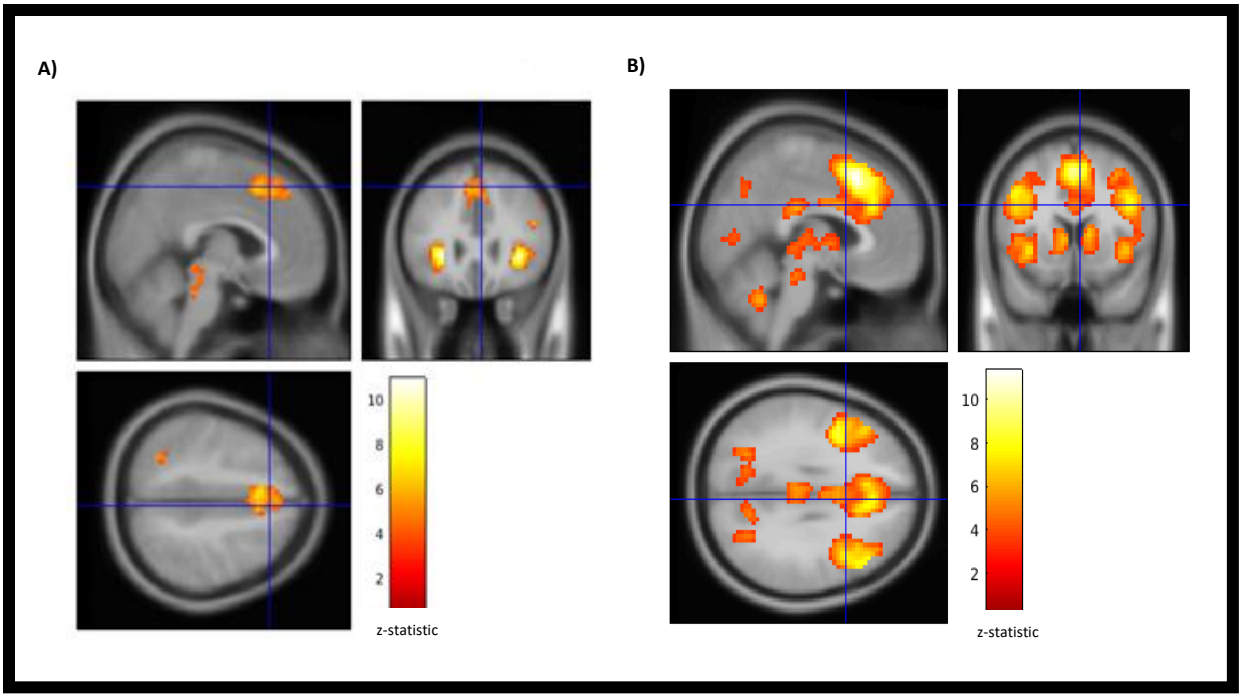

Figure 2-5. One Sample T-tests of the Expectation Violation > Expectation Confirmation trials A*)* healthy controls and B) psychosis patients. Whole brain analysis depicted, though only regions within the prediction error meta-analysis based mask were reported.

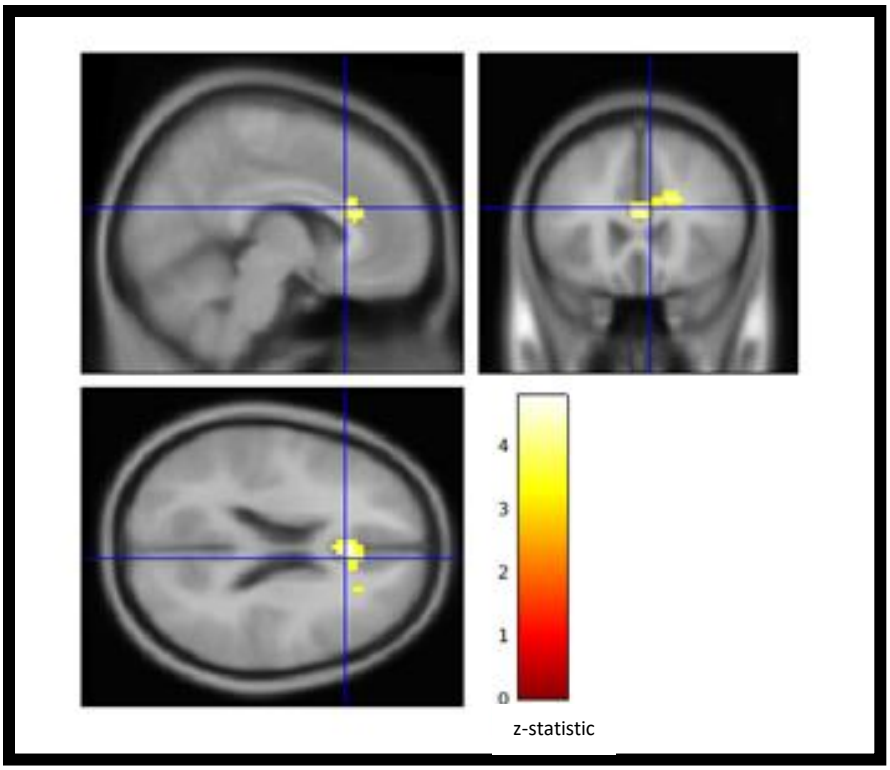

Figure 2-6. Difference in Prediction Error Activation between Healthy Controls and Patients.

A significant difference is found in BOLD response to expectation violation trials compared to expectation confirmation events (e.g., in prediction error related activation) in anterior cingulate cortex (patients > healthy controls).

PE Association with Delusion Symptoms

Results of correlational analyses between delusion symptoms and prediction error activation indicated no significant association. This was true for both current delusion severity (PANSS P1), lifetime delusion severity (LDPS), and self-assessed delusion symptoms (PDI Total Score). As this outcome was discordant with the findings of association in the literature, a few supplementary analyses were conducted to help rule out whether the negative finding may be related to methodological differences from the prior studies. Additional correlation analyses of prediction error activation were conducted using the PANSS item "Unusual Thought Content," and then using distress associated with delusional beliefs (the PDI Distress score).

Next, all correlations were re-ran after excluding bipolar subjects, since no prior work included such subjects (instead being more consistent with a schizophrenia/schizoaffective disorder spectrum). Finally, evaluation of results was restricted to spherical clusters centered on the previously published D-PE neurocorrelates, as the previous studies used more restrictive masks to assess results compared to those planned for use presently in delusion correlations. No association with delusion measures was observed in these additional analyses that attempted to align the findings more closely with prior reports.

Task Functional Connectivity in Patients and Healthy Controls

As no association was found with delusions and prediction error response in the sample, the planned gPPI analysis could not be conducted. Instead an exploratory, post hoc gPPI analysis using the previously published D-PE neurocorrelates was conducted (Corlett et al., 2007; Corlett et al., 2010; Romuniuk et al., 2010). First, groups were compared for connectivity. No significant difference was found for any D-PE region. Groups were assessed individually and then combined (1 sample t tests) to determine whether connectivity was significant for any ROI. There was no significant common connectivity observed across groups during expectation violation, although there was trend-level significant relationship between the rVLPFC and posterior cingulate cortex (Table 2-4; pFWE=0.022).

Association of Task Functional Connectivity in D-PE regions with Delusions Symptoms

Delusion severity was observed to be a significant predictor of connectivity changes for expectation violation feedback events for the D-PE regions (Table 2-5). Current delusion severity (PANSS P1) was positively associated with the gPPI connectivity between the r DLFPC ROI and the left intracalcarine cortex (pFWE=.002). Patient self-assessed history of delusion

burden (PDI Total) was negatively associated with gPPI of the R caudate with the left orbitofrontal cortex. (pFWE=.005). The clinician assessed lifetime severity of delusions (LDPS) was not found to be significantly associated with expectation violation gPPI, but there was a trending relationship to gPPI between the R Caudate ROI and the left thalamus (pFWE=.025).

Table 2-1. Demographics and Clinical Characteristics for Included Subjects

	PTS	HC	P-value
N	47	15	
Male/Female	21/26	6/9	.750
Avg. Age (years)	38.8 (11.4)	36.9 (13.6)	.594
Avg. Daily CPZ	280.7 (319.8)		
PANSS Delusion (P1)	2.8 (1.8)		
PANSS Positive	15.6 (6.4)		
PANSS Negative	12.6 (5.0)		
PANSS General	28.5 (8.6)		
PANSS Total	56.7 (17.0)		
LDPS Delusion Severity	2.80 (1.1)		
PDI Total	106.7 (70.7)	52.4 (43.5)	.007
PDI Endorse	10.0 (5.4)	5.4 (4.0)	.004
PDI Distress	28.3 (20.3)	12.5 (10.5)	.006
PDI Preoccupation	30.3 (20.8)	14.4 (12.2)	.008
PDI Conviction	33.8 (23.7	17.2 (14.8)	.016

Values in parenthesis are standard deviations. *Abbreviations*: CPZ – Chlorpromazine Equivalents of Antipsychotic Medication Dose, HC – Healthy Control, NS – Not Significant, PANSS -Positive and Negative Symptom Scale, PDI – Peters' Delusion Inventory, LDPS – Lifetime Dimensions of Psychosis Symptoms, PTS - Patients

Table 2-2. Neural Activation associated with Prediction Error in Healthy Controls and Psychosis Patients

Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE (Cluster Level)
Neurocorrelates of Prediction Error Events in Healthy Controls			
77	-27, 26, -1	L Anterior Insula – Inferior Frontal Gyrus – Frontal Operculum	<0.001
53	33, 26, -1	R Anterior Insula – Inferior Frontal Gyrus – Frontal Operculum	0.004
36	45, 29, 23	R Middle Frontal Gyrus – Inferior Frontal Gyrus	0.024
219	-6, 17, 47	Supplementary Motor Cortex – medial Superior Frontal Gyrus	<0.001
38	-3, -31, -19	Brainstem – Thalamus	0.020
Neurocorrelates of Prediction Error Events in Patients			
1171	6, 20, 47	Supplementary Motor Cortex – medial Superior Frontal Gyrus – Middle Cingulate gyrus	<0.001
1128	33, 23, -4	R Anterior Insula – Inferior Frontal Gyrus – Middle Frontal Gyrus – Frontal Operculum – Precentral Gyrus	<0.001
1055	-33, 17, -1	L Anterior Insula – Inferior Frontal Gyrus – Middle Frontal Gyrus – Precentral Gyrus – Frontal Operculum – Oribital Frontal Cortex	<0.001
646	12, 5, 11	Caudate – Putamen – Pallidum – Thalamus – Brainstem	<0.001
79	-3, -28, 26	Posterior/Middle Cingulate Gyrus	0.004

Abbreviations: L – Left, pFWE – p-value Family Wise Error, R – Right

Table 2-3. Differences in Neural Activation associated with Prediction Error Between Healthy Controls and Psychosis Patients

Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE (Cluster Level)
Neurocorrelates of Prediction Error Events in Healthy Controls vs Patients			
45	0, 23, 20	Anterior/Middle Cingulate Gyrus	0.043

Abbreviations:– p-value Family Wise Error

Table 2-4. gPPI of Expectation Violation Events in literature-based Delusion-associated Prediction Error ROIs in Psychosis Patients and Healthy Controls

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE
Patients				
R VLPFC	138	-30, -68, 40	L Superior Lateral Occipital Cortex	.003906
L Midbrain	109	42, -40, 50	R Posterior Central Gyrus – SMG	.010763
Healthy Controls				
R DLPFC	173	-24, -22, 50	L Pre-Central Gyrus – White Matter	.005524
Healthy Controls and Patients				
R VLPFC	146	-0, -32, 32	Posterior Cingulate Gyrus	.022151

Abbreviations: DLPFC – Dorsolateral Prefrontal Cortegx, DLPFC – Dorsolateral Prefrontal Cortex, D-PE – Delusion associated Prediction Error, EV – Expectation Violation, gPPI – generalized psychophysiological interactions, L – Left, pFWE – p-value Family Wise Error, , R – Right, ROI -Region of Interest , VLPFC – Ventrolateral Prefrontal Cortex

Table 2-5. gPPI of Prediction Error Events in D-PE ROIs associated with Delusions

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE
Associated with Current Delusion Severity (PANSS P1)				
R DLPFC	200	-14, -80, 20	L Intracalcarine Cortex	.002133
Associated with Lifetime Delusion Severity (LDPS)				
R Caudate	127	-08, -20, 02	L Thalamus	.025303
Associated with Self-Assessed Delusion Severity (PDI Total)				
R Caudate	185	-24, 14, -28	L Orbitofrontal Cortex	.005139

Abbreviations: DLPFC – Dorsolateral Prefrontal Cortex, D-PE – Delusion associated Prediction Error, EV – Expectation Violation, gPPI – generalized psychophysiological interactions, L– Left, pFWE – p-value Family Wise Error, PANSS – Positive and Negative Symptom Scale, PDI – Peters Delusion Inventory, R – Right, ROI -Region of Interest

Discussion

This study aimed to provide insight into a possible neurocognitive pathway underlying delusions by identifying the brain networks involved in prediction error response and investigating their dynamics. Principally, it sought to extend the findings of prediction error neurocorrelates being associated with delusion symptoms in a large transdiagnostic psychosis sample. This meant including bipolar with psychosis subjects, a group not previously incorporated. An additional aim was to explore the relationship of delusion symptoms to the task based functional connectivity of brain regions involved in prediction error, a neural characteristic

that also has not been studied. The findings were: (1) in both patient and healthy controls, expectation violation events were associated with activation in several brain regions identified in previous prediction error studies. (2) Patients had greater activation than healthy controls in the anterior cingulate cortex during prediction error. (3). There was no evidence that prediction error neural activity, including activation and connectivity, was associated with delusion symptoms. (4) An exploratory analysis of task connectivity using the D-PE regions identified by previous studies found suggestive but inconsistent evidence that altered neural dynamics, in the form of connectivity during prediction error, is associated with delusion symptoms.

Prediction Error Response

Prediction error activation was identified in several region throughout the brain during the causal learning association task. For this study specifically, prediction error response was conceptualized as BOLD response during expectation violation events compared to expectation confirmation events. Within both patients and healthy controls, the prediction error response was seen most robustly in anterior insula, midbrain, and middle frontal cingulate cortex. This activation pattern was concordant with many of the prediction error neurocorrelates identified in prior neuroimaging studies of reinforcement learning in healthy participants (Garrison et al., 2013; Chase et al. 2015; Astolfo and Rief, 2017). This confirms the task was likely engaging the intended neurocognitive system.

When examining differences between patients and healthy subjects on prediction error related activation, the anterior cingulate was found to have a stronger prediction error response in the patient group. This region was not examined in the Corlett studies, so this may be a novel finding. While other studies have also observed abnormal prediction error response in the

cingulate cortex in psychosis groups, those observations were of reduced activation in anterior cingulate, and subjects were only inclusive of those with schizophrenia (Kerns et al. 2005) and or were first episode psychosis patients (Murray et al. 2011). The latter study also reported decreased prediction error activation in the midbrain, striatum, hippocampus and insula. Examination of other neuroimaging studies investigating prediction error in psychosis shows further mixed findings, with decreased midbrain and striatal activation frequently reported (Romanuik et al., 2010; Gradin et al., 2011) but not consistently found in the striatum (Corlett et al., 2007; Ermakova et al., 2018), midbrain (Koch et al., 2010; Schlagenhauf et al., 2013; Reinen et al., 2016) or both (Culbreth et al.2017; Katthagen et al. 2018).

One source of discrepancy could be task-dependent. Differences in processing of reward vs non-reward prediction error may be an important factor (Waltz et al., 2009; Morris et al, 2012). Reinen et al., (2016) reported that unmedicated schizophrenia patients had attenuated prediction error response in the medial prefrontal cortex, striatum, and medial temporal lobe when learning to predict rewards but not to avoid losses. In another study of medicated chronic schizophrenia subjects similar reductions were reported in reward prediction error response for the anterior cingulate cortex in addition to striatum, medial prefrontal cortex, medial temporal gyrus, but found loss avoidance prediction error to be associated with reduced activation in the hippocampus and insula (Koch et al. 2010). The subjects recruited in the current study received no direct reward or loss feedback, but instead received an implicit acknowledgement of error (signaled by the food stimuli being highlighted in red which designating an allergic reaction occurred). Thus, it is worth noting that the causal learning association task used in this study and related previous studies (Fletcher et al., 2001; Turner et al.; 2004; Corlett et al., 2007; Corlettt and Fletcher) generates unsigned prediction errors that are more characteristic of surprise

prediction errors than positive or negative valence reward prediction error. This may result in a more muddled signal, when taking into consideration the studies suggesting reward vs avoid loss tasks result in different neural activation alterations for errors.

A recent meta-analysis of prediction error studies in healthy subjects provides evidence supporting the existence of two systems encoding valence prediction errors and a third distinct neural system of the anterior cingulate cortex, anterior insula and dorsal striatum that registers surprise, regardless of valence (Fouragnan et al, 2018). According to this breakdown the allergy task used for this study would more strongly engage the latter neural system, which is consistent with the strong activation in the insula and cingulate cortex. However, a combination of subtle reward learning and reality monitoring was likely at play in the participants, as subjects were driven by a mix of internal motivation to be correct, please investigators, and receive full compensation for participating in the task. This form of motivational salience for learning is likely characteristic of the implicit learning that subjects engage in the real world and may account for the broad range of prediction error associated neural circuits engaged by the subjects.

Delusion Symptoms Not Associated with Prediction Error Response

Prediction error during associative learning was not significantly correlated with history of lifetime delusions or current delusional severity in this chronic psychosis sample. This finding contrasts with the observation by Corlett et a.l (2007) that decreased prediction error response in ventrolateral prefrontal cortex was associated with unusual thought content in first episode psychosis patients and the report by (Gradin et al., 2011) that prediction error activity in the insula, midbrain and amygdala-hippocampal complex was negatively associated with positive psychosis symptoms (hallucinations plus delusions) in chronic schizophrenia patients. However,

other studies have also failed find a direct relationship with prediction error and delusion symptoms in first episode psychosis (Murray et al, 2008; Ermakova et al., 2018) and chronic psychotic patients (Romaniuk et al., 2010). The conflicting results may be due to several reasons. The previous studies employed diverse methods, including differences in sample selection, associative learning fMRI tasks, and regional analysis of prediction error response. The sample size of psychosis patients was notably small for all the earlier studies (all $n<25$), elevating the risk for false positives (Eklund et al., 2016; Turner et al., 2018). Also, the previous studies examined delusions using a range of measures. An exploratory analyses was conducted of unusual thought content (measured by PANSS) and delusion dimensions including distress, preoccupation, and conviction (measured by PDI) to provide more comprehensive assessment of delusion severity, boosting likelihood of overlapping with prior studies, and potentially determining an optimal delusion severity associate of prediction error activation. However, prediction error response was not associated with any of the delusion measures. An important consideration was the use of a transdiagnostic psychosis population spanning the schizophrenia-bipolar spectrum. This was chosen because these diagnostic groups share delusions symptoms and genetic risk for these illnesses (Badner & Gershon, 2002; Pini et al., 2004; Purcell et al., 2009). While there is some evidence that the propensity and severity of delusions subtypes differ across the groups (Kempf et al., 2005; Mancuso et al., 2015, Picardi et al., 2018), it is unknown whether delusion types may have differential neural underpinnings. Hence, a starting point is to assess across delusion types in a sample of patients reporting them in a similar manner phenomenologically. However, delusion or other symptom characteristics aside, both common and different brain pathologies have been observed between psychosis diagnoses (Tamminga et al., 2014). Thus, the initial analysis assumption that a homogenous prediction error response

exists across the diagnoses and is associated with general delusion pathology may not be warranted. As none of the previous studies incorporated bipolar patient with psychosis, a restricted analysis excluding these subjects from the sample was conducted, but no significant relationship was observed within the remaining patient group that more closely resembled samples in prior studies.

Suggestive Evidence Linking Prediction Error Dynamics with Delusions

Our preliminary investigation into task related connectivity during prediction error found no significant differences between patients and healthy controls in the five prediction error regions explored. Although no significant patient difference was observed, a significant positive association was found such that greater current delusion severity predicted greater connectivity between the dorsolateral prefrontal cortex and primary visual cortex during expectation violation events. This may be related to the modulation in visual attention given to surprising events. Although no directly comparable studies have been done in psychosis, Schott et al. (2015) did report in a small sample of patients with schizophrenia that activation of the orbitofrontal cortex was increased in response to recognition of novel stimuli in a visual memory paradigm. However, they found that acute delusion symptoms were associated with interaction of the hippocampus and orbitofrontal cortex with anterior cingulate cortex and not with the ventral striatum. The result was not replicated in the other assessments of delusions. Instead a separate finding was that connectivity of caudate with orbitofrontal cortex was negatively associated with PDI Total (the patient's self-assessed measurement of historical delusion burden both in terms of prevalence of delusional beliefs and cognitive/emotional impact).

On the other hand, despite Type 1 error protection steps taken, it is also possible the findings are spurious. Results from the connectivity study are severely limited for a methodological reason. PPI is not optimal for detecting effects in event related designs, which was the type of fMRI task conducted. This is due to a low signal to noise ratio. Limitations on the ability to fully explain task variance for both the prior and current study make them at heightened risk for spurious results when PPI is used to detangle subtle cognitive effects (Orielly et al, 2012).

Limitations

There are several caveats to this study which should be noted. First, there are several challenges for the field in assessing prediction error response. One obstacle is providing an fMRI task that provides enough expectation violations experiences while accounting for normal habituation to surprise and that is of reasonable length and difficulty for both healthy and clinically ill psychiatric subjects. Thus, the administered task, though closely modeled from the Corlett et al 2012 study, was lengthened to provide participants more surprise events and enable greater power for analysis of prediction error response and its associated dynamics. Secondly, use of antipsychotic medication is a relevant confound in the analysis. This is particularly true, as many of the investigated prediction error regions are the target of D2 receptor antagonism mediating theorized therapeutic relief. Furthermore, prediction error response was investigated in patients who were still experiencing various psychotic symptoms after being medically stabilized. How medication may restore or transform prediction error functioning needs further clarification. Third, as gPPI is a generally low powered for event related design such as this study, it makes the risk high for false negatives and false positives. Coupled with the relatively low number of expectation violation events available for analysis in prediction error studies the

findings should be treated with significant caution. Future investigations would greatly benefit from a fMRI task that can powerfully capture prediction error response during associative learning within a design more suitable to study both the functional and effective connectivity dynamics of prediction error. In addition, although there was not a significant difference in delusion severity between groups, the excluded participants were more impaired according to total severity of positive symptoms. Hence the findings may not be generalizable to more severely ill patients. Lastly, although the study was much larger than many previous prediction error studies in psychosis, it was not powered to detect potential heterogeneous effects between diagnostic subgroups and the notably smaller healthy sample limits the power to characterize differences between patients and controls.

Conclusions

This study expands upon a growing field of literature investigating the association of prediction error with psychosis symptoms, specifically delusions. The data does not support that delusion symptoms are associated with a common neural mechanism of prediction error dysfunction in psychosis patients. Although, preliminary results suggest delusion symptoms may be related to subtle prediction error dynamics.

Chapter III: Prediction Error Resting State Functional Connectivity and its Association with Delusion Symptoms

Introduction

The disconnection hypothesis asserts that psychotic symptoms emerge from abnormally connected functional brain networks (Friston and Frith, 1995). Numerous neuroimaging studies have found evidence of abnormal connectivity in psychosis patients compared to healthy controls (Satherswaite and Baker, 2015; Mwansisya et al., 2017). Widespread dysconnectivity has been observed in psychosis, with results including the frontal cortex, basal ganglia, sensory cortex, language association areas, and cerebellum. These findings have been reported in both neuroimaging studies during cognitive tasks and during resting state. Resting state functional magnetic resonance imaging (rs-fMRI) allows for the characterization of intrinsic connectivity patterns of neural systems while individuals are not engaged in a specific task. Resting state networks have been found to have a high degree of reliability (Shehzad et al., 2009; Jann et al., 2015) and good spatial correspondence with networks connected during task-based cognition (Smith et al., 2009). These studies provide evidence that networks of the brain continue to be dynamically active even when seemingly at rest. Further, as resting state scanning can be more consistently implemented across sites and studies, it is a useful method for investigating brain networks involved in cognition.

The potential relationship between spontaneously occurring delusion symptoms and spontaneous neural activity has not been extensively explored. It has been theorized that

delusions may arise as an epiphenomenon of impaired salience and self-attribution mechanisms during periods of mind wandering and periods of self-reflective cognition (Cahill et al., 1996; Bentall et al., 2011; Shin et al., 2015). Examination of the resting state dynamics provide an approach to investigate how endogenous brain activity during these periods may be disrupted during delusional states. It is possible that in psychosis the normal resting state networks may be altered in ways that cause delusions or increase vulnerability to experiencing them. This may be supported by the investigations of psychotic delusions in schizophrenia patients that have implicated altered resting state connectivity in several brain regions, though this work is limited thus far. One approach conducted a whole brain network analysis of 176 schizophrenia patients which found delusion severity - measured by the Positive and Negative Syndrome Scale (PANSS) - was associated with altered connectivity of the thalamus with the pre/postcentral gyrus, superior medial and middle frontal gyrus (Li et al., 2017). In a different study of 46 patients examining the insula, resting state connectivity of the right posterior insula to thalamus was found to be associated with delusions as measured by the PANSS (Chen et al., 2016). An independent component analysis of the salience resting state network in 26 psychosis patients reported that hypoconnectivity within the striatum was associated with delusions as measured by the PANSS (Orliac et al.,2013). This contrasted with the findings from another study examining intrinsic connectivity within the basal ganglia of 21 patients, which reported hyperconnectivity within the dorsal striatum as associated with delusions as measured by the PANSS (Sorg et al., 2013). The assorted findings suggest abnormal engagement of the salience and central executive resting state network may be associated with delusions. However, the sparse samples and potentially inconsistent results suggest that more research is needed to clarify how resting state brain activity may be associated with delusion symptoms.

Despite the mixed findings, there is general support for dysconnectivity in the striatum and frontal cortex as a potential mechanism for delusions particularly when resting state studies are coupled with task-based neuroimaging observations, as these brain regions have also previously been associated with delusions in cognitive tasks (Corlett et al. 2007; Romaniuk et al., 2010; Corlettt and Fletcher). Within these fMRI studies, prediction error associated regions – specifically the right lateral prefrontal cortex, midbrain, and striatum – were reported as abnormally engaged in delusional patients during reinforcement learning paradigms. Prediction error within these studies is a neural measure of the difference between an individual's expectations and the outcomes they experience. Prediction error has been posited as an important neural correlate for both learning and reality monitoring with broad evidence that it is also disrupted in psychosis (Murray et al., 2008; Morris et al., 2012; Gradin et al., 2013). However, the mechanism for how these regions may interact to engender psychotic symptoms is not yet understood. As the dynamics of the brain networks associated with prediction error remain unclear, more in-depth research of the system in robustly sized samples of both healthy and psychotic individuals is necessary to answer if and how the prediction error system is a relevant mediator of delusion symptoms.

To understand the specific role of prediction error neurocorrelates in psychotic symptoms, it is important to characterize both the context-dependent activity and context-independent activity of the prediction error network. This can be done by investigating brain connectivity in neuroimaging tasks that induce prediction error cognitive responses (context-dependent) and by investigating the intrinsic connectivity of the prediction error system (context-independent), respectively. For the latter, region-of- interest (ROI) or seed-based connectivity analysis is one method which can be useful to investigate targeted hypotheses about resting state

brain activity. This approach informs us of how the selected brain region of interest is functionally connected – via correlation of endogenous neural activity - with other brain regions. Based on the prior task-based and resting state findings, it is reasonable to predict that the resting state connectivity of delusion-associated prediction error (D-PE) brain regions (e.g. the midbrain, striatum, and lateral prefrontal cortex) will be disrupted in psychosis. Such a prediction, if true, may provide needed perspective on how delusions may arise in psychosis. Starting with the hypothesis of dysregulated salience as a core mechanism for psychosis proposed by (Kapur, 2003), delusions may arise due to abnormal endogenous connectivity of the striatum and midbrain - regions identified as important in numerous salience detection tasks (Schultz, 2000; Zink et al., 2003; Wise, 2004; Bromberg-Martin et al., 2010). In this scenario both a potential disruption in endogenous connectivity and in salience processing could result from thhe abnormal dopaminergic transmission observed within the striatum for psychotic patients (Howes et al., 2012). Thus, it is hypothesized that patients may have decreased intrinsic connectivity of striatum to with regulatory regions such as the prefrontal cortex, and increased connectivity to perceptual sensory and association cortices. One way to address these hypotheses is via examining the whole brain resting state connectivity of prediction error regions and their association with delusions symptoms.

Another question that arises is whether the relationship of delusions to endogenous brain activity within these regions is a transdiagnostic or disorder-specific phenomenon. It is possible the same or distinct neural alterations in the prediction error system underlie delusion symptoms across psychotic disorders. Delusion symptomology and treatments are shared across psychotic disorders and current research suggests there is significant overlap in psychotic disorder genetics and neurobiology (Appelbaum et al., 1999; Badner and Gershon, 2002; Purcell et al., 2009;

Keshavan et al., 2011). These observations suggest that a common neural mechanism for delusions may exist within psychotic disorders. This is further supported by the shared observation from the prediction error task studies that midbrain and lateral prefrontal cortex activation is linearly associated with delusion symptom severity using varied cohorts (e.g. delusion spectrum beliefs in non-clinical participants, first-episode psychosis, and chronic schizophrenia patients). A key metric of interest that would lend further support to the role of connectivity of prediction error brain regions in delusions is one of corresponding magnitudes: greater connectivity alteration should correlate with greater delusion severity.

Study Aims

Therefore, a useful way forward may be to study the intrinsic connectivity of the delusion associated prediction error system in a large transdiagnostic psychotic sample. Such an opportunity exists in the Bipolar Schizophrenia Network for Intermediate Phenotypes (B-SNIP) study (Tamminga et al., 2014), a multi-site study of schizophrenia (SCZ), schizoaffective (SAD), and bipolar disorder with psychotic features (BDP). In the present study, a seed-based analysis of B-SNIP1 resting state data is conducted to address the following questions: (1) what is the intrinsic connectivity of delusion-associated prediction error regions in healthy and transdiagnostic psychosis subjects, and (2) is the intrinsic connectivity of these regions associated with severity of current delusions.

Methods

Participants

Data for resting state analyses was obtained from the completed Bipolar & Schizophrenia Network for Intermediate Phenotypes 1 (B-SNIP 1) multisite study. Subjects were recruited

following IRB approval from each of the five study sites, and the larger study has been described elsewhere in detail (Tamminga et al., 2014). Psychosis patients were clinically characterized, and all subjects had a panel of biomarkers assessed including the neuroimaging reported here. The psychosis patients and healthy volunteers were recruited using local advertising. Inclusion and exclusion criteria were age 18-60, able to provide written informed consent, estimated IQ > 60, no current substance abuse disorders or major neurological/cognitive/cerebrovascular-affecting disorders, and no significant head trauma history. Healthy controls had no personal history of any psychiatric disorder or first-degree relative with schizophrenia, schizoaffective disorder, or mood disorder.

Clinical Assessments

Trained clinical raters confirmed DSM-IV diagnosis of Schizophrenia, Schizoaffective, or Bipolar Disorder with Psychosis using the SCID-IV (First, 2000a; First, 2000b). Raters also administered the Positive and Negative Syndrome Scale (PANSS) which assesses the severity of a range of psychotic symptoms in the last week (Kay et al., 1987). Current delusional severity in patients was characterized using the "Delusions" PANSS item (score range 1-7).

Imaging Data Acquisition and Preprocessing

Subjects underwent a 5-min rs-fMRI scan in 3T scanners with closely aligned acquisition parameters (Appendix B - Table 1). Subjects were instructed to remain still, stay awake and keep their eyes focused on a crosshair for the scan's duration. Wakefulness was confirmed with the subjects following the scan. To allow for scanner stabilization, the initial 6 images were discarded. Using the SPM based CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012), the time series was aligned, slice-time corrected, normalized to MNI space

(Montreal Neurological Institute), and smoothed with a 8mm FWHM Gaussian kernel, with 2x2x2 mm resampled voxel size. The data was denoised using regression of subject's white matter and CSF (aCompCor), scrubbing signal, motion + 1st order derivatives, and a linear and 2nd order polynomial drift term, and subsequently band-pass filtered at 0.008-.1 Hz based on recent reports on the effect of filtering on resting state (Goto et al., 2015). To additionally ensure scan quality, subjects with visually identified artifacts and framewise motion > 3mm were excluded from analyses.

Delusion associated Prediction Error (D-PE) Neurocorrelates

Whole brain connectivity maps for the delusion associated prediction error (D-PE) regions were created for each subject. Five D-PE regions of interests (ROI) were based on neural coordinates from published task-based fMRI studies (Corlett et al., 2007, Romaniuk et al., 2010; Corlettt and Fletcher). The MNI space coordinates were (x=34, y=34, z=26) for the right Dorsolateral Prefrontal Cortex [r DLPFC]; (x=54, y=18, z=24) for the right Ventrolateral Prefrontal Cortex [r VLPFC]; (x=15, y=15, z=4) for the r Caudate; (x=-15, y=15, z=4) for the l Caudate; and (x=-11, y=-23, z=-9) for the l Midbrain. Masks for each seed were generated by creating a 7mm radius sphere centered on the MNI coordinate (Fig 3-1).

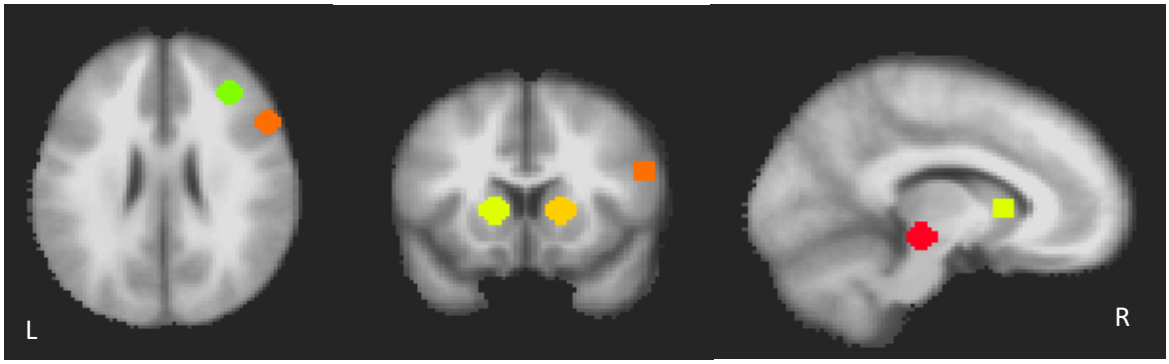

Figure 3-1. Region of Interests for D-PE Resting State Functional Connectivity Analyses

Spherical ROIs based on peak neurocoordinates indentified from task based prediction error association with delusion symtpoms. Red = midbrain. Orange = ventrolateral prefrontal cortex. Green = dorsolateral prefrontal cortex. Yellow and Green: left and right caudate. Maske shown overlaid upon an MNI-152 T1 template.

Functional Connectivity Analysis of D-PE Resting State Seeds

Using the CONN toolbox, timeseries across voxels within each D-PE region were averaged, and then correlated with all remaining voxels in the brain to create a whole brain functional connectivity map for each seed region. These maps were converted to Fisher z- scores. The primary analyses tested 1) whether the psychosis group differed from the healthy group on connectivity, and 2) where connectivity to the D-PE region was significantly predicted by current delusion severity across the transdiagnostic sample. For (1), differences between healthy and psychotic subjects were tested using ANCOVAs (with age, sex, site, and motion [FDpower] covariates). Following these results, a supplemental analysis was conducted for descriptive purposes given the groups appeared to have more connectivity similarity than difference. For this, a 1-sample t test was conducted on the entire sample combined, with the same covariates as in the primary between group comparison and by additionally using a weighted average to equalize effect of patient and healthy cohorts. For (2), each set of

connectivity maps for patients was entered into a separate multiple linear regression with the delusion severity item score as an independent predictor variable. Age, sex, recruitment site, mean framewise displacement – FDpower - as a measure of micromotion (Power et al., 2012) were additional predictors of no interest. Exploratory analyses examined for the potential differential group effects with D-PE connectivity and delusion severity, and an adjustment for the effect of antipsychotic medication (using average daily chlorpromazine – computed per Andreasen et al., 2010). In the exploratory analyses, regions with significant diagnostic interaction effects were followed up with post-hoc pairwise comparisons. For all analyses, significance was set at $p < 0.05$, familywise corrected, obtained by using a voxel threshold of $p \leq 0.01$ and cluster pFWE< .01, controlling for multiple comparisons in the five conducted seed connectivity analyses.

Results

Sample Characteristics

In total, 338 psychosis patients (245 of whom had information on current antipsychotic usage) and 186 healthy controls were included, while 39 patients and 27 healthy controls were excluded due either to excessive motion or image artifacts. Demographic information and clinical characteristics are presented for the included subjects (Table 3-1), excluded subjects (Appendix B -Table 2), diagnostic subgroups (Appendix B -Table 3) and medication subsample (Appendix B - Table 4). Within the primary analysis the included patients spanned a range of delusion severity, though no subject had the highest severity rating, consistent with the more stable community sample targeted for recruitment (Fig 3-2, Appendix B – Fig 1).

Table 3-1. Demographics and clinical characterization of included participants

	Included Patients	Included Healthy	P-value
N	338	186	
Male/Female	160/178	72/114	.057
Avg. Age (year)	35.9 (12.2)	37.8 (12.4)	.091
PANSS Delusion	2.7 (1.4)		
PANSS Positive	15.9 (5.3)		
PANSS Negative	14.6 (5.1)		
PANSS General	31.9 (8.6)		
PANSS Total	62.4 (16.5)		
GAF	52.1 (13.3)	85.9 (6.8)	<.001
mFDpower (mm)	0.21 (0.13)	0.17 (0.10)	<.001

Abbreviations: Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, GAF – Global Assessment of Function, PANSS -Positive and Negative Syndrome Scale. Values in parenthesis are standard deviations.

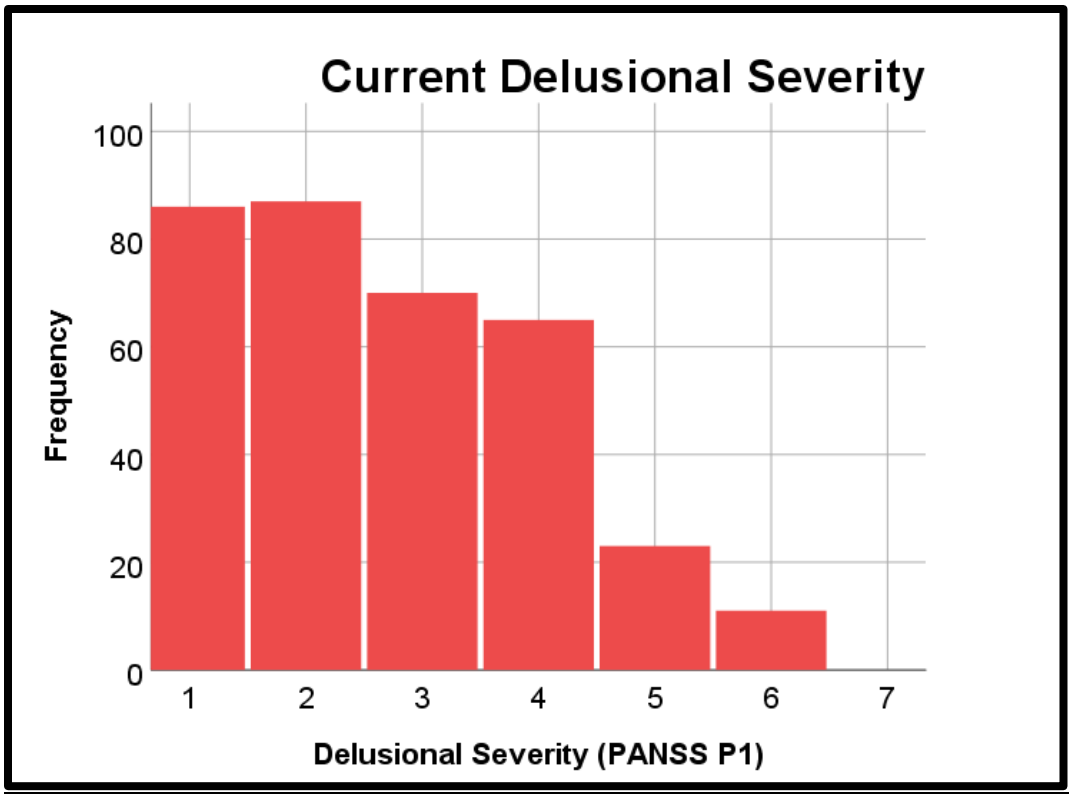

Figure 3-2. Distribution of Delusion Severity in B-SNIP1

Current delusion severity was measured by item P1 from the Positive and Negative Syndrome Scale (PANSS). Patients were represented across the severity spectrum with delusion severity ranging from no present symptoms (1) to severe delusion symptoms (6).

D-PE RS Functional Connectivity Differences between Patients and Healthy Controls

Results of ANCOVA to compare the connectivity maps of psychosis patients to those created for healthy controls revealed that the full patient sample had significantly weaker negative connectivity for r and l caudate seeds to the precuneus (Fig 3-3, Table 3-2). When controlling for the effects of anti-psychotic medications in the subsample with this information available, a similar finding was present as well as significant reductions in the positive connectivity of the r

DLPFC with pre- and post-central gyri, and of r caudate with the cerebellum in patients compared to healthy controls (Appendix B - Table 5).

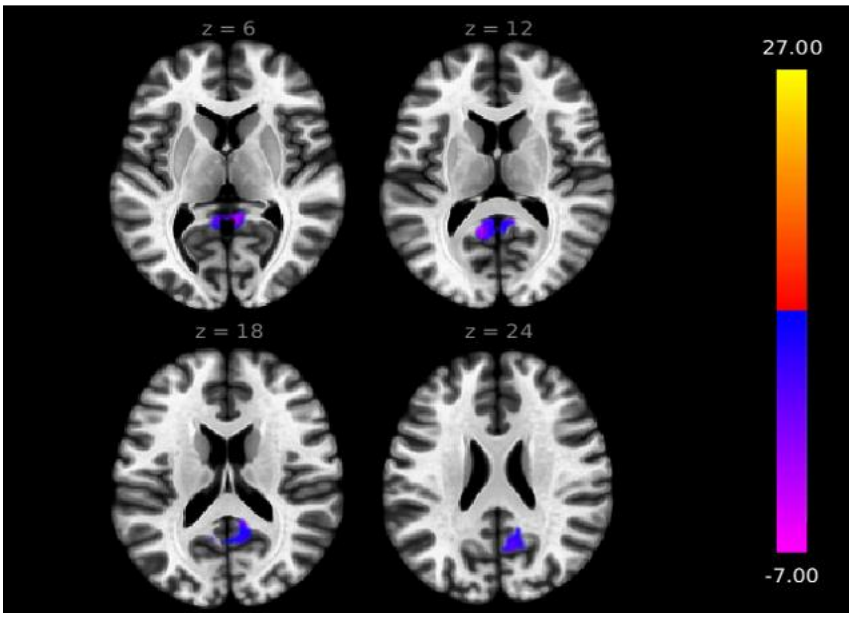

Figure 3-3. Differences in D-PE Connectivity between Healthy Controls and Psychotic Patients

A decrease in anti-correlated activity between the bilateral caudate seeds and precuneus was observed in patients relative to healthy controls (right caudate seed depicted). Results overlaid upon an MNI-152 T1 template shown as significant z-statistics. Regions of positive connectivity show greater significance from red-to-yellow and regions of negative connectivity show greater significance from blue-to-purple.. Abbreviations: D-PE – Delusion associated Prediction Error

D-PE RS Functional Connectivity Similarities between Patients and Healthy Controls

A supplemental analysis was conducted to verify connectivity was generally as expected, and to describe the connectivity that is detectable in the combined groups. This will complement results of primary analyses showing very little group difference, as results here suggest group commonalities and may contextualize group difference findings. There were several areas showing significant connectivity (Fig 3-4, Table 3-2). As expected, the lateral prefrontal nodes were positively connected with other prefrontal regions, part of the central executive network, as

well as parts of the salience network (Figs 3-4A and 3-4B). They were negatively connected (anticorrelated) with midline hubs of the default mode network (e.g., precuneus and medial prefrontal cortex)

Within the left and right caudate there was significant positive connectivity with a large subcortical cluster spanning much of the striatum including caudate, putamen, pallidum, and nucleus accumbens, and extending to adjacent subcortical structures- and additional positive connectivity with the superior frontal gyrus. Subjects showed significant negative connectivity of the caudate seeds with the regions covering precuneus and cuneal cortices (Fig3-4C). For the midbrain seed the main connectivity was positive and was to a brain region covering much of the brainstem, with coverage extending to a few adjacent structures (Fig 3-4D). The midbrain was also found to be positively connected to the medial prefrontal cortex and negative connected to the right frontal pole and right middle frontal gyrus and bilateral angular gyrus. Similar connectivity patterns for just the patient group were observed in the exploratory analysis when controlling for antipsychotic medication (Appendix B - Table 5).

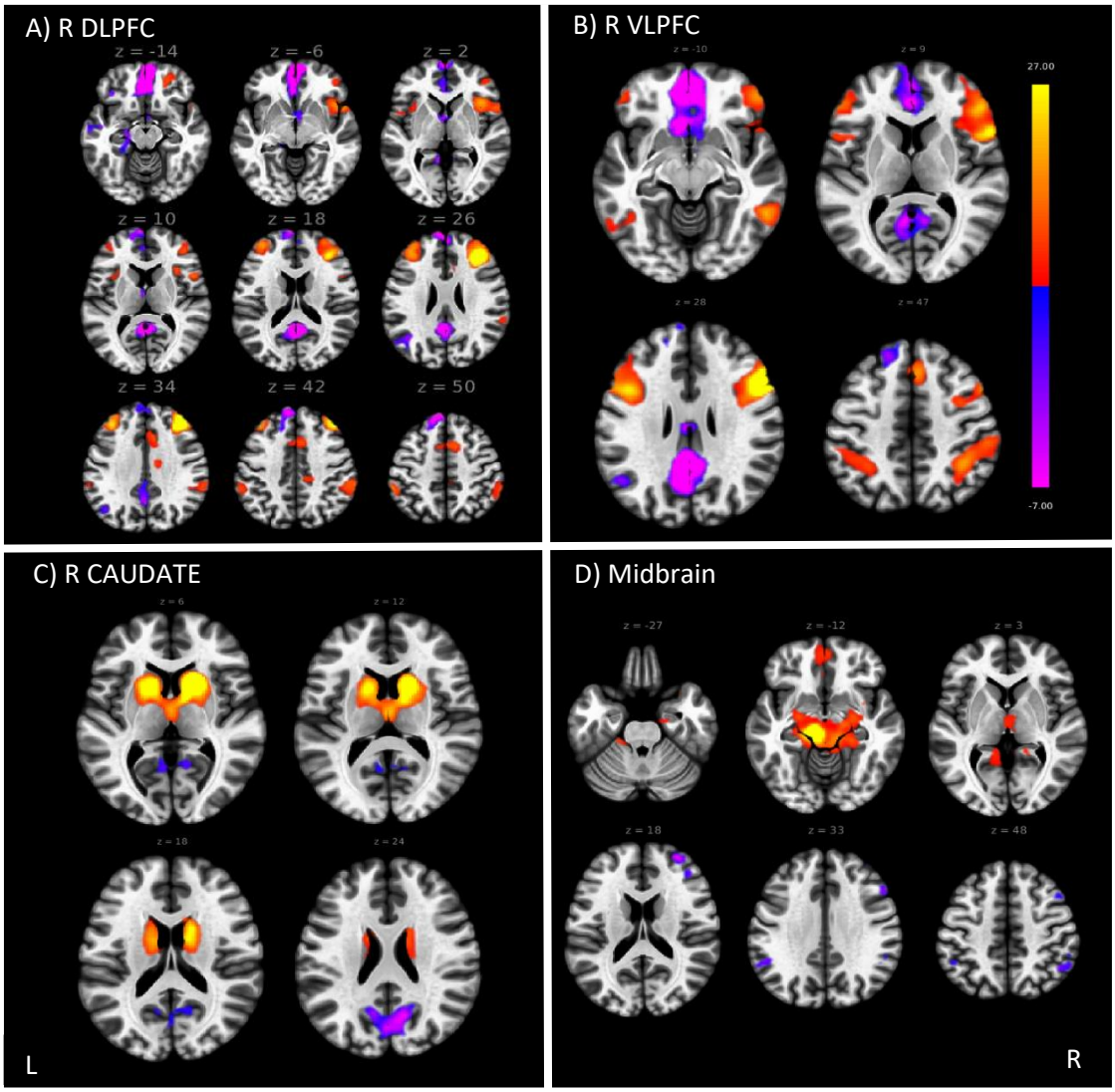

Figure 3-4. D-PE Intrinsic Connectivity in Healthy Controls and Psychotic Patients

Regions of significant connectivity healthy controls and psychotic subjects combined, shown for most of the D-PE seeds. A) Whole brain resting state connectivity of the R DLPFC. B) Whole brain resting state connectivity of the R VLPFC. Positive connectivity is depicted between R VLPFC and R DLPFC with other bilateral prefrontal regions, e.g., the central executive network A & B also depict negative connectivity (anticorrelation) of the seeds to regions within the default mode network. C) Whole brain resting state connectivity of the right caudate (left not shown but very similar results). Significant bilateral positive connectivity is depicted in striatal regions, and negative connectivity with the posterior cingulate and precuneus, nodes of the default mode network D) Whole brain resting state connectivity of the midbrain seed. Positive connectivity seen within the midbrain and medial prefrontal cortex and sparse negative connectivity to regions of the central executive network. Results overlaid upon an MNI-152 T1 template shown as significant z-statistics. Regions of positive connectivity show greater significance from red-to-yellow and regions of negative connectivity show greater significance

Figure 3-4 continued
from blue-to-purple. Abbreviations: D-PE – Delusion associated Prediction Error, R DLPFC – Right Dorsolateral Prefrontal Cortex, R VLPFC – Right Ventrolateral Prefrontal Cortex

D-PE RS Functional Connectivity Association with Delusions

Results of regressions testing whether any of the connectivity maps of the D-PE regions can predict current delusion severity were negative. Delusion severity was not a significant predictor of connectivity for any of the D-PE regions. However, a near significant finding was noted between the r VLPFC and lobule VI of the cerebellum (pFWE = .0148, Table 3-3, Fig 3-5). Next, in exploratory analyses of diagnostic group effects, there was a significant interaction for diagnostic group for the connectivity of r VLFPC and the precentral gyrus (pFWE < .0001, Table 3-3, Appendix B - Fig 2A). Post-hoc comparisons revealed greater delusion severity was associated with lower connectivity between theses brain regions for schizophrenia compared to bipolar with psychosis (pFWE < .0001). Similar results of the regression analyses were obtained when controlling for antipsychotic medication. Delusion severity was not a significant predictor of connectivity for the D-PE regions. In the exploratory analyses of diagnostic group effects after controlling for antipsychotic medication, there was a significant interaction for diagnostic group for the connectivity of r DLPFC to a large cluster centered on the midbrain (pFWE=.0001, Appendix B - Table 6, Appendix B - Fig 2B). A similar but non-significant trend was also observed in the full sample between r DLPFC and midbrain (pFWE=.0270). Post-hoc comparisons revealed greater delusion severity was associated with reduced connectivity between the r DLPFC and midbrain for both schizophrenia compared to bipolar with psychosis (pFWE<.0001) and schizoaffective compared to bipolar with psychosis (pFWE=.0049).

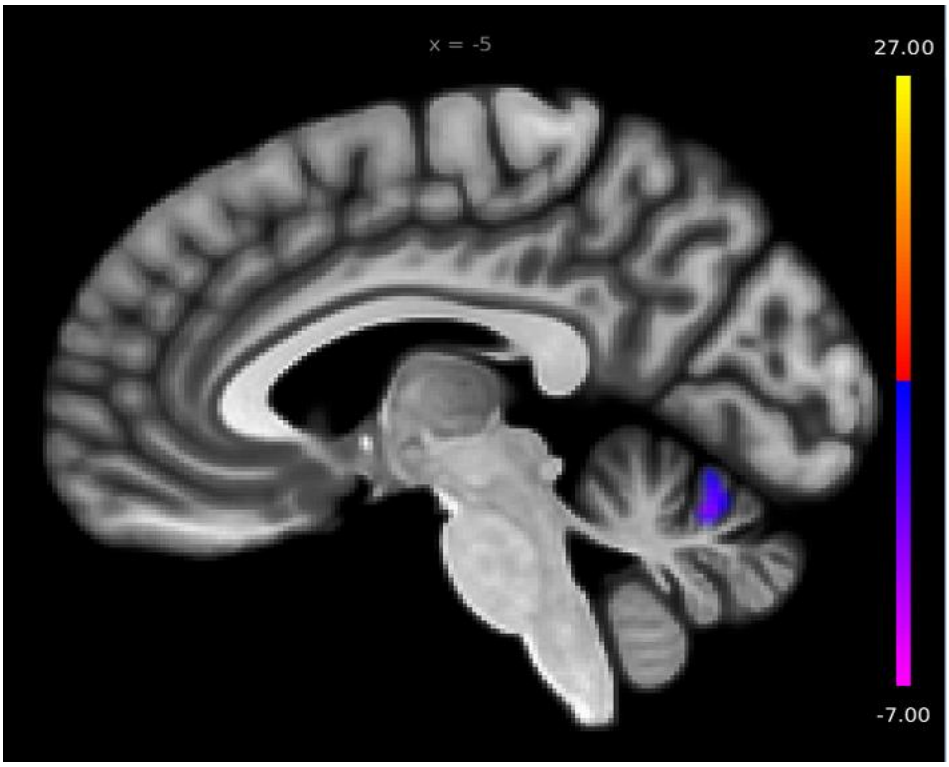

Figure 3-5. Association of with D-PE Connectivity with Delusion Severity

A transdiagnostic association of delusions was observed with resting state connectivity at a trend level of significance between the right ventrolateral prefrontal cortex and cerebellum (pFWE=0.014). Results overlaid upon an MNI-152 T1 template shown as significant z-statistics. Regions of positive connectivity show greater significance from red-to-yellow and regions of negative connectivity show greater significance from blue-to-purple. Abbreviations: D-PE – Delusion associated Prediction Error, pFWE – p-value Family Wise Error

Table 3-2. Significant D-PE Connectivity found in the Combined Group (top), and Differences (bottom) between Patients and Healthy Controls.

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE	Type of Connectivity (+/-)
Patients and Healthy Controls					
R DLPFC	4470	00, 48, -14	Frontal Pole-Medal Prefrontal Cortex – Superior Frontal Gyrus – Paracingulate Cortex – Anterior Cingulate Cortex – Subcallosal Cortex	<.000001	-
	2633	34, 34, 26	R Frontal Pole – Middle Frontal Gyrus	<.000001	+
	2154	16, 12, 66	R Superior Frontal Gyrus – Paracingulate Cortex – Anterior Cingulate Cortex – Supplementary Motor Area	<.000001	+
	1868	00, -54, 18	Precuneus – Posterior Cingulate Cortex	<.000001	-
	1479	38, 20, 02	R Insula – Inferior Frontal Gyrus – Frontal Operculum -Temporal Pole	<.000001	+
	1350	-38, 36, 34	L Frontal Pole – Middle Frontal Gyrus	<.000001	+
	1018	60, -36, 42	R Supramarginal Gyrus – Angular Gyrus	<.000001	+
	750	-40, 34, -20	L Temporal Pole – Frontal Orbital Cortex	.000003	-
	620	-20, -20, 18	L Hippocampus – Para Hippocampal cortex	.000021	-
	458	04, 02, -10	L Thalamus	.000250	-
	433	-58, -42, 50	L Supramarginal Gyrus	.000376	+
	430	-40, -62, 28	L superior Lateral Occipital	.000395	-
	390	16, -12, 36	Posterior Cingulate Cortex	.000772	+
	389	22, 46, -16	R Frontal Pole	.000786	+
	306	-32, 18, 06	L Insula – Frontal Operculum	.003417	+
	300	-58, -08, -20	L Middle Temporal Gyrus – anterior Superior Temporal Gyrus	.003818	-

Table 3-2. Significant D-PE Connectivity found in the Combined Group (top), and Differences (bottom) between Patients and Healthy Controls, Continued

R VLPFC	6572	54, 18 24	R Frontal Pole – Middle Frontal Gyrus – Interfior Frontal Gyrus – Frontal Operculum – Frontal Orbital Cortex - Insula – Precentral Gyrus	<.000001	+
	5344	-04, 48, 06	Frontal Pole – Anterior Cingulate Cortex – Paracingulate Gyrus – Subcallosal Cortex – Medial Prefrontal Cortex - Caudate - Nucleus Accumbens	<.000001	-
	4116	-04, -48, 26	Precuneus – Posterior Cingulate Cortex – Intercalcarine Cuneal Cortex – Cuneal	<.000001	-
	3029	-46, 12, 24	L Middle Frontal Gyrus – Frontal Pole – Inferior Frontal Gyrus – Precentral Gyrus – Frontal Operculum	<.000001	+
	2397	38, -46, 48	R Supramarginal Gyrus – Superior Parietal Lobe – Angular Gyrus – superior Lateral Occipital Cortex – Postcentral Gyrus	<.000001	+
	1266	-32, -42, 40	L Superior Parietal Lobe – Supramarginal Gyrus	<.000001	+
	940	56, -48, -10	R Inferior Temporal Gyrus - Middle Temporal Gyrus	.000001	+
	736	-16, 40, 36	L Frontal Pole – Superior Frontal Gyrus	.000007	-
	501	-44, -54, -14	L Inferior Temporal Gyrus - Middle Temporal Gyrus – Fusiform Cortex	.000201	+
	482	-24, -74, -50	L Cerebellum lobules 7b/2/1/8	.000267	+
	478	04, 26, 48	R Superior Frontal Gyrus– Paracingulate Gyrus	.000284	+
	369	-46, -64, 34	L superior Lateral Occipital Cortex – Angular Gyrus	001596	-

Table 3-2. Significant D-PE Connectivity found in the Combined Group (top), and Differences (bottom) between Patients and Healthy Controls, Continued

L Caudate	5736	-14, 14, 04	Putamen-Caudate-Thalamus-Pallidum-Nucleus Accumbens-Amygdala-l Hippocampus	<.000001	+
	742	-10, -84, 34	Cuneal – Precuneus Cortex	.000003	-
	482	30, -60, -34	R Cerebellum lobules Crus1/7b	.000163	+
	438	60, -30, 36	R Supramarginal Gyrus	.000332	-
	288	-08, 14, 72	Superior Frontal Gyrus	.004628	+
R Caudate	5907	16, 14, 04	Putamen-Caudate-Thalamus-Pallidum-Nucleus Accumbens-Amygdala - Frontal Orbital Cortex	< .000001	+
	2068	00, -76, 32	Precuneus - Cuneal Cortex	<.000001	-
	525	04, 16, 56	Superior Frontal Gyrus	.000107	+
	276	24, -36, -06	R Hippocampus- Para Hippocampal Cortex	.006810	-
Midbrain	6074	-10, -24, -10	Brainstem-Hippocampus-Thalamus-Amygdala-Para Hippocampal Cortex - Cerebellum	<.000001	+
	666	00, 54, -14	Medial Prefrontal Cortex	.000011	+
	469	30, 56, 18	R Frontal Pole	.000216	-
	454	52, 20, 38	R Middle Frontal Gyrus	.000275	-
	420	46, -54, 46	R Angular Gyrus - posterior Supramarginal Gyrus	.000480	-
	279	10, -42, -40	Brainstem	.005778	+
	265	-46, -58, 44	L posterior Supramarginal Gyrus - Angular Gyrus-	.007549	-
	Patients vs Healthy Controls				
L Caudate	277	-02, -52, 12	Precuneus	.005705	-
R Caudate	1330	04, -46, 02	Precuneus - Posterior Cingulate Cortex	<.000001	-

Abbreviations: BPD – Bipolar Disorder, DLPFC – Dorsolateral Prefrontal Cortex, HC – Healthy Controls, L – Left, FWE – p-value Family Wise Error, R – Right, SCZ – Schizophrenia, SAD – Schizoaffective Disorder, VLPFC – Ventrolateral Prefrontal Cortex

Table 3-3. D-PE Connectivity Associated with Delusion Severity

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE	Association (+/-)
Connectivity associated with Delusion Severity in Patients					
		Main Effect			
R VLPFC	241	-08, -68, -18	Bilateral Cerebellum (lobule VI)	.014797	+
		Diagnosis Interaction			
R VLPFC	353	-16, -20, 66	L Precentral Gyrus	.000907	+
SCZ vs BDP	563	-10, -16, 72	Precentral Gyrus – L Superior Frontal Gyrus	.000055	+
R DLPFC	292	-10, -34, -16	Midbrain – Cerebellum – Para Hippocampal Cortex	.027000	-
SCZ vs BDP	395	-08, -32, -14	Midbrain – Cerebellum – Para Hippocampal Cortex	.000640	-

Abbreviations: BPD – Bipolar Disorder, DLPFC – Dorsolateral Prefrontal Cortex, L – Left, PaHC – Para-hippocampal Cortex, FWE – p-value Family Wise Error, R – Right, SCZ – Schizophrenia, SAD – Schizoaffective Disorder, VLPFC – Ventrolateral Prefrontal Cortex

Discussion

Though analysis of the B-SNIP1 resting state existing dataset, this study aimed to investigate the intrinsic connectivity of delusion-associated prediction error regions in healthy and transdiagnostic psychosis subjects. It further addressed whether the intrinsic connectivity of these regions is associated with current delusions symptoms for the psychosis subjects. Several key observations can be made following the analyses. 1) Reduced connectivity was observed between the caudate and precuneus in psychotic subjects compared to healthy controls, providing evidence of abnormal connectivity of both striatum and default mode network in psychosis, both observed in prior work. However, no other group differences were detected, suggesting more similar connectivity between the groups than different connectivity. 2) There was no evidence that resting state connectivity of D-PE regions is significantly associated with delusion severity, contrary to hypotheses that alterations in the intrinsic connectivity of the prediction error brain

regions may be linearly associated with severity of delusions as transdiagnostic mechanism. 3) Exploratory analysis provided suggestive evidence in support of diagnostic heterogeneity, as delusion severity was found to be associated with resting connectivity changes in the schizophrenia patients.

Intrinsic Connectivity of Striatum Nodes of D-PE Network Disrupted in Psychosis Patients Relative to Healthy Subjects

There was essentially one difference in intrinsic connectivity of D-PE regions between patients and healthy controls: bilateral connectivity of the caudate with the precuneus was altered in patients. Specifically, the anticorrelation between the activity of the precuneus and caudate seen in healthy subjects was not observed in the psychosis subjects. This finding is consistent with studies that have found default mode resting state network alterations in schizophrenia. Less work has been conducted that evaluates the relationships among major resting state networks in schizophrenia, which would be pertinent to the present observation. One very study reported that schizophrenia patients show reduced time in which their brains show strong anticorrelation of default mode and task-positive networks, including basal ganglia (Weber et al. 2020). This is consistent with the present observation of lack of anticorrelation of caudate (of the basal ganglia) to precuneus (a default mode network node). The dysconnectivity of caudate to precuneus is entirely consistent with notions of dysconnectivity between networks potentially underlying psychosis, where an imbalance in task-positive and restful/introspective functions associated with default mode network could lead to aberrant experiences of reality. In terms of affecting prediction error processes, a weak anti-coupling of default mode and caudate could be a clue regarding how the striatum may fail to differentiate internally generated stimuli, as its function remains somehow indistinct from that of precuneus, or not modulated properly by it (Northoff,

2014; Bolton et al., 2020). However, the importance of the default mode network to prediction error brain region function should not be overstated, as there were no other connectivity alterations to it, although other seeds had significant connectivity to DMN nodes in the descriptive view of connectivity of the PFC and midbrain seeds.

A prior study identified aberrant interactions between the central executive network - which encompasses the r VLPFC and r DLPFC – and the default mode network in schizophrenia (Manoliu et al, 2014) but significant differences were not found in our transdiagnostic cohort. Counter to expectation, abnormal frontostriatal resting state connectivity did not distinguish patients from healthy controls. One possibility is that these changes underlie the deficits of these in prediction tasks are not jointly coupled or salient in resting state activity.

To supplement the aim of comparing whole brain intrinsic connectivity patterns of the D-PE regions, a descriptive analysis of the connectivity observable in the entire sample was conducted to assess whether it followed generally predicted patterns. This also served as a data quality check and offered clues that the groups had a great deal of similarity of connectivity patterns and strengths, which was suggested by the group comparison result of essentially one group difference in connectivity. Overall, the connectivity observed was well in line with expectations based on extensive studies of normal resting state connectivity for these regions. In the combined group, resting state activity in R VLPFC and R DLPFC nodes of the D-PE network were positively correlated with each other and to areas more broadly covering the lateral and middle prefrontal cortex. This corroborates the predictions of similar influence, or similar role, of DLPFC and VLPFC in the prediction error models being tested. In addition, for both patients and controls, the resting state activity in the two prefrontal regions were anti-correlated with that of the medial prefrontal cortex and precuneus – two major hubs of the default mode network

(Raichle, 2015). This is consistent with numerous findings of anti-correlated activity between "task positive" central executive network and the "task negative" default mode network brain areas (Fox et al., 2005; Greicius et al., 2003), including during resting state.

Similarly, the caudate seeds both were observed to be anticorrelated with the precuneus in the total-sample analysis, another instance of such relationships of default mode nodes to nearly any other "task positive" network, which would include basal ganglia networks. For example, the finding of a negative relationship between the caudate seeds and the extended precuneus area is consistent with reports from Di Martino et al. (2008) who observed in a sample 35 healthy participants that the dorsal caudate has negative correlations with precuneus, posterior cingulate, and occipital cortices. On the other hand, caudate seeds were positively connected with the entire set of basal ganglia nuclei - the contralateral caudate, and bilateral putamen, pallidum, and nucleus accumbens. This tight connectivity of caudate to other basal ganglia structures is well in line with known networks reliably detectable in resting state data, and consistent with our understanding of the interconnections of basal ganglia nuclei. It is perhaps notable that there was no significant connectivity between caudate and the PFC seeds. Di Martino et al., (2008) did report increased positive connectivity of the dorsal caudate with the control and attention areas including the DLPFC and VLPFC. However, connectivity of basal ganglia, including caudate, to prefrontal targets in resting state data is not a connection found using typical network parsing of resting state data, such as with independent components analysis (data-driven analyses that parse resting state data into networks of brain areas showing strong covariance over time). This suggests frontal-striatal resting state connectivity is not among the most clearly connected regions, consistent with our findings. The relationship may be complex, however, potentially explaining why some seed-based studies have reported the

connectivity such as Di Martino et al. (2008) and that is why it is not usually detected in resting state: in a finely parsed mapping of the precuneus, Zhang and Chiang-shan (2012) reported a complex relationship with the basal ganglia, observing similar anti-correlated activity between the entire precuneus and dorsal caudate, but positive connectivity between ventral precuneus areas the ventral caudate.

The midbrain seed was positively connected to the brainstem and medial prefrontal cortex in both patients and controls. This is consistent with the known structural links of the midbrain with the basal ganglia and prefrontal cortex (Verger et al. 2020). The finding also corroborates results from other recent studies investigating the resting state connectivity of midbrain nuclei (Hadley et al., 2014, Murty et al., 2014, Tomasi and Volkow, 2014, Zhang et al., 2015; Bär et al., 2016). However, the midbrain was not found to be functionally connected with the striatum at rest which would be predicted by the mesolimbic and nigrostriatal pathway and has been reported previously (Hadley et al., 2014; Tomasi and Volkow, 2014, Zhang et al., 2015). The divergent finding may be partially accounted for by reduced specificity in discriminating amongst midbrain nuclei in the current analysis. This is supported by Murty et al. (2014) report of increased resting connectivity of the ventral tegmental area to the nucleus accumbens relative to the substantia nigra. Overall, no significant difference was observed in the midbrain intrinsic connectivity between the patients and healthy controls. This contrasts with observations by Hadley et al. (2014), who within 21 unmedicated patients compared to 21 healthy controls reported reduced connectivity of the midbrain with a number of regions including the precuneus, anterior cingulate cortex, and basal ganglia. They also showed evidence that after one week of antipsychotic treatment midbrain connectivity with the thalamus normalized. Although their results were based on a fairly small sample, it suggests that potential

midbrain abnormalities within the current study's chronic and stably medicated patients may also be normalized as a result of antipsychotic medication.

Diagnostic Heterogeneity observed association of D-PE RS Functional Connectivity with Delusions

Within the full psychosis sample, connectivity of the D-PE regions was not found to be associated with delusion severity and this remained true when controlling for anti-psychotic medication. This runs counter to the predictions that abnormal intrinsic resting state of D-PE regions may be a transdiagnostic mechanism that underlies psychotic delusions. The study's large sample reduced the risk of likely alternative explanations such as inadequate power. Subtle variations within prediction error network though may not be captured by the stringent whole brain analysis, nor would non-linear relationships have been detected. This was the first study to attempt to identify linear relationships of connectivity of these brain regions previously shown to have correspondence between their task-evoked activation and delusion severity (Corlett et al. 2007; Romaniuk et al., 2010; Corlettt and Fletcher). Hence, the results suggest the abnormalities reported previously are specific to context, e.g., prediction error in action, rather than also being related to context-independent connectivity alterations, as would have been depicted by the correlational approach of the present analysis.

Although not significant after multiple comparisons corrections, there was a trend relationship between increased delusions and decreased connectivity between the r VLPFC and lobule VI of the cerebellum. This portion of the cerebellum has been found previously to be structurally and functionally linked to the prefrontal cortex and to be active in non-motor cognitive tasks such as language, spatial tasks, executive function and affective processing

(Stoodley & Schmahmann, 2010). Indeed, cerebellar regions are part of cortical-subcortical-cerebellar loops, for which research exists linking abnormality in these extended loops to schizophrenia (Andreasen et al., 2008). Hence, this finding suggests there may be a modest relationship between reduced connectivity of the lateral prefrontal cortex and cerebellum nodes of these loops, which might be expected to be found given this prior literature suggesting such abnormality is an illness-related rather than symptom related, correlate. The fact that this relationship was not found when controlling for anti-psychotic medication further supports this interpretation, given the association between dose and symptom severity such that covarying for antipsychotic dose may remove illness-related variance. However, as it was a trend-level finding in an already robust sample, future studies are needed to replicate this possible association.

An exploratory analysis of an interaction between D-PE connectivity and diagnosis suggest delusions maybe be associated with specific intrinsic connectivity abnormalities within specific disorders. This tests an alternate possibility to our assumption of transdiagnostic similarity, and results suggest this alternative view may be worth considering. Reduced connectivity between the midbrain and the right dorsolateral prefrontal cortex was shown in both schizophrenia and schizoaffective patients to be associated with increased delusion symptoms. This whole brain connectivity finding is consistent with the prediction error tasks studies, in particular with the report by Corlett et al. (2012) that activation in both the dorsolateral prefrontal cortex and midbrain were associated with distress related to delusional beliefs. As the original prediction error task studies did not incorporate bipolar subjects, it is possible a similar association would not be in these subjects, which would be consistent with bipolar patients not showing a similar association of delusions with resting state alterations of D-PE network. This result may suggest distinct circuitry may underlie the delusional experiences which occur in

bipolar patients experiencing psychosis outside of those assessed in the present study. As there are some variations in presentation of delusion experiences, such as higher frequency of grandiose themed delusions in bipolar illness than schizophrenia, different neural system abnormalities might be present (although there are more phenomenological similarities between bipolar and schizophrenia delusions than differences).Overall, however, interpretation of this result is tempered by the smaller representation of bipolar subjects at the more extreme end of the delusion severity in the sample.

A second connectivity association was observed for schizophrenia: reduced connectivity of the right ventrolateral prefrontal cortex with the precentral gyrus correlated with increased delusions. Direct association of frontomotor connectivity with delusions severity is a novel observation to our knowledge, although aberrant motor connectivity has been reported numerous times more generally for schizophrenia (Shinn et al. 2015, Bernard et al. 2017, Du et al., 2019). It has been suggested that failures in motor cortex integration may underlie experiences of alien control and passivity delusions in schizophrenia (Schnell et al., 2008; Corlett et al., 2010). This is supported by evidence of sensorimotor deficits and reduced capacity to identify self-generated actions in schizophrenia (Frith et al., 2000; Shergill et al., 2005). It is possible diagnostic differences in results for the study here may be due to the increased prevalence of these types of delusions in schizophrenia (Junginger and Coe, 1992; Appelbaum et al., 1999).

Moreover, heterogeneity in delusions symptoms, beyond potential diagnostic differences, may also account for the absence of a significant transdiagnostic association with resting state connectivity. The limited studies examining neurocorrelates of delusional subtypes (e.g. paranoid, grandiose, control delusions) suggest distinct neural circuits are engaged (Gallagher and Frith, 2003; Blackwood et al., 2004, Kimhy et al. 2005). It may be that

abnormalities in the prediction error circuit may underlie a distinct dimension of the delusional experience. Overall, the data suggest that changes in the resting state functional connectivity of the defined prediction error circuit are not universally related to delusions symptoms across psychotic disorders, though results of the supplementary diagnostic analyses suggest that specific sub-disorder mechanisms may exist. Further research will be useful to investigate how symptom and disorder heterogeneity potentially affect the relationship with connectivity in prediction error and related brain circuits.

4.3 Limitations

There are several caveats to acknowledge in this study. Firstly, the measurement of delusion symptoms was via the single item of the PANSS interview, which assesses delusion severity within the last week. An additional limitation was that PANSS assessments were not systematically conducted on the day of the scan, as data in the original B-SNIP1 was collected over many visits, which for some cases were separated by a few days up to a couple weeks. Thus, severity is only approximately related to the scan day. These two factors may have reduced the ability of the analysis to precisely capture acute delusional state associated variance related to intrinsic connectivity. However, significant changes in delusion symptoms during study participation are not expected as all recruited participants were medically stabilized with no recent history of drug abuse. A separate limitation was that the selection of ROIs for connectivity analysis was based on peak voxels of prior literature. This enabled a focus on regions directly related to delusions in prior fMRI studies. However, a more precise localization of prediction error associated regions for each subject may improve precision in examining the resting state connectivity of these regions in the future and the relationship with symptoms. Finally, although no functional imaging study, to our knowledge, has directly addressed delusions in bipolar

disorder previously, the findings presented here should be considered tentative given the smaller number with greater active delusion symptoms.

Conclusions

This study provided a novel test of the association of current delusion severity to prediction error system connectivity using the resting state fMRI data of B-SNIP. Resting state connectivity was not found to be a transdiagnostic predictor of delusions. However, exploratory analyses suggest diagnostic specific associations particularly between the r DLPFC and midbrain in schizophrenia subjects. Additionally, resting state connectivity of the prediction error network was tested in order to characterize the normal intrinsic dynamics and elucidate an abnormal connectivity of prediction error system in psychosis irrespective of symptoms. The prediction error network in patients and healthy controls similarly engaged task positive and task negative resting state networks, but differences were observed between groups in the connectivity of the striatum with the precuneus.

Chapter IV: Prediction Error Related Effective Connectivity in Psychosis and its Association with Delusion Symptoms

Introduction

Psychosis is a serious brain disorder that effects an estimated 1-3% of the population (Perälä, et al., 2007; van Os et al., 2009). Delusions, false beliefs that are firmly held despite contrary evidence, are a major clinical manifestation of psychotic disorders. The prevalence of delusions across psychotic disorders and their partial remediation by pharmacological antipsychotic therapy suggests that their experience may be mediated through common means of neural circuit dysfunction (Pini et al., 2003, Howes et al., 2012; Picardi et al., 2018, Cheng et al., 2020). However, further advances in the treatment of delusions is severely limited by poor insight into the causal neural mechanisms underlying them. Disruption in normal predictive coding processes and aberrant salience have been put forth as a theoretical framework for the emergence of delusions (Kapur, 2003; Fletcher and Frith, 2009; Corlett et al., 2010; Adams et al., 2014; Sterzer et al., 2018). Specifically, abnormal prediction error response, neural activity associated with the difference between prior expectations and observed outcome, is thought to underlie misattribution of beliefs in delusions. However, greater empirical evidence of biological pathways linking prediction error processes with delusions is needed to validate these theories along with research characterizing the exact nature of the neurophysiological changes occurring in psychotic delusions.

Prediction error response is intricately linked to the function of basal ganglia – corticothalamic loops. The basal ganglia incorporate a number of subcortical structures which include the striatum (comprised of the caudate and putamen), limbic, and midbrain regions. Several parallel pathways through the basal ganglia are known to exist, and culminative evidence in recent decades has identified associative and limbic basal ganglia – corticothalamic loops as critical for reinforcement learning (Fig 4-1, Alexander et al., 1986; Haber and Knutson, 2010). Numerous studies have demonstrated that dopaminergic neurons in the midbrain and striatum play a central role in mediating salience related processes, and activity in these regions has been associated with various domains of salience including response to reward, aversive, and novel stimuli (Schultz, 2000; Zink et al., 2003; Wise, 2004; Bromberg-Martin et al., 2010). In addition, electrophysiological studies of laboratory animals undergoing reinforcement learning paradigms have identified a specific prediction error associated response in the midbrain and striatum that corresponds to phasic firing of the dopaminergic neurons during expectation violation conditions (Schultz, 2000, Waelti et al., 2001, Bayer and Glimcher, 2005). Parallel observations of prediction error related BOLD activation have been reported in both reward and non-reward based associative learning tasks administered to healthy subjects (O'Doherty et al., 2003a; McClure et al., 2003; O'Doherty et al., 2006).

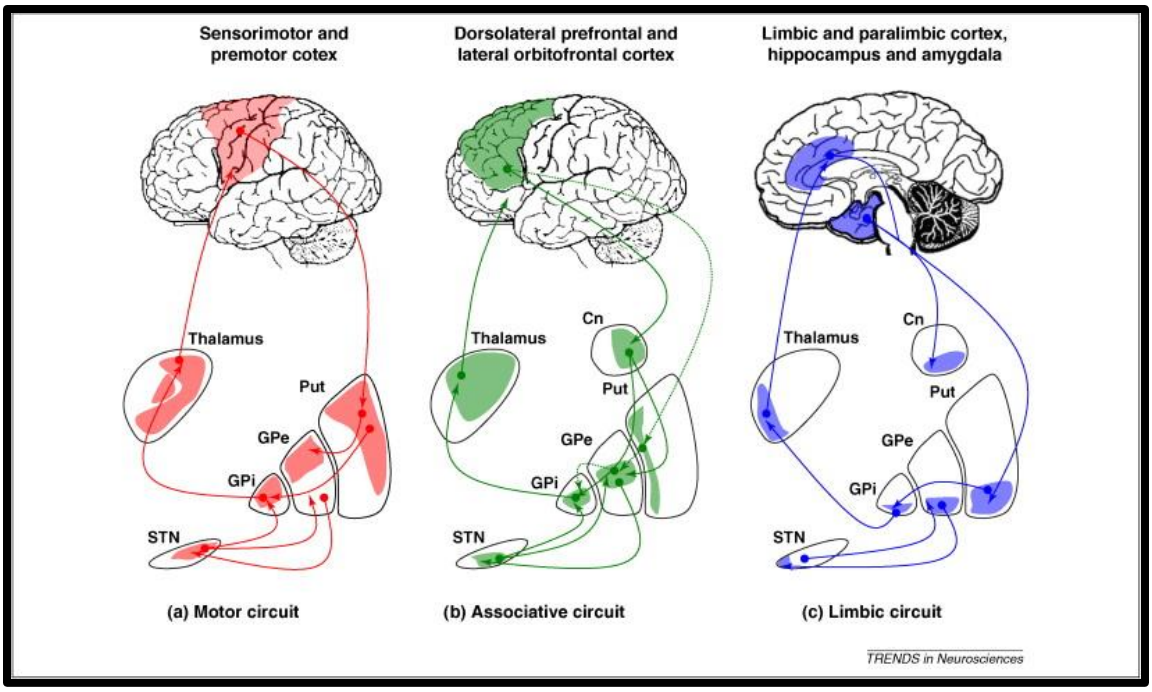

Figure 4-1. Basal Ganglia – Corticothalamic Loops

Three parallel basal ganglia circuits have been commonly described (a) Motor, (B) Associative, (C) Limibic Loops.
Image Credit - Paul Krack, Marwan I. Hariz, Christelle Baunez, Jorge Guridi, Jose A.Obeso Deep brain stimulation: from neurology to psychiatry? Trends in Neuroscience. Volume 33, Issue 10, October 2010, Pages 474-484

The cortical portions of the loops for prediction error regulation have been identified as located in prefrontal cortex. This is supported by a number of neuroimaging studies which report prediction error neurocorrelates in dorsolateral prefrontal cortex and ventrolateral prefrontal cortex (Fletcher et al., 2001; Corlett et al., 2004; Turner et al., 2004; Garrison et al., 2013). However, the exact role of the lateral prefrontal cortex in prediction error during association learning remains unclear. It is less directly related with rewards in comparison with more midline cortical structures like the medial prefrontal cortex; however, there is evidence suggest that the ventrolateral and prefrontal cortex have largely overlapping roles in mediating working memory and action selection (Haber and Knutson, 2010, Sakurai et al., 2015). Top down cortical inputs to

the basal ganglia in sensorimotor loops during prediction errors suggest that their role may be to help mediate action selection and movement, and it is hypothesized the lateral prefrontal cortex may plays a similar role within the associative basal ganglia circuits during prediction error in reinforcement learning (den Ouden et a., 2012).

Specific but limited evidence has been reported linking delusions in individuals with psychotic disorders to abnormal prediction error (PE) activity of the prefrontal cortex, striatum, and midbrain. Corlett et al. (2007) reported decreased PE activation in the right prefrontal cortex as being associated with delusional severity in a small psychosis sample. In a similar investigation of delusion spectrum beliefs within healthy subjects, bilateral caudate activity was found to be negatively associated with odd beliefs, while decreased PE activation in the prefrontal cortex, striatum, and midbrain were all associated with distress related with odd beliefs (Corlettt and Fletcher). However, a study by a different group reported increased midbrain activation was found to be associated with delusions in psychosis subjects (Romaniuk, 2010), yielding a mixed literature regarding greater or lower PE related activity associating with delusions. Some recent psychosis studies have identified prediction error abnormalities in the patient groups, but failed to find positive symptom associations, which delusion severity would contribute to (Ermakova et al., 2018; Katthegan et al., 2018). Of course, these latter negative findings with positive symptoms may be due to lower sensitivity for delusion severity associations due to the combined positive symptom measure used. Finally, our replication efforts (chapter 2) also failed to corroborate prior reports of delusion association to PE related brain activity.

One source of inconsistency in the observations of PE activity and delusions may be related to ignoring aspects of neural function, referred to as neural dynamics, influencing the PE-

related systems targeted in those studies. Further any PE-related activation identified as altered in association with symptoms falls short of offering a clear mechanistic view of the neural abnormalities given the limited description of them as essentially overly active or too inactive. This is a limited picture of the realities of brain system function, where responses during a cognitive operation like PE are generated and integrated with ongoing dynamics of brain systems, processes that cannot be discovered by activation magnitude-based methods. Greater understanding of both the context-dependent (task-evoked) and context-independent dynamics of the prediction error regions is needed to move toward establishing a more thorough, as well as causal, relationship of neural system function yielding psychotic delusions.

One useful way forward is investigating within both psychotic and healthy individuals the causal neural dynamics involved in prediction error, and whether such dynamics associated with delusions. Effective connectivity analysis provides an advantageous way to study causal dynamics of intrinsic connectivity, and so can be applied to study the prediction error system to address the questions posed. As neuronal nodes can have direct and indirect influences on each other within a network, effective connectivity metrics, versus more ubiquitous functional connectivity metrics (such as those conducted in chapter 3), seek to model the directed and causal connections of a neural network (Friston, 2011). Spectral Dynamic Causal Modeling (spDCM), a method specifically developed to test the effective connectivity within task-independent resting state neuroimaging data, can be used to investigate the directed connections of the prediction network using subjects' endogenous activity (Friston et al., 2003*).*

Consideration of the hypotheses regarding how neural system alterations may lead to delusions is needed to determine what effective connectivity models should be tested. As has been reviewed in prior chapters, false beliefs are thought to arise as a result of increased salience

given to either externally or internally generated stimuli, which may then lead to misattribution of causal relationships (Kapur, 2003). This process could be biologically induced by distinct disruptions in the neural systems that support normal salience labeling processes. For example, impairments in top down control of sub-cortical systems that mediate motivational salience during prediction error could lead to reduced fidelity of appropriate salience detection (Adams et al., 2013). Alternatively, an alteration could be tonic hyperactivation of the striatum, which may result in increased bottom up signaling that drives a hyper associative state in delusional patients (Howes and Nour, 2016; Sorg et al., 2013). Furthermore, the dynamic feedback within these systems may lead to combined effects that result in delusions (Broyd et al., 2017). Analyses addressing the causal dynamics of the prediction error system may most properly differentiate between the directions of neural system regulation suggested by these models. This may provide clearer understanding, not to mention testable models, of how delusions emerge and/or are sustained in psychosis.

In this study, it is hypothesized that alteration in the intrinsic connectivity of the prediction error circuit will be associated with delusion severity, as there is evidence that salience and cognitive control networks that support normal learning and reward processing are also abnormal at rest in individuals with psychosis (Sarpal et al., 2015; Raij et al., 2018; Karcher et al., 2019; Zhang et al., 2019). The aberrant salience theory of delusions suggests delusional beliefs may arise in psychosis in response to misdirected salience given not only irrelevant stimuli (being improperly tagged as salient) but also internally generated signals that, under normal neurocognitive operations, would be inhibited. This set of processes could be neuronally mediated by a reduction in top-down inhibition from the lateral prefrontal cortex to the basal ganglia, resulting in adequate suppression during salience detection. Furthermore, as several

studies have also reported abnormal perceptual and attribution biases (Maher, 2006; Uhlhaas et al., 2007; Dudley et al., 2016, McLean et al., 2017) in relation to delusion symptoms in psychosis, it is reasonable to speculate that alteration in endogenous activity may be precursor to abnormal salience underlying delusions.

Insight into context-independent (e.g., intrinsic connectivity), dynamics of the prediction error system is additionally of interest as prior research has failed to provide biological evidence accounting for what often appears to be the spontaneous emergence of delusional beliefs. While it has been hypothesized that delusions may arise as an epiphenomenon of impaired salience and self-attribution mechanisms during periods of mind wandering and periods of self-reflective cognition, there has been little empirical research to substantiate this theory (Kean, 2009; Shin et al., 2015; Iglesias-Parro et al. 2020). Increased understanding of the intrinsic resting state dynamics of the prediction error system through effective connectivity analyses can better inform how the network functions is disrupted in psychosis and specifically address if changes in these pathways at rest increase vulnerability for greater delusion symptoms.

Study Aims

The current study takes advantage of recently developed neuroimaging analysis methods to test how various models of neural circuitry dynamics within prediction error brain regions are associated with delusions in psychotic disorders. Within a sample of transdiagnostic psychotic patients and healthy controls, an effective connectivity analysis of resting state fMRI data is conducted to address the following questions: (1) Are the endogenous connectivity dynamics between prediction error regions altered in psychosis patients, and (2) are any such effective connectivity changes in psychosis associated with severity of acute delusion symptoms? It is

predicted that 1) psychosis patients will show reduced top down inhibition of prefrontal cortex to the striatum and midbrain, and 2) this will be associated with increased delusion symptoms.

Methods

Participants

Data for resting state analyses was obtained from the completed Bipolar & Schizophrenia Network for Intermediate Phenotypes 1 (B-SNIP 1) multisite study. Subjects were recruited following IRB approval from each of the five study sites, and the larger study has been described elsewhere in detail (Tamminga et al., 2014). Psychosis patients were clinically characterized, and all subjects had a panel of biomarkers assessed including the neuroimaging reported here. The psychosis patients and healthy volunteers were recruited using local advertising. Inclusion and exclusion criteria were age 18-60, able to provide written informed consent, estimated IQ > 60, no current substance abuse disorders or major neurological/cognitive/cerebrovascular-affecting disorders, and no significant head trauma history. Healthy controls had no personal history of any psychiatric disorder or first-degree relative with schizophrenia, schizoaffective disorder, or mood disorder.

Clinical Assessments

Trained clinical raters confirmed DSM-IV diagnosis of Schizophrenia, Schizoaffective, or Bipolar Disorder with Psychosis using the SCID-IV (First al., 2002a; First et al., 2002b). Raters also administered the Positive and Negative Symptom Scale (PANSS) which assesses the severity of a range of psychotic symptoms in the last week (Kay et al., 1987). Current delusional severity in patients was characterized using the “Delusions” PANSS item (score range 1-7).

Imaging Data Acquisition and Preprocessing

Subjects underwent a 5-min rs-fMRI scan in 3T scanners with closely aligned acquisition parameters (Appendix B - Table 1). Subjects were instructed to remain still, stay awake and keep their eyes focused on a crosshair for the scan's duration. Wakefulness was confirmed with the subjects following the scan. To allow for scanner stabilization, the initial 6 images were discarded. Using the SPM based CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012), the time series was aligned, slice-time corrected, normalized to MNI space (Montreal Neurological Institute), and smoothed with a 8mm FWHM Gaussian kernel, with 2x2x2 mm resampled voxel size. The data was denoised using regression of white matter and CSF (aCompCor), scrubbing signal, motion + 1^{st} order derivatives, and a linear and 2^{nd} order polynomial drift term, and subsequently band-pass filtered at 0.008-0.1 Hz based on recent reports on the effect of filtering on resting state (Goto et al., 2015). To additionally ensure scan quality, subjects with visually identified artifacts and framewise motion > 3mm were excluded from analyses.

Delusion associated Prediction Error (D-PE) Regions of Interest

Five D-PE regions of interest (ROI) were based on neural coordinates from published task-based fMRI studies (Corlett et al., 2007, Romaniuk et al., 2010; Corlettt and Fletcher). The MNI space coordinates were (x=34, y=34, z=26) for the right Dorsolateral Prefrontal Cortex [rDLPFC]; (x=54, y=18, z=24) for the right Ventrolateral Prefrontal Cortex [rVLPFC]; (x=15, y=15, z=4) for the right Caudate; (x=-15, y=15, z=4) for the left Caudate; and (x=-11, y=-23, z=-9) for the Midbrain. Masks for each seed were generated by creating a 10mm radius sphere centered on the MNI coordinate (Fig 4-2) and then excluding from the sphere any voxels classified as CSF or white matter, using individual subject brain segmentation. For each subject,

the first eigenvariate was extracted from the denoised rs-fMRI timeseries of each D-PE region to use for effective connectivity analysis.

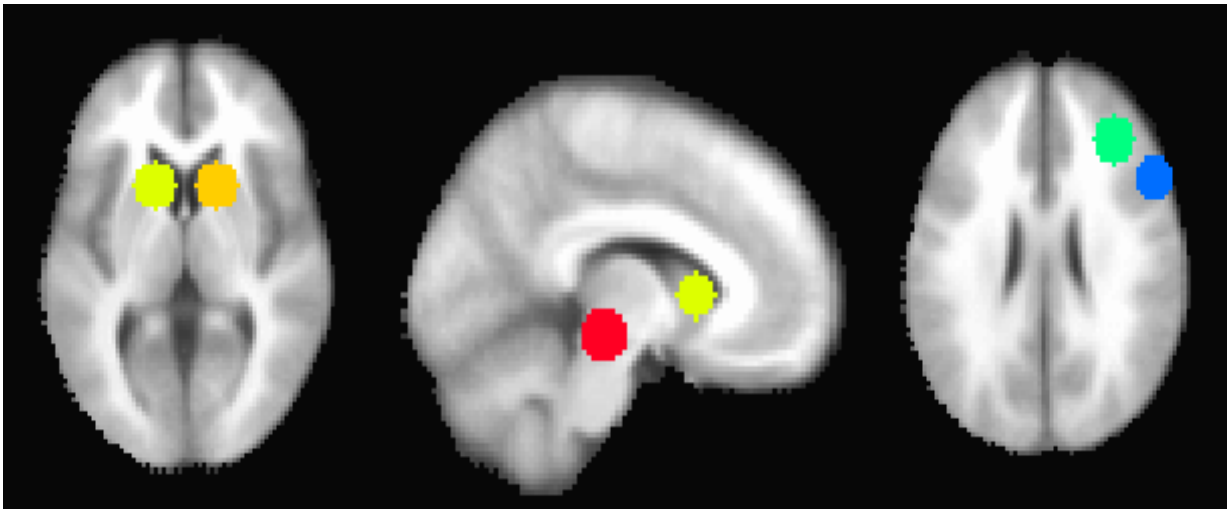

Figure 4-2. Regions of Interest for D-PE Effective Connectivity Analysis

Spherical ROIs based on peak neurocoordinates indentified from task based prediction error association with delusion symtpoms. Red = midbrain. Blue = ventrolateral prefrontal cortex. Green = dorsolateral prefrontal cortex. Yellow and Orange: left and right caudate.

Spectral Dynamic Causal Modeling of D-PE Resting State Seeds

Zeidman et al. (2019a) summarizes the process of conducting dynamic causal modeling (DCM) as follows: "A DCM…can be conceptualized as a procedure that generates neuroimaging timeseries from the underlying causes (e.g., neural fluctuations and connection strengths). The generated timeseries depend on the model's *parameters*, which generally have some useful interpretation; for example, a parameter may represent the strength of a particular neural connection. Having specified a …model, one can then simulate data under different models (e.g. with different connectivity architectures), and ask which simulation best characterizes the

observed data." The specified model for this study was a fully connected model of the five D-PE ROIs, meaning excitatory or inhibitory influence of each region on each other region is modeled, and the model also includes self-inhibition for each region, reflecting inhibitory interneurons within brain regions that downregulate the region's activity once activated. The best fit of these biologically-informed (e.g., inclusion of the hemodynamic response functions, noise from scanning, etc.) generated models of connectivity to the data is then determined. For the analysis a specific DCM variation is used called spectral dynamic causal modeling (spDCM). This approach models the rs-fMRI data in the frequency domain rather than time domain. This has been argued to be more appropriate for assessing connectivity of the low amplitude fluctuations within resting state data that are of interest for connectivity and for detecting group differences of connectivity (Friston et al., 2014; Razi et al., 2015).

For the fully connected D-PE network (Fig 4-3), spDCM was implemented with DCM12 (revision 6801) in SPM12 (revision 6778) following recommended settings per Zeidman et al.(2019a) for subject-level processing. Activity within each brain region is modeled as a single state (either excitatory or inhibitory, not both simultaneously, which would be a two-state model), and only linear modulatory effects (e.g., connectivities) were permitted between each region. Parameters for the fully connected model were fit using the Parametric Empirical Bayes (PEB) framework, where the probability distributions of each effective connectivity parameter for individual subjects were then passed to the group level following procedures and SPM12 settings recommended by Zeidman et al., 2019b. Free energy – a measure approximating the model evidence accumulated through the computations to assess the fit of connectivity models– seeks to find the parameter values that most accurately explain the brain data while minimizing model complexity (Friston et al. 2007). Significant effective connectivity was identified as

parameters with free energy evidence greater than 95% posterior probability. The association of effective connectivity with patient vs healthy group differences, and then with current delusion severity in patients, was modeled as a general linear model for each between-subject effect (e.g., the patient-control difference, then the delusion severity association). For each of these general linear models, effects of no interest were included: antipsychotic medication dosage, sex, age and scanner site. To examine these group level effects, the connections (i.e. effective connectivity parameters) were reduced and comparison of the evidence for each reduced GLM model was conducted using the Bayesian Model Reduction (BMR) procedure. BMR iteratively eliminates parameters that reduce model evidence (Friston and Penny, 2011; Rose et al., 2012; Fristen et al., 2016; Zeidman et al., 2019b). For this study, a summary of the best model was identified by calculating the weighted average of the parameters based on their model evidence using Bayesian Model Averaging (BMA) (Zeidman et al., 2019b). Ultimately, this method used the BMR automatic greedy search to comparethe evidence of reduced models and then averaged over the final 256 best models using PEB-BMA to test the association of any possible effective connectivity effect in the D-PE network - either with patient healthy control difference (hypothesis 1) or with delusion symptoms (hypothesis 2).

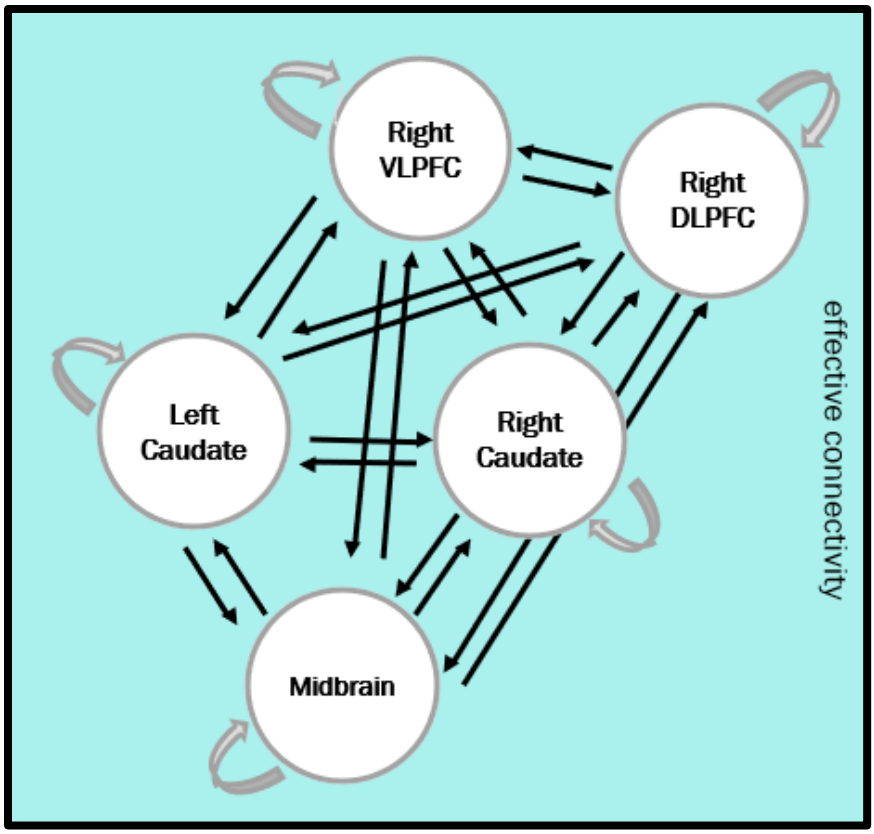

Figure 4-3. Fully Connected D-PE Network Entered Into Spectral DCM

The initial fully connected model was specified including either excitatory or inhibitory influence between each region (black arrows), and self-inhibition modulation for each region (gray curved arrows).

Results

Sample Characteristics

In total, 324 psychosis patients (237 with antipsychotic usage information) and 182 healthy controls were included, while 53 patients and 31 healthy controls were excluded due either to excessive motion or image artifacts (Appendix C-Table 1). Clinical and demographic characteristics for included subjects are available in Table 4-1 and Appendix C - Table 2. The sample spanned a range of delusion severity, though no subject had the highest severity rating,

consistent with the more stable community sample targeted for recruitment (Fig 4-4, Appendix C-Fig 1).

Table 4-1. Demographics and Clinical Characteristics for Included Subjects

	Patients	Healthy Controls	P-value
N	324	182	
Male/Female	155/169	70/112	0.042
Avg. Age (years)	35.8 (12.1)	37.8 (12.3)	0.076
Avg. Daily CPZ	389.1 (371.8)		
PANSS Delusion (P1)	2.7 (1.4)		
PANSS Positive	15.9 (5.4)		
PANSS Negative	14.7 (5.1)		
PANSS General	31.9 (8.9)		
PANSS Total	60.7 (16.5)		
mFDpower (mm)	0.21 (0.13)	0.17 (0.10)	<.001

Abbreviations: CPZ – Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, PANSS -Positive and Negative Syndrome Scale**.** Values in parenthesis are standard deviations.

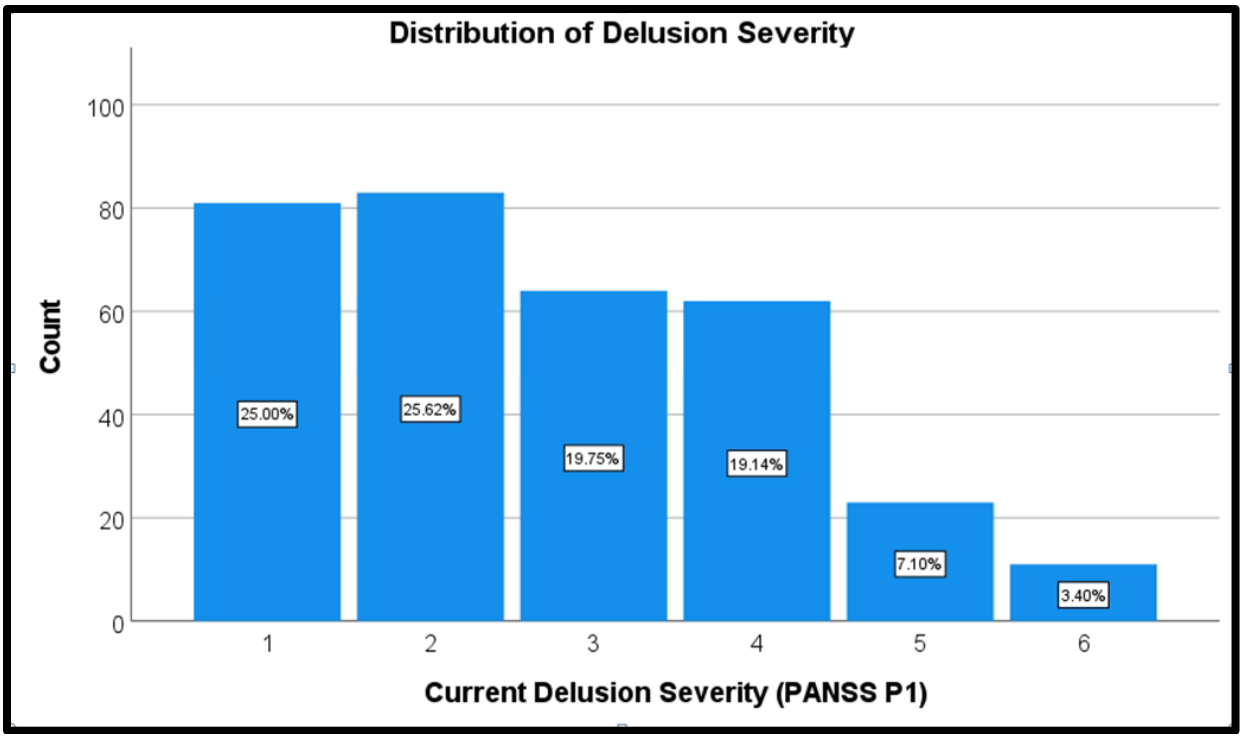

Figure 4-4. Distribution of Delusion Severity

Current delusion severity was measured by the P1 item on the Positive and Negative Syndrome Scale (PANSS). Patients represented most of the severity spectrum, from no current symptoms (P1=1, leftmost column) to severe symptoms (P1=6, rightmost column). Percentage of the total sample that was rated at each symptom severity level is shown.

Differences between Patients and Healthy Controls in Resting State Effectivity Connectivity of D-PE Regions

Prior to evaluating the hypotheses, the mean effective connectivity among the ROIs was assessed within just the healthy subjects to provide context for what significant "normal" connectivity of these regions may look like. Connectivity was characterized by inhibition of r Caudate by r DLPFC, and of r DLPFC by r VLPFC. There was excitation of: r Caudate by r VLPFC and l Caudate, the l Caudate by the r Caudate, and the r VLPFC by the r Caudate. Finally, there was disinhibition of each ROI (Table 4-2, Fig 4-5A). Next, the mean effective

connectivity was characterized within just the psychosis patients., There was inhibition of the Midbrain by the l Caudate. There was excitation of r Caudate by the l Caudate, and of the l Caudate by the r Caudate and r DLPFC. There was disinhibition of all nodes except r Caudate (Table 4-3, Fig 4-5B).

Next the results of the GLM comparing healthy controls to patients was evaluated (hypothesis 1). Patients' connectivity was significantly distinguished from controls by inhibition of: the r Caudate by the r VLPFC, the r DLPFC by the Midbrain, and the Midbrain by the r Caudate, and further self-inhibition of all nodes except the l Caudate (Table 4-3, Fig 4-5C). The same effective connectivity analysis restricted to psychotic probands with antipsychotic medication information (so that the medication covariate could be included) found a very similar pattern, with the only difference being that R Caudate excitation of midbrain was not significant (Appendix C-Table 3, Appendix C-Fig 3C).

Table 4-2. Summary of Effective Connectivity Results for Healthy Controls

Connection →	Parameter Estimate (typical is 0.1 Hz)	Connection Valence: + excitatory - inhibitory
Group Mean Effective Connectivity		
rCaud to rCaud	-0.134	-
rCaud to lCaud	0.198	+
rCaud to rVLPFC	0.103	+
lCaud to rCaud	0.214	+
lCaud to lCaud	-0.236	-
rVLPFC to rCaud	0.132	+
rVLPFC to rVLPFC	-0.277	-
rVLPFC to rDLPFC	-0.074	-
rDLPFC to rCaud	-0.173	-
rDLPFC to rDLPFC	-0.263	-
Midbr to Midbr	-0.178	-

N= 182 Healthy Controls Subjects

Table 4-3. Summary of Effective Connectivity Results in Association with Patient Status

Connection →	Parameter Estimate (typical is 0.1 Hz)	Connection Valence: + excitatory - inhibitory	Connectivity: ↑Increased ↓Decreased
Patient Status Association			
rCaud to rCaud	-0.195		↓
rCaud to Midbr	-0.074		↓
rVLPFC to rCaud	-0.118		↓
rVLPFC to rVLPFC	-0.306		↓
rDLPFC to rDLPFC	-0.158		↓
Midbr to rDLPFC	-0.116		↓
Midbr to Midbr	-0.102		↓
Group Mean Effective Connectivity			
rCaud to lCaud	0.200	+	
lCaud to rCaud	0.232	+	
lCaud to lCaud	-0.282	-	
rVLPFC to rCaud	0.087	+	
rVLPFC to rVLPFC	-0.346	-	
rDLPFC to rCaud	-0.078	-	
rDLPFC to rDLPFC	-0.264	-	
Midbr to Midbr	-0.199	-	

N= 506 Subjects – 182 Healthy Controls and 324 Psychosis Patients

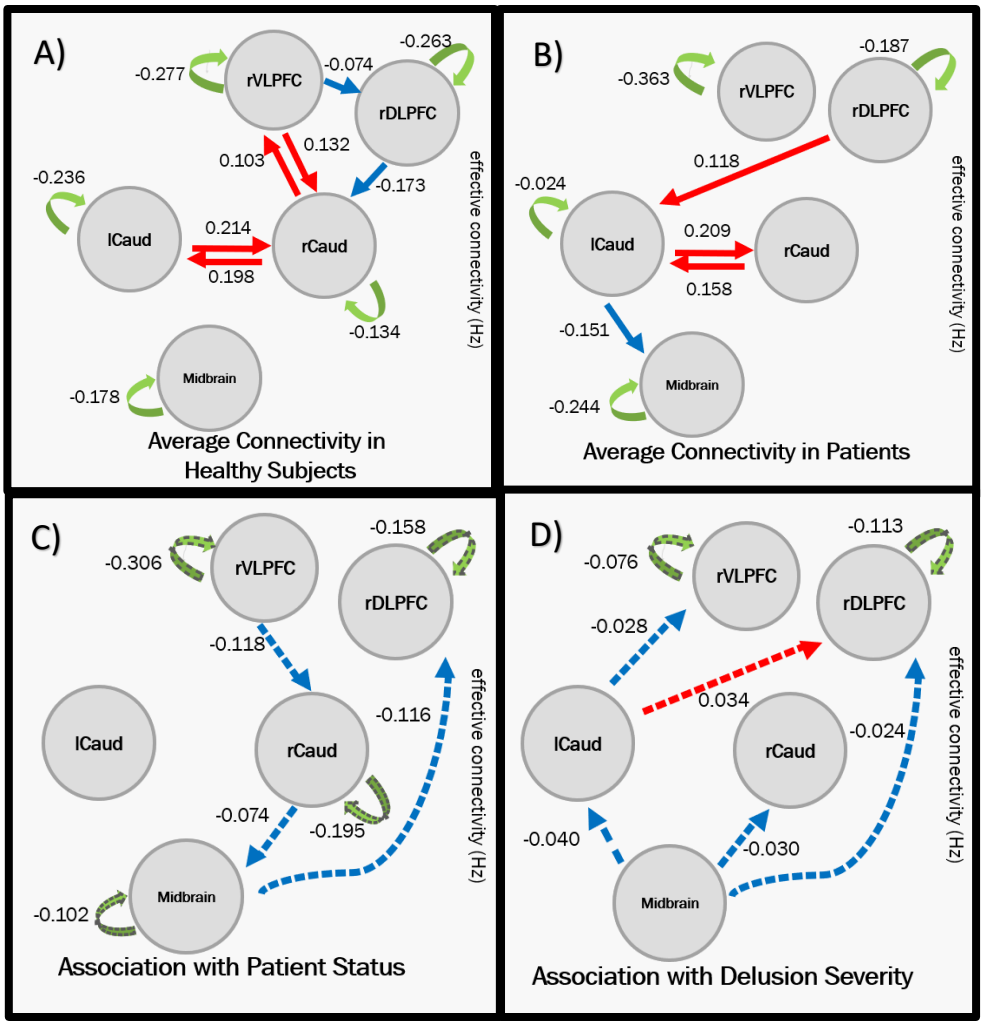

Figure 4-5. D-PE Network Effective Connectivity Relationships

Between node connection value represent the rate of change in neural response in one region due to activity from the connected regions, measured in hertz (Hz). Red arrows/positive values are excitatory connections; blue arrows/negative values are inhibitory connections. Self-connection values represent self-inhibition in each region, measured of sensitivity to input measured in a unitless log scaling parameter (multiplying the default value of -0.5Hz that is part of the DCM biologically based priors for self-inhibition). The self-connections are depicted as green arrows starting and ending at the same node – negative values imply disinhibition (positive would indicate stronger self-inhibition, though none were positive in this analysis). Solid lines indicate average connectivity within group and broken lines indicate changes in average connectivity due to effect of interest: group difference (c) or association with delusion severity (d).
A) Average Effective Connectivity in Healthy Subjects
B) Average Effective Connectivity in Psychosis Subjects
C) Changes in Effectivity Connectivity associated with Patient Status
D) Changes in Effectivity Connectivity associated with Delusion Severity

Association of Resting State Effectivity Connectivity in D-PE Regions with Delusion Severity

Finally, we evaluated results of effective connectivity associations with delusional severity (hypothesis 2). Greater delusional severity was significantly associated with excitation of the r DLPFC by the l Caudate; inhibition of the r VLPFC by the l Caudate; inhibition of the r DLPFC, r Caudate and l Caudate by the Midbrain; and greater disinhibition of the r VLPFC and r DLPFC (Table 4-4, Fig 4-5D). When additionally controlling for antipsychotic medication, current delusion severity was associated with a somewhat similar pattern as in the full sample, although some connections were no longer significant, including excitation of l Caudate to r VLPFC, Midbrain to L Caudate and to r DLPFC; while unique to the medication-controlled subsample, r VLPFC had significant excitation of r Caudate. (Appendix C -Table 4, Appendix C – Fig 3D).

Table 4-4. Summary of Effective Connectivity Associated with Delusion Severity

Connection →	Parameter Estimate (typical is 0.1 Hz)	Connection Valence: + excitatory - inhibitory	Connectivity: ↑Increased ↓Decreased
Delusion Association			
lCaud to rVLPFC	-0.028		↓
lCaud to rDLPFC	0.034		↑
rVLPFC to rVLPFC	-0.076		↓
rDLPFC to rDLPFC	-0.113		↓
Midbr to rCaud	-0.030		↓
Midbr to lCaud	-0.040		↓
Midbr to rDLPFC	-0.024		↓
Group Mean Effective Connectivity			
rCaud to lCaud	0.158	+	
lCaud to rCaud	0.209	+	
lCaud to lCaud	-0.264	-	
lCaud to Midbr	-0.151	-	
rVLPFC to rVLPFC	-0.363	-	
rDLPFC to lCaud	0.118	+	
rDLPFC to rDLPFC	-0.187	-	
Midbr to Midbr	-0.244	-	

N= 324 Psychosis Patients

Discussion

This study aimed to explore neural dynamics associated with psychotic delusions by investigating the resting state effective connectivity in brain regions involved in prediction error response. The prediction error neural system was selected as it appears to be the leading neural system for understanding delusions, though to date it has not provided clear or consistent understanding of delusions. The analysis of relationships among key nodes of this system in a large sample of psychosis patients and healthy controls was intended to fill in these knowledge gaps. In summary, significant effective connectivity associations of the prediction error nodes

were found that distinguished psychosis patients from healthy individuals, and that associated with delusion severity.

The specific findings were that, for hypothesis 1, psychosis patients differed from healthy controls in effective connectivity within the prediction error network via (1a) inhibition of the r caudate by the r VLPFC, e.g., increased top down control of the subcortical region by a prefrontal region, and yet also there was (1b) greater bottom up inhibition of DLPFC by the midbrain. There was also (1c) greater disinhibition of nearly all nodes including both prefrontal nodes, and (1d) greater inhibitory effective connectivity of the r caudate on the midbrain. For the second aim of assessing delusion severity associations, there was (2a) only one excitatory connectivity association, which was bottom-up excitation from the l caudate to the r DLPFC, (2b) increased inhibition from the midbrain to both other subcortical ROIs and one cortical ROI (r DLPFC), (2c) increased inhibition from l caudate to r VLPFC, and (2d) greater disinhibition of both prefrontal nodes. Broadly, these effective connectivity results provide novel insight into the causal paths which may underlie delusions in psychotic disorders. Additionally, the results provide further evidence that altered connectivity of the prediction error system may be a useful transdiagnostic biomarker of delusions, implicating several abnormalities in basal ganglia – cortical circuitry.

Effective Connectivity Alterations in D-PE network associated with Psychosis and Delusion Severity

Prior to reviewing the findings from the primary analyses assessing psychosis patients relative to controls, and assessing connectivity associations to delusion severity, it is helpful to address the findings of resting state D-PE network dynamics within healthy subjects, for context.

Excitatory connections were observed between the left and right caudate, and between the r VLPFC and the r caudate. All other between-node connectivities were inhibitory: r VLPFC inhibited r DLPFC, r DLPFC inhibited r Caudate. All nodes had significant self-disinhibition, and the midbrain was not significantly connected to any other node (except itself). Excitatory relationships between the right and left caudate are consistent with expected synchrony between bilateral structures at rest, and the inhibitory input from r DLPFC to r caudate is consistent with the top down control model of prefrontal cortex over subcortical structures. Other aspects of the connectivity found in healthy controls depict a more complicated picture, including no contra-lateral PFC inhibitory connectivities to l Caudate, but an excitatory connectivity from r VLPFC to r Caudate, and vice versa. This more complicated set of connectivity findings may be closer to ground truth of directed connectivity relationships, as much of the understanding of top down control connectivity has been developed with simpler correlational approaches, rather than through the testing of directed models, as was done here with DCM. While correlational studies would detect that these systems are connected (e.g., fluctuate over time in a similar manner), they would not distinguish whether the connections are inhibitory or excitatory. Interpreting the significant connectivities of the healthy group is somewhat beyond the scope of this study. For the most part, it serves as a useful baseline for interpreting changes in the patient group. Similarly, the patient group significant average connectivity schematic is helpful to reference in interpreting differences between controls and patients. Any apparent differences between the mean patient connectivity and mean control connectivity upon visual inspection comparing the upper panels of Fig 3 should be sought for statistical test-based substantiation, e.g., the results of the between-group comparison of connectivities depicted in Fig 3c. Furthermore, the differences between the groups here were tested in the hierarchical PEB framework, which incorporated the

magnitude of effective connectivities within individual subjects as well as their precision via probability distributions, so all results will not translate into simple group mean comparisons.

Several commonalities were found when assessing across the two primary analyses. In general, these suggest that there are abnormalities that distinguish patients from controls, and that the more aberrant the abnormality is, the more severe the delusion. First these overlapping observations from the analyses are discussed, then additional observations that were only in one or the other analysis are considered. It is notable that such a specific model being tested as the one used in this study will have very limited direct comparisons in the literature. The findings are novel by virtue of the relatively infrequent use of the spDCM method in psychosis research, offering new insights into directed relationships among the brain systems.

Connectivity that Both Distinguishes Patients from Controls and is Associated with Delusion Severity

Greater disinhibition of the right ventrolateral and dorsolateral prefrontal regions was a prominent, consistent finding, seen in distinguishing patients from controls, and associated with delusion severity (green self-referring arrows in Fig 4-3). This finding was also observed when controlling for anti-psychotic medication (Appendix C - Fig 3), extending the robustness of the observation. The state of being disinhibited in this DCM model context suggests increased sensitivity to inputs. This implies the self-inhibition of the right prefrontal cortex -conventionally mediated by GABAergic interneurons- is impaired. One way to understand this finding of impaired self-regulation of these lateral frontal nodes is as a corroboration of numerous studies reporting dysfunction of the lateral prefrontal cortex in psychosis (Dolan et al., 1993; Lewis and Hashimoto, 2007). There are conflicting reports of both hyper- and hypo-activation of prefrontal

cortex in psychotic disorders during tasks, though evidence leans toward hypoactivation during cognitively demanding tasks (Taylor, 1996; Callicott et al., 2000; Perlstein et al., 2001; Potkin et al., 2009; Koike et al., 2011). The observation of increased disinhibition/sensitivity could correspond to either hyper- or hypo- activity in task-based studies because evoked activity could be enhanced in an increased sensitivity state, or may be inadequately differentiated by task conditions due to a ceiling effect, and the lack of change compared to controls, who can show change between task conditions, may appear as "hypoactivity."

The finding of increased disinhibition of prefrontal prediction error nodes is interesting to consider in terms of being either a *primary* feature of psychosis pathophysiology or a downstream consequence of other circuitry alterations. In the former case, if disinhibition of lateral prefrontal cortex is a primary disease feature, an expected downstream consequence might be greater inhibition of subcortical circuits. The study results provide mixed evidence in support of this, as the data shows increased top down inhibition in association with psychosis but only for the r VLPFC. Alternatively, rather than being a primary pathophysiology, in the context of the aberrant salience framework that relies on a hypothesis of increased dopaminergic tone in basal ganglia, a consequence may be increased prefrontal disinhibition as a feedback response. This may be more consistent with evidence supporting hypo-frontal activity may be unassociated with positive psychosis symptoms; this is further supported by its more frequent association of prefrontal cortical dysfunction with motor and cognitive deficits in schizophrenia (Dolan et al., 1993).

A second observation from both the between-group comparison and delusion severity associations was inhibition of r DLPFC by the midbrain. First, the finding of inhibitory influence from the midbrain in patients contrasts with the healthy sample, where there is no evidence of

significant resting state effective connectivity between the midbrain and frontal nodes. This lack of connectivity in the healthy group may seem contrary to well-established evidence of dopamine signaling from the midbrain to prefrontal cortex, which for prediction error can serve as a gating signal to facilitate updating the representation of the expected outcome (Braver and Chhen, 2000; Seamans and Yang, 2004). However, our lack of finding effective connectivity between these regions could be due to the use of resting state data, less sensitive to this signaling than task-based data. In any case, given that midbrain signaling to prefrontal cortex is important during prediction error, the observation of context-independent inhibition of midbrain to prefrontal targets in patients, and in association with delusion severity, is exactly the type of novel observation sought in this analysis. It suggests that an intrinsic neural system dynamic abnormality in psychosis patients - inhibition of prefrontal cortex by the midbrain - may be a neural system alteration impacting prediction error processes when evoked. This abnormality would be overlooked when examining prediction error processes directly, e.g., during a task. Inhibition from the midbrain to the caudate bilaterally was also observed in association with delusion severity (but not patient status) but may be interpreted similarly. It may be an important background neural dynamic abnormality that may interfere with prediction error processes when invoked, but in this case, the abnormality is expected to be prominent in only those with greater delusional severity. Follow-up studies to test these interpretations should couple within participant analysis of intrinsic connectivity within the prediction error system and an analysis of effective connectivity during prediction error task performance within a suitably matched design paradigm.

Findings Unique to Each Primary Analysis

In addition to prefrontal disinhibition, increased disinhibition of the r caudate and midbrain was associated with patient status. The broad observation of increased in disinhibition in nearly all the prediction error nodes may reflect a core pathology of excitatory-inhibitory imbalance in patients. Self-inhibition is conceptualized to be mediated by pyramidal cells and interneurons (Bastos et al., 2012, Friston et al., 2019). As such, findings of increased disinhibition in psychosis may be consistent with reports of cellular abnormalities in excitatory/inhibitory imbalance in schizophrenia (Benes and Berretta, 2001, Kehre et al., 2008; Gao and Penzes, 2016; Selten et al., 2018).

Two additional observations unique to patient status were inhibition of the r caudate by the r VLPFC. This result is concordant with our initial hypothesis of increased top-down inhibition of subcortical nodes in psychosis. However, this finding was not found to extend to association with delusion severity. However, when analyzed in the subsample where antipsychotic medication was available and covaried (Appendix C - Fig 3), frontostriatal top-down inhibition (r VLFPC to the r caudate) was associated with both patient status and delusions. This may be in accordance with the partial efficacy of antipsychotic medication in normalizing D2 receptor-mediated dopaminergic transmission in frontostriatal connections (Kresby et al., 2018). In any case, the observation of top down inhibition distinguishing patients from controls confirmed predictions of hypothesis 1. The association to delusion severity (hypothesis 2) is less clear, though is more strongly supported than refuted by finding the association after co-varying for medication dose, as remaining variance may be more attributable to illness-related pathophysiology.

In addition to inhibition of the r DLPFC, delusion severity was associated with increased inhibition of the bilateral caudate nodes by the midbrain. Mechanistically, dopamine can act as

an inhibitory or excitatory transmitter depending on the post-synaptic receptor (Seamans and Yang, 2004; Mendoza and Foundas, 2008). A neuropharmacological pathway for the observed increases in inhibitory control by the midbrain is that increased release of dopamine in patients leads to increased occupancy and overstimulation of D2 receptors in the striatum (Howes et al., 2012; Seeman, 2013). An alternative pathway for inhibition is supported by research from Tritsch et al., (2014) who show in mice models that both ventral tegmental area and substantia nigra midbrain dopamine neurons mediate GABAergic signaling that inhibits firing of the striatum.

Another significant delusion association was increased bottom up excitation from l Caudate to r DLPFC. This finding remained when controlling for anti-psychotic dose. This is consistent with predictions within the aberrant salience hypothesis that delusions are a result of increased bottom up activity from the dopaminergic striatum. While there was some evidence of top down excitation of the l Caudate by r DLPFC in patients' effective connectivity, it was not a significant predictor of patient status, and only bottom-up excitation from the l caudate to the r DLPFC was associated with delusion severity. Together these findings suggest that delusions have an association with a connectivity pattern that is a fairly extreme perturbation of normal top down control – this connectivity is one of bottom-up influence from striatum to prefrontal cortex However, the observation is tempered by the parallel but opposite finding of inhibition of the r VLPFC by the l Caudate being associated with delusion severity, again a perturbation (though of a different kind) of the presumed normal state of top down control from prefrontal cortex to subcortical areas. Both are observations in which it is the *caudate* activity showing causal relationship to the prefrontal regions. In this scenario, salience detection occurring in the caudate

may occur and then those with greater delusion severity have more excitation to DLPFC, and more inhibition of VLPFC.

The lateral prefrontal cortex is part of the associative basal ganglia loop and is directly linked to the striatum via the mesocortical pathway. Furthermore, both resting state and task-based fMRI studies link activity in the striatum with activity of the dorsolateral and ventrolateral prefrontal cortex (Di Martino et al., 2008, Jarbo et al., 2015). Significant research supports a generalized role for DLPFC in supporting executive function which include cognitive control, working memory, and flexibility (Niendam et al., 2012). While studies suggest several overlaps in the function of the DLPFC and VLPFC, evidence also suggests the VLPFC may be specialized to mediate response inhibition by working as hub to integrate contextual representations of stimuli with motivational information that helps guide behavior (Sakagami and Pam, 2007, Hampshire et al., 2010). Thus, one possibility is that increased excitation by the caudate to the r DLPFC observed in both psychosis and delusion mediates increased engagement with irrelevant stimuli and non-contextual cues while increased inhibition of the r VLPFC impairs inhibitory control.

In summary, the observed increased striatal inhibition by the midbrain along with changes in the contralateral connectivity between the prefrontal cortex were associated with delusion symptoms and not psychosis more broadly. This might suggest that increased bottom up excitation may be an important biomarker of active delusional state. Though disinhibition of the lateral prefrontal cortex nodes and increased inhibition from the midbrain were also associated with delusion severity, these changes may represent a more core pathological feature of psychosis that helps mediate the risk of delusions. A useful next step may be to investigate the relevance of D-PE network alterations with other psychosis symptoms, particularly positive

ones. Furthermore, although studies have suggested that dysfunction in the right lateral prefrontal cortex plays a critical role in delusion arising in both primary and secondary psychosis (Corlett et al., 2007; Darby et al. 2017; Joyce, 2018) further study is needed to support evidence of lateralized frontostriatal connectivity as a driving mechanism in psychotic delusions.

Top-Down or Bottom-Up Account of Delusions

Although it was initially hypothesized that decreased top-down inhibition of the basal ganglia could be an important mediator of delusion symptoms, the results suggest top-down inhibition of endogenous striatal activity is increased in psychosis overall, and that increased bottom up influence from subcortical structures may play a more relevant role for delusions. These results could provide a possible explanation for decreased prediction error signal frequently observed in task-based studies of psychosis. Several studies have reported reduced activation in the striatum in response to unexpected outcomes in comparison with healthy subjects (Murray 2008, Koch et al., 2010, Morris et al., 2011). As presynaptic hyderdominerigia in the midbrain and ventral striatum is a repeated finding within schizophrenia patients (Howes et al., 2012), it is possible that an increase in presynaptic dopamine could lead to decreased "signal to noise ratio" in PE signaling, resulting in the relatively minimal disambiguation in striatal response that is seen between unexpected vs expected events in psychosis (Broyd et al., 2017). While the current results point to role for bottom up dysfunction of intrinsic brain activity in psychotic delusions, further investigation is needed to know if a similar relationship would be observed in a decision making or task-dependent contexts. The unexpected finding of significantly greater disinhibition of prefrontal nodes adds insight into understanding potential neural alterations leading to delusions, where regional self-regulation mechanisms are ineffective. While it was "normal" to show disinhibition in nodes (averages for healthy controls

indicated disinhibition for all nodes), patients had significantly greater disinhibition. This serves as a novel observation for testing in future studies and building into a more sophisticated model of neural dynamics underling delusions in psychosis.

Caveats and Considerations

There are several limitations that should be considered when interpreting the results for this study. The investigation was guided by an interest in understanding how endogenous connectivity with the prediction error system is related to psychosis and delusion symptoms. As such the focus was on the described D-PE sub-circuit due to the prior evidence supporting dysfunction in these regions with both psychosis generally and delusions specifically. However, several other brain regions encode prediction error and may be important for regulating prediction error response. For example, there is evidence that interactions with the insula, anterior cingulate cortex, thalamus, and parts of the default mode network may be critical to regulating normal salience process and are altered in psychosis (Limongi et al., 2013; Goulden et al., 2014; Nekovarova et al., 2014). Secondly, it is acknowledged that the use of antipsychotic medication remains an important confound in this study. This is specifically relevant as some of the investigated PE regions (the bilateral caudate nodes) are the target of D2 receptor antagonism of such medication, and action there is considered the primary mechanism of therapeutic relief (Kapur et al., 1999). Although there were some subtle differences, results were largely the same and medication did not independently predict changes in the effective connectivity within the D-PE circuit. The greatest difference was in changes to frontostriatal connectivity findings in which some results were no longer significantly associated with delusion severity. However, as symptom severity and medication dosage are positively associated, interpreting such findings could be akin to "throwing the baby out with the bathwater." Still, the data are available in the

appendix for comparison with future investigations that may better parse the question, such as in first episode psychosis studies. Lastly, a more complete picture of alterations in prediction error circuitry in psychotic delusions should also examine context dependent interactions which to be done will require using a task-based fMRI design that can powerfully capture prediction error and is sufficiently robust for both functional and effective connectivity analyses.

Conclusions

To understand the specific role of prediction error related neural system function in delusions it is important to characterize both the context-dependent dynamics and context-independent dynamics of the prediction error system. This study is the first to do the latter by reporting effective connectivity of resting state function within the prediction error system and its association with delusions in a large sample of transdiagnostic psychosis patients. The results provide evidence that both psychosis itself and delusion symptom severity are associated with altered connectivity dynamics. Compelling evidence included observations of connectivity that both distinguished patients from controls and, the more abnormal it was, the more severe delusions were. Specifically, these observations were of increased inhibition from the midbrain to DLPFC, and increased disinhibition of prefrontal regions. Additional observations of connectivity changes related to just delusion severity suggest more state (more severe illness), rather than trait, features of neural dynamics. These included bottom-up influence of the caudate on prefrontal cortex. Overall, this investigation supports prior research that that regions active during prediction error and salience processing are disrupted in psychotic illness that are characterized by active delusions and provides new insight into how intrinsic connectivity within this network is altered. It reveals a potentially complex picture that must be validated and investigated further in future studies.

Chapter V: Discussion

Delusions are illogical false beliefs that remain fixed even when presented with disconfirming evidence. Although delusional experiences can occur in both healthy and mentally ill individuals, they are most prevalent in and are often signature features of psychotic disorders. Within the primary psychotic disorders - which include schizophrenia, schizoaffective disorder, and bipolar with psychosis - delusions share many features. They tend to be polythematic in nature, often co-occur with other positive psychosis symptoms such as hallucinations and disorganized thoughts, and for some patients, may be alleviated partially or even fully by anti-psychotic medication therapy. However, more often than not, patients with prominent delusions to their illness usually remain at least somewhat delusional. Understanding the etiology of delusions is therefore urgently needed.

The biological basis of delusions remains unclear. This is despite culminative evidence over the past decades supporting disrupted dopamine functioning as a prevalent observation in psychosis. In particular, it is recognized that D2-receptor antagonism within the striatum is the primary correlate of efficacy of antipsychotic medications. Accordingly, cognitive theories attempting to link the presentation of delusions to disrupted dopamine systems have emphasized the importance of the basal ganglia for motivated behaviors and learning. Dopamine has been shown to function as a biological substrate within the brain to encode prediction errors, conceptualized as the difference between an expected outcome and the actual outcome, a process critical to reward learning. Within the interconnected incentive salience and predictive coding accounts of psychosis, delusional beliefs are thought to arise as aberrant salience is given to irrelevant stimuli and reduced attention is allotted to unexpected or normally surprising stimuli.

This theory is supported by evidence of abnormal prediction error patterns in psychosis patients and some preliminary evidence linking this to delusion symptoms.

In terms of evidence of neural system alterations, a few prior works have reported that abnormal brain activation in prediction error related brain regions is associated with psychosis and specifically with delusion symptoms. Overall, while promising, the evidence has been limited, and the explicit neurophysiological mechanisms underlying these changes remain opaque. In this regard one aspect that still lacks clarity is the interaction of the components of prediction error circuitry in psychosis and how these interactions may be related to delusion symptoms. In this project, several neuroimaging analysis methods were used to characterize the interactions of the prediction error sub-circuit in psychosis and to investigate their associations with delusion severity. Moreover, this project has been done in a transdiagnostic sample of patients diagnosed with the range psychotic disorders where delusions are commonly present, and for 2 of the 3 analyses, with fairly large samples, which the prior work has lacked. Throughout, I focused on a subset of the prediction error circuit, termed the D-PE network, that has been found both to have abnormal BOLD activation in psychosis and to be associated with delusion symptoms.

I aimed to investigate the following central questions. 1) Is the prediction error a biomarker of delusions? Specifically, can the association of delusions with prediction error-related brain activation be replicated within a larger transdiagnostic psychosis sample? 2) Within associative learning, is task dependent connectivity in the D-PE circuit altered in psychosis and/or delusions? 3) Outside the context of prediction error tasks, is the intrinsic connectivity of prediction error regions altered in psychosis generally, and in line with delusion severity? 4) Is there empirical evidence to support a specific network model of prediction error dysfunction in

delusions – such as reduced top-down inhibition in frontostriatal circuitry? In the prior chapters I demonstrated that activation and connectivity patterns in prediction error circuits are trans-diagnostically altered, and found novel evidence supporting an association of current delusion severity with alterations in the intrinsic effective connectivity of patients.

Prediction Error Activation Not Found To Be A Significant Predictor of Delusions

First, although a few prior studies have found abnormal prediction error activation to be associated with delusion symptoms, the evidence behind this observation has been both sparse and inconsistent. In chapter 2, I aimed to replicate these prior reports within a larger transdiagnostic psychosis sample (n=47), and further extend the finding by examining the relationship of delusion severity to task dependent connectivity changes. However, I did not find any evidence that prediction error neural activation, was associated with delusion symptoms, as has been previously reported. Several supplemental analyses of the present data were conducted to try to align more closely with the prior studies, such as using the same delusion measure, or excluding bipolar subjects, but replication remained elusive. The present study seemed robust relative to prior works in terms of task design with potentially improved signal/noise in the task design due to more prediction error trials included in the design, and a larger sample, so the lack of replication evokes consideration of the prior studies themselves. They investigated prediction error and symptoms using fMRI, and while there was some consistency in the findings, there was also variation in brain regions reported to be related to delusions, and importantly, these prior studies relied on small samples. It is also possible that the inconsistent results are additionally due to psychosis heterogeneity, including an unappreciated range of ways participants engage the associative learning task. Although the fMRI prediction error task study conducted here was

much larger than many previous prediction error studies in psychosis, it was not powered to detect such potential heterogeneous effects of prediction error aberrations.

An additional analysis of interest for the task was whether connectivity of the regions activated by prediction error task performance would show association to delusion severity, a novel analysis to date with respect to the question of how prediction error circuitry may be altered in association with delusions. The majority of results of this analysis indicated no such association. However, a post hoc analysis of connectivity employing the literature-based regions of interest yielded a few possible associations, current delusions being associated with increased connectivity of the visual cortex with r DLPFC during expectation violation and severity of lifetime delusion associated with decreased connectivity between the r Caudate and the orbital prefrontal cortex. The results were not consistent with one another (although that could be due to the fact that the delusion severity ratings were different from one another and not expected necessarily to show the same results). In any case, these findings require replication or direct testing in future studies.

D-PE RS-Functional Connectivity Not a Transdiagnostic Predictor of Delusions

Prior work and the present task-based study yield an inconsistent picture regarding the role of the magnitude of evoked prediction error neural system activation for delusions. Another possibility is that there are alterations to prediction error neural circuitry that may be more robustly observed outside of task performance. This has not previously been directly examined. So, in order to obtain this potentially fuller picture of how endogenous brain activity within the prediction error system might be associated with delusions, the whole brain functional connectivity of each D-PE region was tested in chapter 3, and the effectivity connectivity among the D-PE set of regions was evaluated in chapter 4. These evaluations were conducted with

resting state fMRI data from the multi-site BSNIP-1 cohort, a significantly larger dataset (n= 338 psychosis patients) to investigate the endogenous activity of the D-PE network. In chapter 3, reduced connectivity between the caudate and precuneus in psychotic subjects compared to healthy controls was the only significant finding. Although this is consistent with adjacent research demonstrating abnormal connectivity between the striatum and the default mode network in psychosis (Di Martino et al., 2008), owing to the prominent role of precuneus in the default mode network, the reduced connectivity was not a significant predictor of delusions. I concluded, therefore, that alterations in the *intrinsic connectivity* of the previously identified delusion-associated PE regions is not a trans-diagnostically identifiable mechanism underlying delusions. However, post hoc analyses exploring whether diagnostic group was a significant factor in connectivity associations to delusion severity suggested that the transdiagnostic approach may have been suboptimal. The results indicated the schizophrenia and schizoaffective groups had some significant connectivity associations with increased delusion severity.

Effective Connectivity within D-PE Network Is Associated with Delusions

In the final study I used novel dynamic causal modeling methods to investigate the directed neural dynamics of the resting state data within the D-PE network. Through this analysis I found not only that patients differed from healthy controls significantly in their effective connectivity of D-PE regions, but also there was also significant association of effective connectivity with delusion severity. These findings provide new perspectives to understand brain circuit changes in psychosis, as very few studies have examined prediction error circuits in psychosis using this method.

Some effective connectivity alterations were found to both distinguish psychosis from health and associate with delusion severity. One such finding was greater disinhibition (lower

self-inhibition) of the prefrontal nodes of the D-PE network. This is consistent with disruption of prefrontal cortex observed frequently in psychosis research and may reflect an underling excitatory/inhibitory imbalance in psychosis (Gao and Penzes, 2016; Selten et al., 2018). In this case, reduced self-inhibition is an instance of excitation being poorly regulated by inhibitory processes. This imbalance may be a general pathology in psychosis, as such reduced self-inhibition observations were present broadly in the D-PE network, distinguishing patients from controls. However, only the prefrontal reduction in self-inhibition associated with delusion severity. Increased sensitivity in these executive control nodes could have this specific association as a reflection of the ongoing presence and persistence of more severe delusional beliefs, more strongly present in working memory and more influential on behavior.

A second finding distinguishing patients from controls and associating with delusion severity was greater inhibition of the DLPFC by the midbrain. This is a bottom-up regulatory association. Although this connection was a specific finding for psychosis patients and delusions, there is evidence supporting that such a directional influence is present normally, such as from molecular research of the mesocortical pathway, showing inhibition of the spontaneous firing in the PFC by the midbrain VTA (Lapish et al., 2007). The fact that it was found to distinguish patients from controls and associate with delusion severity may reflect a downstream consequence of increased endogenous midbrain activity in psychosis (Howes et al., 2012). Delusions were more broadly associated with bottom up influence from the midbrain, with greater inhibitory connection also seen towards the striatal (caudate) nodes. The only excitatory connectivity association observed in relation to delusion symptoms was increased bottom-up excitation of the r DLPFC by the l Caudate, which contrasts with increased bottom up inhibition of the r VLPFC by the l Caudate. Within patients, greater inhibition of the r Caudate by r VLPFC

was also observed. Broadly, these effective connectivity results provide new evidence supporting abnormal connectivity in prediction error circuits for both psychosis and delusions and implicates several specific changes in frontostriatal and midbrain intrinsic connectivity as common neural pathologies which may underlie delusions in psychotic disorders. Future work can validate if these are replicable biomarkers and more directly associate their relationship to underlying neuronal and neurotransmitter systems.

While resting state effective connectivity was indeed found to be associated with delusion severity (chapter 4), resting state functional connectivity was not (chapter 3). What might the reason for these seemingly disparate results of intrinsic (task unrelated) connectivity of the D-PE network, especially considering the two analyses used predominantly overlapping datasets (95% shared subjects)? The primary explanation is that functional and effective connectivity capture two unique types of relationships of activity between brain regions. Functional connectivity is a correlation metric that simply informs on how coherent the activity is between two regions over time, while effective connectivity evaluates the directed influences between neuronal populations, measuring how activity in one brain region drives activity in another. Activity is quantified very differently in effective connectivity (e.g., is summarized by power of the frequency spectra, combined with priors in the model such as expected influences of the scanner and of biological signals, etc.). Second, the effective connectivity analysis was a restricted analysis within the D-PE network, possibly rendering it more suitable for detecting small effects, whereas the functional connectivity analyses searched the entire brain (conducting thousands of tests) to find alterations of connectivity to each seed, a process that is likely not sensitive to small effects. This may explain the lack of significant findings in the functional connectivity analysis for any of the D-PE regions being connected differentially to another D-PE region as it would

have to be found among all brain regions to be altered in patients relative to controls, or in association with delusion severity. For example, there was no significant functional connectivity difference between the midbrain and DLPFC or the VLPFC and caudate, although these did show significant relationships in the DCM. In addition to these findings depending on different mathematical relationships being identifiable, like most brain measures, the magnitude of effect in relation to clinical symptoms was also likely modest.

Converging Evidence of Prediction Error Dysfunction in Psychosis?

Prior neuroimaging studies have shown that prediction error in reinforcement learning is associated with a number of regions. These most consistently include the basal ganglia, midbrain, portions of prefrontal cortex, anterior cingulate, and insula (Garrison et al., 2013; Chase et al., 2015). Broad prediction error abnormalities have been observed throughout these regions in psychosis, with studies most often reporting psychosis groups showing increased prediction error or salience response to outcomes that are neutral or unsurprising, and a diminished response in these brain regions to surprising outcomes. In Chapter 2, the fMRI task engaged many of the same regions identified by prior prediction error tasks, overlapping with reinforcement learning and salience detection paradigms as well. However, the only significant difference between healthy controls and patients in our study was for the increased psychosis activation in the anterior cingulate cortex, rather than the more extensive findings prediction differences that have been reported. In addition, previous studies which observed differences in anterior cingulate reported decreased activation in patients. There are key aspects in the present study to consider relative to the literature. First, the prediction error task study was designed primarily to look at prediction error cognition in relation to delusions, so the sample size of healthy controls was relatively smaller than the psychosis group. The prior literature suggested

prediction error activation group differences should be relatively robust, so a modest healthy group size was presumed adequate in exchange for the larger patient sample in the present study. However, results comparing activation between the two groups may have been insensitive to group differences, in retrospect, due to the smaller healthy group. Earlier studies found patient vs. control differences in prediction error activation in midbrain, basal ganglia and frontal cortex (Murray et al., 2008; Waltz et al., 2008; Gradin et al., 2011; Koch et al., 2010; Romaniuk et al., 2010; Morris et al., 2011; Schalagenhauf et al., 2013), though not all regions were different in all studies. Also, many of these studies utilized currently liberal statistical thresholds in addition to recruitment of small sample sizes (n <25 for each cohort group). Together these factors increase the risk of false positives (Eklund et al., 2016, Turner et al., 2018). Interestingly, the present study may be one to add to a very recent and growing list of studies of large samples, published over the years the present project was conducted, each failing to find, as was the case in the present study, a difference in task-related prediction error response for the basal ganglia and striatum. This includes a study assessing prediction error in reversal learning tasks, in which strong prediction error responses were evoked by the task, but there were no group differences between 87 chronic schizophrenia patients and 61 controls (Culbreth et al., 2017). Haarmsa et al. (2020) report differences in activation for unsigned prediction error (error not associated with reward valence) in the right superior frontal cortex for first episode psychosis compared to healthy controls, but failed to find prediction error activation at all in midbrain or striatum, nor did patients and controls differ in these latter regions. Thus, considering these recent studies along with the present one, results have been inconsistent, with no one region clearly reflecting a source of prediction error dysfunction in psychosis. One possibility is that there may be heterogeneous underlying neural changes in psychosis that no study has yet been large enough to

parse effectively. There are also potentially important variations in the prediction error tasks utilized, as well as analysis methods differentiating studies.

Mixed Evidence of Prediction Error Dysfunction Associated with Delusions

In the effort to evaluate the relationship of prediction error and delusions, aspects of the rationale for doing so are worth re-assessing. While this study takes the relatively novel approach of seeking association of a neurocognitive function alteration to the specific psychosis symptom of delusional thought, much of the prior work has been done examining prediction error in relation to combined positive psychosis symptoms (hallucinations, disorganized thought, and delusions) more broadly, with fewer studies looking directly at delusions. The results from Chapter 2 do not strongly support a direct relationship between prediction error activation and delusions in the D-PE network. It is worth considering similarities and differences between this study and prior work. The original Corlett et al. (2007) finding of delusion severity being associated with r DLPFC activity was conducted in a small sample (n=12) of patients recently diagnosed with first episode psychosis. In addition to sample size and different population, a key difference from the present study was that they investigated prediction error activation in a sample in which all patients were not medically stabilized, which increases the risk that that antipsychotic medication was generating acute dopaminergic mediated changes in the prediction error system. Along the same lines, it is also possible that long-term anti-psychotic usage may work to regularize prediction error activation during the task. Longitudinal medication studies are needed to address this question. Thus, prior findings of activation of prediction error brain regions as being associated with delusional beliefs in psychosis (Corlett et al. 2007; Romaniuk et al. 2012) and healthy subjects with subclinical delusional thought features (Corlettt and Fletcher) were not replicated in the transdiagnostic chronic psychosis sample. This was true when

excluding bipolar subjects, a group that has been less represented in prior works, looking at the relationship to schizotypal beliefs in healthy controls, examining different dimensions of delusions such as distress and conviction. Even more focused, post-hoc analysis within the D-PE regions did not identify a significant relationship with prediction error activation and current delusion severity. However, the results are consistent with more recent, larger studies that failed to detect a relationship in both chronic psychosis (Culbreth et al., 2017) and first episode patients (Ermakova et al., 2018; Haarmsa et al., 2020). Another important consideration for prediction error research is specificity of the construct. Though the present prediction error study was modeled closely after the paradigm utilized by Corlett et al. (2007, 2012), the span of results in the literature may reflect the imprecise targeting of the prediction error construct. The prediction error tasks that studies use include a large variety of designs, from temporal difference modeling to neurocorrelates of surprise. Prediction errors exist in multiple forms, from low level sensory-perceptual to higher order reward prediction errors (Sterzer et al., 2018). Furthermore, there is some evidence suggesting signed and unsigned prediction errors may engage distinct cortical and subcortical regions (Klavir et al., 2013; Katthagan et al., 2018; Haarmsa et al., 2020). This variety may make it more challenging to identify consistent associations to delusions.

The significant challenges such as those mentioned above while investigating the prediction error system in action during a task are potentially addressed by an alternative approach of assessing the prediction error system in a more stable and controlled context, such resting state fMRI. This permits testing of the hypothesis that the D-PE nodes are altered continuously, outside of the context of evoked activity during a prediction error task, a state that seems likely in the context of ongoing chronic illness such as the psychotic disorders. Greater alteration of this aspect of the prediction error system may be a more sensitive correlate of

delusion severity. So far, very few studies have explored resting state activity with respect to delusions in psychosis, or with respect to the integrity of the prediction error system function. Regarding the D-PE network nodes specifically, there has been an inconsistent relationship with delusions and resting state functional connectivity of the striatum. Two studies using independent component analysis to look at with network connectivity reported both increased connectivity within the striatum (Sorg et al., 2013) and decreased connectivity within the left (but not right) striatum has been reported (Orliac et al., 2013) in association with delusion severity in schizophrenia patients. However, neither study captured potential changes in connectivity outside the defined network. When whole brain functional connectivity to the dorsal striatum (the caudate) was examined in Chapter 3, no relationship was found with current delusion severity. Furthermore, no relationship was seen with functional connectivity to midbrain or prefrontal nodes. This is consistent with more broad whole brain network analyses that have also not reported an association with delusions and networks containing these regions (Rotarska-Jagiela et al., 2010; Li et al., 2017).

Symptoms have been examined in relation to effective connectivity changes of prediction error circuits in a few prior studies used either during prediction error tasks or resting state paradigms. However, all of these investigations have been small in scope and the results are mostly incomparable as they largely investigated different sub-circuits (Limongi et al., 2020a; Csukly et al. 2020). The most related works are a study of 17 schizophrenia patients and 24 healthy controls exploring effective connectivity during a dynamic probabilistic inference task (Kaplan et al., 2016) and another study which investigated intrinsic effective connectivity (during resting state) between the left DLPFC and the left striatum in 19 first episode psychosis patients (Limongi et al. 2020b). The former reported differences in task modulated effective

connectivity for patients and controls. Additionally, they found that patients with delusions had decreased connectivity of left DLPFC to left anterior PFC during the events where patients inferred context changes in learning rules, and separately, there was increased connectivity of the left DLPFC to midbrain for patients following feedback trials in which uncertainty was resolved. Abnormal effective connectivity between the midbrain and l DLPFC in psychosis and in association with delusions is consistent with the findings reported in chapter 4. However, the direction and valence of influence is reversed. This may due to the fact that Kaplan et al. investigated effective connectivity during task performance, whereas the present study evaluated effective connectivity at rest, and these conditions, therefore, define the direction influence of one region to the other. The latter study (Limongi et al. 2020b) reported increased inhibitory tone within left DLPFC and left striatum and decreased excitatory connectivity between them to be associated with positive symptoms. Though left DLPFC was not examined within D-PE circuit this was consistent with delusions being associated with abnormal connectivity between l caudate and r DLPFC, however, it differed we observed increased bottom-up excitation.. Differences may be due to the DCM methodologies employed, which were somewhat divergent. Limongi et al. employed different assumptions about the neuronal microcircuit (e.g., that of a "two-state", where two nodes may have both excitatory and inhibitory relationships at the same time as a set of possible models; whereas the present study used a one-state assumption in its tested models), which make direct comparison of the results difficult .

In sum, the prior work on the association of prediction error neural system function and delusions has offered a few positive findings, but they are sparse and somewhat inconsistent. The prior work utilized fairly conventional imaging and analysis methods, and the parts of the present study that employed similar conventional approaches essentially only added to the inconsistency

in the literature. If anything its strengths lend to a conclusion of no strong association of task-related activation nor conventional functional connectivity of prediction error brain regions as associated with delusions. On the other hand, the effective connectivity analysis yielded a range of intrinsic connectivity relationships within the prediction error brain regions as associated with delusion severity. This was a novel method and yielded information that is consistent with the prior positive findings, but directed network findings, are difficult to juxtapose easily with the prior work. The next steps are to validate the results of the effective connectivity findings through replication and further experimental testing of the model. The effective connectivity results essentially keep the delusion severity association with prediction error circuit alterations a more, rather than less, viable hypothesis to continue to pursue.

Integration with Existing Neurocognitive Frameworks of Delusions

The aberrant salience and predictive coding framework have become prominent neurocognitive perspectives to explain belief updating and disrupted reality monitoring in psychosis (Kapur et al. 2003, Rosier et al., 2009; Adams et al., 2014). The culminative results presented in this dissertation provide mixed evidence in support of these hypotheses. Both the task and resting state results support evidence of changes in neural regions associated with salience. The challenge of recent studies to find whether delusions are associated with prediction error task activations puts into question whether predictive coding deficits are crucial for delusions. However, as many of the prior investigations has been lacking in both sample size and scope, more extensive research is needed to rule out the predictive coding hypothesis. Although, the investigation of intrinsic connectivity within prediction error associated regions cannot provide direct evidence of changes in prediction error processing during inferential learning and reality monitoring as of yet; further understanding remains to be developed to map such

observations of intrinsic activity to task performance. Importantly, effective connectivity during prediction error would be a key measure to endeavor to move forward, but such work is demanding in terms of a task design that would afford robust amounts of data for an adequately powered effective connectivity analysis. Fortunately, abnormal activity in resting state networks has been found to correlate with changes in activation and connectivity patterns in a wide range of cognitive tasks. In fact, presently resting state data may be more useful to investigate the alteration in prediction error circuitry within psychosis as the inconsistencies in modeling prediction error in task paradigms may be especially prohibitive for detecting subtle neurocognitive changes in relations to clinical symptoms.

There is molecular level evidence supporting the possibility of all the effective connectivity changes observed in the D-PE network by the present research. In particular, the effect of dopamine on basal ganglia and the prefrontal cortex may be a key mediator of observed results. One reason for this is that dopamine effects are notably complex. It is, generally conceived as a slow acting neuromodulator that can have both excitatory and inhibitory effects depending on the type of neuron receiving afferent input and which receptor type dopamine binds (Seeman, 2013). Furthermore, many dopaminergic neurons co-release GABA and glutamate which further expand the range of effects (den Ouden et al., 2012, Tritsch et al. 2014). Thus, there is cellular neural circuit evidence supporting the possibility of all the effective connectivity changes observed in the D-PE network by the present research, tying observations to the dopamine hypothesis of psychosis. However, the studies reported here do not directly address the dopamine hypothesis. Still, the findings can fit with the significant evidence of presynaptic hyperdopaminergia in the midbrain and striatum within psychosis (Howes et al., 2016), This may provide a useful perspective to interpret the effective connectivity findings. For

example, the psychosis literature has largely found increased striatal dopamine synthesis and reduced dopamine activity in the prefrontal cortex. The current findings of greater self-inhibition in the prefrontal cortex associating with delusions, may reflect hyperdopaminergia in the midbrain producing increased inhibitory effects via direct action on cortical pyramidal neurons or inhibitory interneurons. Alternatively, similar effects may occur via co-release of GABA in overactive midbrain neurons.

Another related and influential theory of delusions is the two-factor hypothesis, which asserts that delusions arise as result of both a low level perceptual aberration that generates the content of delusions and a second higher order neurocognitive process that maintains the delusions (Colheart et al., 2007; Moritz et al., 2017). In this model, both the aberrant salience and aberrant prediction error model can be fit - increased bottom up hyperactivity in salience systems relates to the first factor while reductions in top-down inhibitory control via diminished prefrontal cortex activity (responding incorrectly to feedback on information marked as salient) relates to the second factor (Corlett et al., 2007; Colheart et al., 2011; Broyd et al., 2017). The effective connectivity results in chapter 4 support increases in bottom up activity – via inhibition – as an important change in delusional patients. However, no evidence is provided of delusions being associated with reduction of top down inhibition by prefrontal cortical nodes. In fact, increased bottom up activity from the caudate nodes were found to be associated with delusions. These finding are consistent with argument that the two-factor hypothesis is less suitable for polythematic delusions and psychotic disorders, than monothematic delusions occurring after stroke or traumatic brain such as Capgras delusion (Broyd et al., 2017). However, it is possible that changes in inhibitory control in delusions are more context dependent or extends beyond the D-PE circuity defined in these analyses.

Limitations and Considerations

There are several overarching limitations that should be considered when interpreting these findings presented here. First, this investigation was guided primarily by an interest in understanding how activity of the prediction error system relates to psychosis and delusion symptoms. The focus on the described D-PE sub-circuit is due to the prior evidence supporting dysfunction in these regions with both psychosis generally and delusions specifically. However, several other brain regions encode prediction error and/or may be important for regulating prediction error response. For example, there is evidence that interactions with the insula, anterior cingulate cortex, thalamus, and parts of the default mode network may be critical to regulating normal salience processes and are altered in psychosis (Limongi et al., 2013; Goulden et al., 2014; Nekovarova et al., 2014). The risk is that prior papers on prediction error and delusions are small in sample size and may fail to detect effects in these additional regions. To that end, these additional regions were not entirely ignored in the present analysis: they could have emerged as significantly altered in connectivity to the D-PE regions in the functional connectivity analysis, or in the analysis of prediction error task-related activation, so this concern is somewhat mitigated.

Secondly, another important consideration for this work is how delusion symptoms were measured. As opposed to a direct capture of patients' delusional experiences (a nearly impossible task given the subjective nature of the experience, presumably), delusional history was quantified with some standardized measures and these were related to the neural measures. Specifically, current delusions were assessed with clinical interviews which assessed severity over the past week, per the Delusion item of a standard clinical assessment (PANSS; Positive and Negative Symptom Scale). This is by far the most common approach to study delusions in the literature,

but this may be imprecise. If delusional state is acutely reflected in brain states, imprecise clinical measures averaging over a week period may obscure relevant effects. Further imprecision may have come from distance between the clinical interview and the scan. The B-SNIP1 resting state scans were not systematically collected on the same day as the clinical symptom interview and that separated in measurement may have introduced greater noise in the analysis. Furthermore, another delusion assessment related concern is that this project did not address potential sources of biological heterogeneity that may be associated with delusional subtypes (paranoid, grandiose, somatic, ….) nor extensively dimensions (level of conviction, level of distress, ...). Instead, it presumed the existence of a common underlying mechanism.

Third, it is acknowledged that use of antipsychotic medication remains a very important confound in the investigation of both the neurobiology of delusions and prediction error. The patients analyzed in the study were all stably medicated, meaning no prescription changes within 30 days of entry into the study, which helps to rule out any acute changes associated with medication changes. In other words, neural systems studied were in fairly stable states. However, such states were influenced by the medications and therefore conclusions cannot be drawn with respect to pure illness-related factors. Still, the study is informative with respect to understanding the circuitry function in relation to prediction error and delusional severity, regardless of etiology that got it there. Supplementary analyses using average daily antipsychotic medication dose as a covariate yielded predominately similar results to those conducted without medication as a covariate. This fits with the fairly low correlation of antipsychotic dose with delusion severity (see appendix C – Figure 2). It also fits with recent work by our group, where medication in this sample was explicitly examined for dose-dependent effects on striatal connectivity, but no association was found (Herms et al., 2020). Moreover, it is

not clear that adjusting for antipsychotic dosage using existing methods (chlorpromazine equivalent estimates per Andreasen et al. (2010)) adequately accounts for medication effects. Although antipsychotic medications are efficacious for D2 dopaminergic receptor antagonism, they have additional neuromodulator effects such as serotonergic, cholinergic, and more. Further, patients are on additional classes of medications (antidepressants, mood stabilizers, etc), rendering it challenging to entirely account for such effects. This is also concordant with our observation within the D-PE network that average daily antipsychotic dosage was not a significant independent predictor of effective connectivity. Nonetheless, future work focusing specifically within first episode psychosis (FEP) in a pretreatment then post treatment state may better parse out effects of medication from direct changes in the prediction error network related to delusions.

Lastly, it is important in this work to reflect on the broader limits on the technology and what the statistical approaches can and cannot tell us. Use of fMRI has been a useful tool to study the context-dependent and context-independent engagement of prediction error brain regions. However, several importance nuances may be obscured using this technology. In particular, fine spatial detail may be lost as each voxel can contains hundreds of thousands of neurons, along with diverse cell types and afferent inputs coming from throughout the brain. This can be of special concern for heterogenous regions like the midbrain and subcortical nuclei. This also impedes our ability to make inferences at the neuronal level. Another known challenge of fMRI is the time resolution. The time course of BOLD is on the order of seconds whereas the cognitive operations occur in much shorter time intervals. Thus, important aspects of the prediction error process are lost in the temporal and spatial smoothing of fMRI analyses.

Specific considerations apply to DCM. DCM attempts to model neuronal populations but relies on numerous approximations. This includes a predetermined hemodynamic response function and predetermined assumptions about the nature of the underlying neuronal model e.g. one-state or two-state. Previous versions of DCM required pre-defining a limited small space of models for direction comparison, severely increasing the amount of assumptions about the investigated neural network. Recent development of the Paremetric Emperical Bayes method - used in chapter 3 – eliminate this constraint. However, DCM investigation is still restricted to regions included in the initial model and relationship between regions is based on statistical dependence not electrophysiologically validated connectivity. In this regard, although direct pathways have been structurally and pharmacologically traced between all the regions modeled in the D-PE network, it is clear that intermediary regions such as additional structures in basal ganglia loops and salience network may play an important role in mediating effects within the circuit that have not been captured in this work. Thus, a broader examination of the prediction error circuits may be useful. Moreover, methods such as direct manipulation (e.g. transcranial magnetic stimulation, transcranial direct stimulation, and neuropharmacological studies in animal models) will be essential to precisely tease apart causal relationships within the D-PE network. Because the strategies employed in this project used static connectivity measures that summarize circuitry across the resting state scan, another useful way forward may be examining dynamic connectivity, e.g., assessments of differing connectivity states between the regions over time (over the few minutes of resting state data acquired). This may help clarify the relationship between the various effective connectivity findings and offer a more nuanced understating of prediction error system alterations. It could be that patients have more or less variability in connectivity in association with delusional severity. Indeed, there are an array of methods that

are continuously evolving to characterize connectivity and network properties of the brain, each of which may offer new insights into alterations in psychosis.

Future Research Strategies

Several questions arise from the work which future research can help to address. One important question is whether effective connectivity in task-induced prediction error cognition is altered in similar ways to that of intrinsic brain activity, as found in Chapter 4. The event-related fMRI task utilized in this study was not sufficiently powered to assess such task dependent changes. Future approaches to address this question should look at recruiting large samples and measuring prediction error in a paradigm more suitable for functional and effective connectivity analyses.

Future research can also address whether there are unique subgroup interactions with prediction error as a biomarker of delusions. Suggestive evidence supporting diagnostic heterogeneity was seen in Chapter 3, as delusion severity was found to be associated with resting state functional connectivity in the schizophrenia patients. While a more detailed examination of effective connectivity differences within psychotic disorders will be useful, future work should also look to examine if similar disruptions associated with delusions exist outside the non-primary psychotic disorders (e.g. Parkinson's, depression, dementia, drug-induced psychosis, monothematic delusional disorders). The latter work can help clarify if prediction error deficits are a sufficient and necessary property of all delusional experiences.

An additional question of interest is which of biological changes occurring in delusions are state vs trait dependent? It is unclear whether abnormal prediction error regulation would be stable biomarker in delusion prone individuals or whether prediction error abnormalities occur

only acutely during psychotic experiences. Within the fMRI study in Chapter 2, the relationship of prediction error activation with both current and lifetime psychosis was tested but no significant relationship was found with either measure. The effective connectivity results presented in chapter 4 suggest that many of the changes observed in broadly in psychosis also support the severity of delusions, specific findings point to increased inhibitory control from the midbrain and increased sensitivity/disinhibition within prefrontal cortical nodes. However, as some of the observations were unique to delusions such as increased bottom up frontostriatal excitation, it may suggest these changes are more directly related to acute delusional experiences. Future investigations can go into further depth to clarify which changes in the prediction error circuit are related to lifetime propensity or history of delusions, and which are associated only with current delusional state. However, true investigation of the delusional state is inhibited by the challenge of directly capturing delusional experiences. One recent study attempted to directly assess delusional experiences by using patient self-report to categorize mental experiences across multiples periods of rest in the scanner (Raij et al., 2018). In the context of more precisely delineated delusional mental states that occur while mind wandering, such an approach while still limited, may provide a useful strategy to more deeply investigate effective connectivity changes in intrinsic activity of the prediction error network.

A related question that remains is how specific are the results to delusions? Delusions are positively correlated with many psychosis symptoms, in particular positive symptoms such as hallucinations. While most psychosis patients will experience both delusions and hallucinations at some point in their illness history, it is not clear what common vs divergent mechanisms these symptoms share. The shared connected effective connectivity results with delusion severity and psychosis diagnosis suggest these circuity alterations observed might be common changes that

extend across multiple psychosis symptoms. Future research should work to more extensively delineate which D-PE network changes are specific to delusions versus other symptom domains.

Conclusion

This thesis provides novel evidence that effective connectivity alterations within the intrinsic connectivity of prediction error circuits is associated with both psychosis and delusions, including decreased self-inhibition with the lateral prefrontal cortex along in addition to increased inhibition from the midbrain to the right dorsolateral prefrontal cortex. These are important new insights into the pathophysiology of delusions and may help to guide more directed investigation of neural circuitry in future biological and clinical studies. Moreover, the current work along with prior literature suggest that although alteration in the activation of prediction error associated brain regions does exist in psychosis, these disruptions are not reliably detected, nor reliably associated with delusions. Further research will be needed to validate whether prediction error activation associated with delusions either transdiagnostically or withing specific subgroups.

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). doi:10.1176/appi.books.9780890423349.

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in psychiatry*, *4*, 47.

Adams, R. A., Brown, H. R., & Friston, K. J. (2014). Bayesian inference, predictive coding and delusions. *AVANT. J. Philos. Int. Vanguard*, *5*, 51-88.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, *9*(1), 357-381.

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., & Roepstorff, A. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. Neuroimage, 127, 34-43.

Andreasen, N. C. (1984). Scale for the assessment of positive symptoms (SAPS). Iowa City: University of Iowa.

Andreasen, N. C., & Pierson, R. (2008). The role of the cerebellum in schizophrenia. *Biological psychiatry*, *64*(2), 81-88.

Andreasen, N. C., Pressler, M., Nopoulos, P., Miller, D., & Ho, B. C. (2010). Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs. Biological psychiatry, 67(3), 255-262.

Appelbaum, P. S., Robbins, P. C., & Roth, L. H. (1999). Dimensional approach to delusions: comparison across types and diagnoses. *American Journal of Psychiatry*, *156*(12), 1938-1943.

Argyelan, M., Ikuta, T., DeRosse, P., Braga, R. J., Burdick, K. E., John, M., ... & Szeszko, P. R. (2014). Resting-state fMRI connectivity impairment in schizophrenia and bipolar disorder. Schizophrenia bulletin, 40(1), 100-110.

Arjmand, S., Kohlmeier, K. A., Behzadi, M., Ilaghi, M., Mazhari, S., & Shabani, M. (2020). Looking into a Deluded Brain through a Neuroimaging Lens. The Neuroscientist, 1073858420936172.

Badner, J. A., & Gershon, E. S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Molecular psychiatry*, *7*(4), 405-411.

Bányai, M., Diwadkar, V. A., & Érdi, P. (2011). Model-based dynamical analysis of functional disconnection in schizophrenia. Neuroimage, 58(3), 870-877.

Bär, K. J., de la Cruz, F., Schumann, A., Koehler, S., Sauer, H., Critchley, H., & Wagner, G. (2016). Functional connectivity and network analysis of midbrain and brainstem nuclei. *Neuroimage*, *134*, 53-63.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695-711.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129-141.

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Explaining delusions: a cognitive perspective. *Trends in cognitive sciences*, *10*(5), 219-226.

Benes, F. M., & Berretta, S. (2001). GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. *Neuropsychopharmacology*, *25*(1), 1-27.

Bentall, R. P., Kinderman, P., & Kaney, S. (1994). The self, attributional processes and abnormal beliefs: towards a model of persecutory delusions. *Behaviour research and therapy*, *32*(3), 331-341.

Bentall, R. P., Myin-Germeys, I., Smith, A., Knowles, R., Jones, S. H., Smith, T., & Tai, S. J. (2011). Hypomanic personality, stability of self-esteem and response styles to negative mood. *Clinical Psychology & Psychotherapy*, *18*(5), 397-410.

Bernard, J. A., Goen, J. R., & Maldonado, T. (2017). A case for motor network contributions to schizophrenia symptoms: Evidence from resting-state connectivity. *Human brain mapping*, *38*(9), 4535-4545.

Blackwood, N. J., Howard, R. J., Bentall, R. P., & Murray, R. M. (2001). Cognitive neuropsychiatric models of persecutory delusions. American Journal of Psychiatry, 158(4), 527-539.

Blackwood, N. J., Bentall, R. P., Ffytche, D. H., Simmons, A., Murray, R. M., & Howard, R. J. (2004). Persecutory delusions and the determination of self-relvance: An fMRI investigation. *Psychological medicine*, *34*(4), 591.

Bolton TAW, Wotruba D, Buechler R, et al. Triple Network Model Dynamically Revisited: Lower Salience Network State Switching in Pre-psychosis. *Front Physiol*. 2020;11:66. Published 2020 Feb 11. doi:10.3389/fphys.2020.00066

Bortolotti, L., & Miyazono, K. (2015). Recent work on the nature and development of delusions. *Philosophy Compass*, *10*(9), 636-645.

Bramness, J. G., Gundersen, Ø. H., Guterstam, J., Rognli, E. B., Konstenius, M., Løberg, E. M., ... & Franck, J. (2012). Amphetamine-induced psychosis-a separate diagnostic entity or primary psychosis triggered in the vulnerable?. *BMC psychiatry*, *12*(1), 221.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. *Control of cognitive processes: Attention and performance XVIII*, 713-737.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815-834.

Broyd, A., Balzan, R. P., Woodward, T. S., & Allen, P. (2017). Dopamine, cognitive biases and assessment of certainty: a neurocognitive model of delusions. *Clinical psychology review*, *54*, 96-106.

Buchanan, R. W., & Carpenter, W. T. (1994). Domains of psychopathology: an approach to the reduction of heterogeneity in schizophrenia. Journal of Nervous and Mental Disease.

Cahill, C. (1996). Psychotic experiences induced in deluded patients using distorted auditory feedback. *Cognitive Neuropsychiatry*, *1*(3), 201-211.

Callicott, J. H., Bertolino, A., Mattay, V. S., Langheim, F. J., Duyn, J., Coppola, R., ... & Weinberger, D. R. (2000). Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cerebral cortex*, *10*(11), 1078-1092.

Chase, H. W., Kumar, P., Eickhoff, S. B., & Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, affective, & behavioral neuroscience*, *15*(2), 435-459.

Chen, X., Duan, M., He, H., Yang, M., Klugah–Brown, B., Xu, H., ... & Yao, D. (2016). Functional abnormalities of the right posterior insula are related to the altered self-experience in schizophrenia. *Psychiatry Research: Neuroimaging*, *256*, 26-32.

Cheng, P. W. C., Chang, W. C., Lo, G. G., Chan, K. W. S., Lee, H. M. E., Hui, L. M. C., ... & Mak, K. F. H. (2020). The role of dopamine dysregulation and evidence for the transdiagnostic nature of elevated dopamine synthesis in psychosis: a positron emission tomography (PET) study comparing schizophrenia, delusional disorder, and other psychotic disorders. *Neuropsychopharmacology*, 1-7.

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., ... & Tamminga, C. A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry*, *173*(4), 373-384.

Cloutier, M., Sanon Aigbogun, M., Guerin, A., Nitulescu, R., Ramanakumar, A. V., Kamat, S. A., ... & François, C. (2016). The economic burden of schizophrenia in the United States in 2013. Journal of clinical psychiatry, 77(6), 764-771.

Coltheart, M., Langdon, R., & McKay, R. (2007). Schizophrenia and monothematic delusions. *Schizophrenia Bulletin*, *33*(3), 642-647.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual review of psychology*, *62*, 271-298.

Conley, R. R., & Buchanan, R. W. (1997). Evaluation of treatment-resistant schizophrenia. *Schizophrenia Bulletin*, *23*(4), 663-674.

Consortium, I. S. (2009). Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. *Nature*, *460*(7256), 748.

Cook, J. A. (2006). Employment barriers for persons with psychiatric disabilities: Update of a report for the President's Commission. Psychiatric services, 57(10), 1391-1405.

Corlett, P. R., Aitken, M. R., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., ... & Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*, *44*(5), 877-888.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., ... & Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, *130*(9), 2387-2400.

Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). Toward a neurobiology of delusions. *Progress in neurobiology*, *92*(3), 345-369.

Corlett, P. R., & Fletcher, P. C. (2012). The neurobiology of schizotypy: fronto-striatal prediction error signal correlates with delusion-like beliefs in healthy people. *Neuropsychologia*, *50*(14), 3612-3620.

Corlett, P. R., & Fletcher, P. C. (2015). Delusions and prediction error: clarifying the roles of behavioural and brain responses. *Cognitive neuropsychiatry*, *20*(2), 95-105.

Cox, R. W., Reynolds, R. C., & Taylor, P. A. (2016). AFNI and clustering: false positive rates redux. bioRxiv, 065862.

Coyle, J. T. (1996). The glutamatergic dysfunction hypothesis for schizophrenia. *Harvard review of psychiatry*, *3*(5), 241-253.

Csukly, G., Szabó, Á., Polgár, P., Farkas, K., Gyebnár, G., Kozák, L. R., & Stefanics, G. (2020). Fronto-thalamic structural and effective connectivity and delusions in schizophrenia: a combined DTI/DCM study. Psychological Medicine, 1-11.

Culbreth, A. J., Gold, J. M., Cools, R., & Barch, D. M. (2016). Impaired activation in cognitive control regions predicts reversal learning in schizophrenia. Schizophrenia bulletin, 42(2), 484-493.

Darby, R., & Prasad, S. (2016). Lesion-related delusional misidentification syndromes: a comprehensive review of reported cases. *The Journal of neuropsychiatry and clinical neurosciences*, *28*(3), 217-222.

Darby, R. R., Laganiere, S., Pascual-Leone, A., Prasad, S., & Fox, M. D. (2017). Finding the imposter: brain connectivity of lesions causing delusional misidentifications. *Brain*, *140*(2), 497-507.

D'Astolfo, L., & Rief, W. (2017). Learning about expectation violation from prediction error paradigms–A meta-analysis on brain processes following a prediction error. *Frontiers in Psychology*, *8*, 1253.

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in psychology*, *3*, 548.

Di Martino, A., Scheres, A., Margulies, D. S., Kelly, A. M. C., Uddin, L. Q., Shehzad, Z., ... & Milham, M. P. (2008). Functional connectivity of human striatum: a resting state FMRI study. *Cerebral cortex*, *18*(12), 2735-2747.

Dolan, R. J., Bench, C. J., Liddle, P. F., Friston, K. J., Frith, C. D., Grasby, P. M., & Frackowiak, R. S. (1993). Dorsolateral prefrontal cortex dysfunction in the major psychoses; symptom or disease specificity?. *Journal of Neurology, Neurosurgery & Psychiatry*, *56*(12), 1290-1294.

Du X, Choa FS, Chiappelli J, et al. Aberrant Middle Prefrontal-Motor Cortex Connectivity Mediates Motor Inhibitory Biomarker in Schizophrenia. *Biol Psychiatry*. 2019;85(1):49-59. doi:10.1016/j.biopsych.2018.06.007

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, delusions and the "jumping to conclusions" reasoning bias: a systematic review and meta-analysis. *Schizophrenia bulletin*, *42*(3), 652-665.

Dutta, R., Greene, T., Addington, J., McKenzie, K., Phillips, M., & Murray, R. M. (2007). Biological, life course, and cross-cultural studies all point toward the value of dimensional and developmental ratings in the classification of psychosis. *Schizophrenia Bulletin*, *33*(4), 868-876.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the national academy of sciences, 113(28), 7900-7905.

Ermakova, A. O., Knolle, F., Justicia, A., Bullmore, E. T., Jones, P. B., Robbins, T. W., ... & Murray, G. K. (2018). Abnormal reward prediction-error signalling in antipsychotic naive individuals with first-episode psychosis or clinical risk for psychosis. *Neuropsychopharmacology*, *43*(8), 1691-1699.

Farde, L., Nordström, A. L., Wiesel, F. A., Pauli, S., Halldin, C., & Sedvall, G. (1992). Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine: relation to extrapyramidal side effects. *Archives of general psychiatry*, *49*(7), 538-544.

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive neuropsychiatry*, *12*(1), 46-77.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition (pp. 94-1). New York, NY, USA:: SCID-I/P.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). Structural clinical interview for DSM IV-TR Axis I disorder, Research version, Patient Edition with Psychotic Screen (SCID-I/PW/PSY SCREEN). Biometric Research, New York Psychiatric Institute, New York.

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., ... & Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature neuroscience*, *4*(10), 1043-1048.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48-58.

Fogelson, N., Litvak, V., Peled, A., Fernandez-del-Olmo, M., & Friston, K. (2014). The functional anatomy of schizophrenia: a dynamic causal modeling study of predictive coding. *Schizophrenia research*, *158*(1), 204-212.

Fornito, A., Zalesky, A., Pantelis, C., & Bullmore, E. T. (2012). Schizophrenia, neuroimaging and connectomics. *Neuroimage*, *62*(4), 2296-2314.

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human brain mapping*, *39*(7), 2887-2906.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Freeman D. (2006). Delusions in the nonclinical population. Current psychiatry reports, 8(3), 191–204. https://doi.org/10.1007/s11920-006-0023-1

Friedman, A., & Sienkiewicz, J. (1991). Psychotic complications of long-term levodopa treatment of Parkinson's disease. *Acta neurologica scandinavica*, *84*(2), 111-113.

Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome. *Clin Neurosci*, *3*(2), 89-97.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, *19*(4), 1273-1302.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1-3), 70-87.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage*, *34*(1), 220-234.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211-1221.

Friston, K. J. (2011). Functional and effective connectivity: a review. Brain connectivity, 1(1), 13-36.

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *Neuroimage*, *56*(4), 2089-2099.

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230-1233.

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *Neuroimage*, *94*, 396-407.

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., Van Wijk, B. C., ... & Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage*, *128*, 413-431.

Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2019). Dynamic causal modelling revisited. *Neuroimage*, *199*, 730-744.

Frith CD, Blakemore S, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res Brain Res Rev*. 2000;31(2-3):357-363. doi:10.1016/s0165-0173(99)00052-1

Frith C. D. (2004). Schizophrenia and theory of mind. *Psychological medicine*, *34*(3), 385–389. https://doi.org/10.1017/s0033291703001326

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences*, *7*(2), 77-83.

Gao, R., Russell, T. A., & Penzes, P. (2016). Synaptic Abnormalities and Neuroplasticity: Molecular Mechanisms of Cognitive Dysfunction in Genetic Mouse Models of Schizophrenia. In *Handbook of Behavioral Neuroscience* (Vol. 23, pp. 375-390). Elsevier.

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *The British Journal of Psychiatry*, *203*(5), 327-333.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1297-1310.

Goto, M., Abe, O., Miyati, T., Yamasue, H., Gomi, T., & Takeda, T. (2015). Head motion and correction methods in resting-state functional MRI. *Magnetic Resonance in Medical Sciences*, rev-2015.

Goulden, N., Khusnulina, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P., & Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. *Neuroimage*, *99*, 180-190.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., ... & Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, awr059.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences, 100*(1), 253-258.

Gründer, G., & Cumming, P. (2016). The dopamine hypothesis of schizophrenia: Current status. In The Neurobiology of Schizophrenia (pp. 109-124). Academic Press.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage*, *50*(3), 1313-1319.

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4-26.

Hadley, J. A., Nenert, R., Kraguljac, N. V., Bolding, M. S., White, D. M., Skidmore, F. M., ... & Lahti, A. C. (2014). Ventral tegmental area/midbrain functional connectivity and response to antipsychotic medication in schizophrenia. *Neuropsychopharmacology*, *39*(4), 1020-1030.

Haarsma, J., Fletcher, P. C., Griffin, J. D., Taverne, H. J., Ziauddeen, H., Spencer, T. J., ... & Murray, G. K. (2020). Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. Molecular Psychiatry, 1-14.

Herms, E. N., Bishop, J. R., Okuneye, V. T., Tamminga, C. A., Keshavan, M. S., Pearlson, G. D., ... & Sweeney, J. A. (2020). No connectivity alterations for striatum, default mode, or salience network in association with self-reported antipsychotic medication dose in a large chronic patient group. *Schizophrenia research*, S0920-9964.

Horga, G., Schatz, K. C., Abi-Dargham, A., & Peterson, B. S. (2014). Deficits in predictive coding underlie hallucinations in schizophrenia. *The Journal of Neuroscience*, *34*(24), 8072-8082.

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophrenia bulletin*, *35*(3), 549-562.

Howes, O. D., Kambeitz, J., Kim, E., Stahl, D., Slifstein, M., Abi-Dargham, A., & Kapur, S. (2012). The nature of dopamine dysfunction in schizophrenia and what this means for treatment: meta-analysis of imaging studies. *Archives of general psychiatry*, *69*(8), 776-786.

Howes, O. D., & Nour, M. M. (2016). Dopamine and the aberrant salience hypothesis of schizophrenia. *World Psychiatry*, *15*(1), 3.

Iglesias-Parro, S., Soriano, M. F., Prieto, M., Rodríguez, I., Aznarte, J. I., & Ibáñez-Molina, A. J. (2020). introspective and neurophysiological Measures of Mind Wandering in Schizophrenia. *Scientific reports*, *10*(1), 1-12.

Jann, K., Gee, D. G., Kilroy, E., Schwab, S., Smith, R. X., Cannon, T. D., & Wang, D. J. (2015). Functional connectivity in BOLD and CBF data: similarity and reliability of resting brain networks. *Neuroimage*, *106*, 111-122.

Jarbo, K., & Verstynen, T. D. (2015). Converging structural and functional connectivity of orbitofrontal, dorsolateral prefrontal, and posterior parietal cortex in the human striatum. *Journal of Neuroscience*, *35*(9), 3865-3878.

Joyce, E. M. (2018). Organic psychosis: The pathobiology and treatment of delusions. *CNS neuroscience & therapeutics*, *24*(7), 598-603.

Junginger, J., Barker, S., & Coe, D. A. (1992). Mood theme and bizarreness of delusions in schizophrenia and mood psychosis. *Journal of Abnormal Psychology*, *101*(2), 287.

Kahan, J., & Foltynie, T. (2013). Understanding DCM: ten simple rules for the clinician. Neuroimage, 83, 542-549.

Kapur, S., Zipursky, R. B., & Remington, G. (1999). Clinical and theoretical implications of 5-HT2 and D2 receptor occupancy of clozapine, risperidone, and olanzapine in schizophrenia. *American Journal of Psychiatry*, *156*(2), 286-293.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American journal of Psychiatry*, *160*(1), 13-23.

Karcher, N. R., Rogers, B. P., & Woodward, N. D. (2019). Functional connectivity of the striatum in schizophrenia and psychotic bipolar disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(11), 956-965.

Katthagen, T., Mathys, C., Deserno, L., Walter, H., Kathmann, N., Heinz, A., & Schlagenhauf, F. (2018). Modeling subjective relevance in schizophrenia and its relation to aberrant salience. *PLoS computational biology*, *14*(8), e1006319.

Kay, S. R., Flszbein, A., & Opfer, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin*, *13*(2), 261.

Kean, C. (2009). Silencing the self: schizophrenia as a self-disturbance. *Schizophrenia bulletin*, *35*(6), 1034.

Kehrer, C., Maziashvili, N., Dugladze, T., & Gloveli, T. (2008). Altered excitatory-inhibitory balance in the NMDA-hypofunction model of schizophrenia. *Frontiers in molecular neuroscience*, *1*, 6.

Kempf, L., Hussain, N., & Potash, J. B. (2005). Mood disorder with psychotic features, schizoaffective disorder, and schizophrenia with mood features: trouble at the borders. *International Review of Psychiatry*, *17*(1), 9-19.

Kerns, J. G., Cohen, J. D., MacDonald III, A. W., Johnson, M. K., Stenger, V. A., Aizenstein, H., & Carter, C. S. (2005). Decreased conflict-and error-related activity in the anterior

cingulate cortex in subjects with schizophrenia. *American Journal of Psychiatry*, *162*(10), 1833-1839.

Keshavan, M. S., Morris, D. W., Sweeney, J. A., Pearlson, G., Thaker, G., Seidman, L. J., ... & Tamminga, C. (2011). A dimensional approach to the psychosis spectrum between bipolar disorder and schizophrenia: the Schizo-Bipolar Scale. *Schizophrenia research*, *133*(1-3), 250-254.

Kimhy, D., Goetz, R., Yale, S., Corcoran, C., & Malaspina, D. (2005). Delusions in individuals with schizophrenia: Factor structure, clinical correlates, and putative neurobiology. *Psychopathology*, *38*(6), 338-344.

Klavir, O., Genud-Gabai, R., & Paz, R. (2013). Functional connectivity between amygdala and cingulate cortex for adaptive aversive learning. Neuron, 80(5), 1290-1300.

Knobel, A., Heinz, A., & Voss, M. (2008). Imaging the deluded brain. *European archives of psychiatry and clinical neuroscience*, *258*(5), 76.

Koch, K., Schachtzabel, C., Wagner, G., Schikora, J., Schultz, C., Reichenbach, J. R., ... & Schlösser, R. G. (2010). Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. *Neuroimage*, *50*(1), 223-232.

Koike, S., Takizawa, R., Nishimura, Y., Takano, Y., Takayanagi, Y., Kinou, M., ... & Kasai, K. (2011). Different hemodynamic response patterns in the prefrontal cortical sub-regions according to the clinical stages of psychosis. *Schizophrenia research*, *132*(1), 54-61.

Krawitz, A., Braver, T. S., Barch, D. M., & Brown, J. W. (2011). Impaired error-likelihood prediction in medial prefrontal cortex in schizophrenia. *Neuroimage*, *54*(2), 1506-1517.

Krishnan, R. R., Fivaz, M., Kraus, M. S., & Keefe, R. S. E. (2011). Hierarchical temporal processing deficit model of reality distortion and psychoses. Molecular psychiatry, 16(2), 129-144.

Kurtz, M. M., & Gerraty, R. T. (2009). A meta-analytic investigation of neurocognitive deficits in bipolar illness: Profile and effects of clinical state. Neuropsychology, 23(5), 551.

Kwapil, T. R., & Barrantes-Vidal, N. (2015). Schizotypy: looking back and moving forward. Schizophrenia bulletin, 41(suppl_2), S366-S373.

Lees-Grossmann L. (2010) Delusion. In: Leeming D.A., Madden K., Marlan S. (eds) Encyclopedia of Psychology and Religion. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-71802-6_159

Levinson, D. F., Mowry, B. J., Escamilla, M. A., & Faraone, S. V. (2002). The Lifetime Dimensions of Psychosis Scale (LDPS): description and interrater reliability. *Schizophrenia bulletin*, *28*(4), 683-695.

Lewis, D. A., & Hashimoto, T. (2007). Deciphering the disease process of schizophrenia: the contribution of cortical GABA neurons. *International review of neurobiology*, *78*, 109-131.

Li, T., Wang, Q., Zhang, J., Rolls, E. T., Yang, W., Palaniyappan, L., ... & Gong, X. (2017). Brain-wide analysis of functional connectivity in first-episode and chronic stages of schizophrenia. *Schizophrenia bulletin*, *43*(2), 436-448.

Limongi, R., Sutherland, S. C., Zhu, J., Young, M. E., & Habib, R. (2013). Temporal prediction errors modulate cingulate–insular coupling. *Neuroimage*, *71*, 147-157.

Limongi, R., Jeon, P., Mackinley, M., Das, T., Dempster, K., Théberge, J., ... & Palaniyappan, L. (2020). Glutamate and Dysconnection in the salience network: Neurochemical, effective-connectivity, and computational evidence in schizophrenia. *Biological Psychiatry*.

Limongi, R., Mackinley, M., Dempster, K., Khan, A. R., Gati, J. S., & Palaniyappan, L. (2020). Frontal–striatal connectivity and positive symptoms of schizophrenia: implications for the mechanistic basis of prefrontal rTMS. European Archives of Psychiatry and Clinical Neuroscience, 1-13.

Maher, B. A. (2006). The relationship between delusions and hallucinations. *Current psychiatry reports*, *8*(3), 179-183.

Mancuso, S. G., Morgan, V. A., Mitchell, P. B., Berk, M., Young, A., & Castle, D. J. (2015). A comparison of schizophrenia, schizoaffective disorder, and bipolar disorder: Results from the Second Australian national psychosis survey. *Journal of affective disorders*, *172*, 30-37.

Mann-Wrobel, M. C., Carreno, J. T., & Dickinson, D. (2011). Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. Bipolar disorders, 13(4), 334-342.

Manoliu, A., Riedl, V., Zherdin, A., Mühlau, M., Schwerthöffer, D., Scherr, M., ... & Wohlschläger, A. M. (2014). Aberrant dependence of default mode/central executive network interactions on anterior insular salience network activity in schizophrenia. *Schizophrenia bulletin*, *40*(2), 428-437.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*, *2*(2), 89-101.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339-346.

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage*, *61*(4), 1277-1286.

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2017). Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: a detailed meta-analysis. *Schizophrenia Bulletin*, *43*(2), 344-354.

Mendoza, J., & Foundas, A. (2007). *Clinical neuroanatomy: a neurobehavioral approach*. Springer Science & Business Media.

Mishara, A. L. (2010). Klaus Conrad (1905–1961): delusional mood, psychosis, and beginning schizophrenia. *Schizophrenia Bulletin*, *36*(1), 9-13.

Moran, P. M., Rouse, J. L., Cross, B., Corcoran, R., & Schürmann, M. (2012). Kamin blocking is associated with reduced medial-frontal gyrus activation: implications for prediction error abnormality in schizophrenia. *PloS one*, *7*(8), e43905.

Moritz, S., Pfuhl, G., Lüdtke, T., Menon, M., Balzan, R. P., & Andreou, C. (2017). A two-stage cognitive theory of the positive symptoms of psychosis. Highlighting the role of lowered decision thresholds. *Journal of behavior therapy and experimental psychiatry*, *56*, 12-20.

Morris, R. W., Vercammen, A., Lenroot, R., Moore, L., Langton, J. M., Short, B., ... & Weickert, T. W. (2012). Disambiguating ventral striatum fMRI-related bold signal during reward prediction in schizophrenia. *Molecular Psychiatry*, *17*(3), 280-289.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., Jones, P. B., Bullmore, E. T., Robbins, T. W., and Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol. Psychiatry 13, 267–276.

Murty, V. P., Shermohammed, M., Smith, D. V., Carter, R. M., Huettel, S. A., & Adcock, R. A. (2014). Resting state networks distinguish human ventral tegmental area from substantia nigra. *Neuroimage*, *100*, 580-589.

Mwansisya, T. E., Hu, A., Li, Y., Chen, X., Wu, G., Huang, X., ... & Feng, J. (2017). Task and resting-state fMRI studies in first-episode schizophrenia: A systematic review. *Schizophrenia research*, *189*, 9-18.

Nenadic, I., Langbein, K., Dietzek, M., Forberg, A., Smesny, S., & Sauer, H. (2015). Cognitive function in euthymic bipolar disorder (BP I) patients with a history of psychotic symptoms vs. schizophrenia. Psychiatry research, 230(1), 65-69.

Nekovarova, T., Fajnerova, I., Horacek, J., & Spaniel, F. (2014). Bridging disparate symptoms of schizophrenia: a triple network dysfunction theory. *Frontiers in behavioral neuroscience*, *8*, 171.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(2), 241-268.

Northoff G. (2014). Are Auditory Hallucinations Related to the Brain's Resting State Activity? A 'Neurophenomenal Resting State Hypothesis'. *Clinical psychopharmacology and neuroscience : the official scientific journal of the Korean College of Neuropsychopharmacology*, *12*(3), 189–195. https://doi.org/10.9758/cpn.2014.12.3.189

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *Journal of neuroscience*, *23*(21), 7931-7939.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329-337.

O'Doherty, John P., Tony W. Buchanan, Ben Seymour, and Raymond J. Dolan. "Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum." *Neuron* 49, no. 1 (2006): 157-166.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social cognitive and affective neuroscience*, *7*(5), 604-609.

Orliac, F., Naveau, M., Joliot, M., Delcroix, N., Razafimandimby, A., Brazo, P., ... & Delamillieure, P. (2013). Links among resting-state default-mode network, salience network, and symptomatology in schizophrenia. *Schizophrenia research*, *148*(1-3), 74-80.

Perlstein, W. M., Carter, C. S., Noll, D. C., & Cohen, J. D. (2001). Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *American Journal of Psychiatry*, *158*(7), 1105-1113.

Perälä, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsä, E., Pirkola, S., ... & Härkänen, T. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of general psychiatry*, *64*(1), 19-28.

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia bulletin*, *30*(4), 1005. *tin*, *25*(3), 553-576.

Picardi, A., Fonzi, L., Pallagrosi, M., Gigantesco, A., & Biondi, M. (2018). Delusional Themes across affective and non-affective Psychoses. *Frontiers in psychiatry*, *9*, 132.

Pini, S., de Queiroz, V., Dell'Osso, L., Abelli, M., Mastrocinque, C., Saettoni, M., ... & Cassano, G. B. (2004). Cross-sectional similarities and differences between schizophrenia, schizoaffective disorder and mania or mixed mania with mood-incongruent psychotic features. *European Psychiatry*, *19*(1), 8-14.

Poletti, M., & Sambataro, F. (2013). The development of delusion revisited: a transdiagnostic framework. *Psychiatry research*, *210*(3), 1245-1259.

Potkin, S. G., J. A. Turner, G. G. Brown, G. McCarthy, D. N. Greve, G. H. Glover, D. S. Manoach et al. "Working memory and DLPFC inefficiency in schizophrenia: the FBIRN study." *Schizophrenia bulletin* 35, no. 1 (2009): 19-31.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, *59*(3), 2142-2154.

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience*, *38*, 433-447.

Raij, T. T., Riekki, T., Rikandi, E., Mäntylä, T., Kieseppä, T., & Suvisaari, J. (2018). Activation of the motivation-related ventral striatum during delusional experience. *Translational psychiatry*, *8*(1), 283. https://doi.org/10.1038/s41398-018-0347-8

Razi, A., Kahan, J., Rees, G., & Friston, K. J. (2015). Construct validation of a DCM for resting state fMRI. *Neuroimage*, *106*, 1-14.

Reinen, J. M., Van Snellenberg, J. X., Horga, G., Abi-Dargham, A., Daw, N. D., & Shohamy, D. (2016). Motivational context modulates prediction error response in schizophrenia. *Schizophrenia bulletin*, *42*(6), 1467-1475.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. *European Journal of Neuroscience*, *35*(7), 1190-1200.

Rosa, M. J., Friston, K., & Penny, W. (2012). Post-hoc selection of dynamic causal models. *Journal of neuroscience methods*, *208*(1), 66-78.

Rosenheck, R., Leslie, D., Keefe, R., McEvoy, J., Swartz, M., Perkins, D., ... & Lieberman, J. (2006). Barriers to employment for people with schizophrenia. American Journal of Psychiatry, 163(3), 411-417.

Roiser, J. P., Stephan, K. E., Den Ouden, H. E. M., Barnes, T. R. E., Friston, K. J., & Joyce, E. M. (2009). Do patients with schizophrenia exhibit aberrant salience?. *Psychological medicine*, *39*(2), 199-209.

Rössler, W., Salize, H. J., van Os, J., & Riecher-Rössler, A. (2005). Size of burden of schizophrenia and psychotic disorders. *European Neuropsychopharmacology*, *15*(4), 399-409.

Romaniuk, L., Honey, G. D., King, J. R., Whalley, H. C., McIntosh, A. M., Levita, L., ... & Hall, J. (2010). Midbrain activation during Pavlovian conditioning and delusional symptoms in schizophrenia. *Archives of general psychiatry*, *67*(12), 1246-1254.

Rotarska-Jagiela, A., van de Ven, V., Oertel-Knöchel, V., Uhlhaas, P. J., Vogeley, K., & Linden, D. E. (2010). Resting-state functional network correlates of psychotic symptoms in schizophrenia. *Schizophrenia research*, *117*(1), 21-30.

Satterthwaite, T. D., & Baker, J. T. (2015). How can studies of resting-state functional connectivity help us understand psychosis as a disorder of brain development?. *Current opinion in neurobiology*, *30*, 85-91.

Sakagami, M., & Pan, X. (2007). Functional role of the ventrolateral prefrontal cortex in decision making. Current opinion in neurobiology, 17(2), 228-233.

Sakurai, T., Gamo, N. J., Hikida, T., Kim, S. H., Murai, T., Tomoda, T., & Sawa, A. (2015). Converging models of schizophrenia–Network alterations of prefrontal cortex underlying cognitive impairments. *Progress in neurobiology*, *134*, 178-201.

Sarpal, D. K., Robinson, D. G., Lencz, T., Argyelan, M., Ikuta, T., Karlsgodt, K., ... & Malhotra, A. K. (2015). Antipsychotic treatment and functional connectivity of the striatum in first-episode schizophrenia. *JAMA psychiatry*, *72*(1), 5-13.

Schennach, R., Riedel, M., Musil, R., & Möller, H. J. (2012). Treatment response in first-episode schizophrenia. Clinical Psychopharmacology and Neuroscience, 10(2), 78.

Schlagenhauf, F., Rapp, M. A., Huys, Q. J., Beck, A., Wüstenberg, T., Deserno, L., ... & Kienast, T. (2013). Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Human brain mapping*, *34*(6), 1490-1499.

Schott, B. H., Voss, M., Wagnr, B., Wüstenberg, T., Düzel, E., & Behr, J. (2015). Fronto-limbic novelty processing in acute psychosis: disrupted relationship with memory performance and potential implications for delusions. *Frontiers in behavioral neuroscience*, *9*.

Schnell, K., Heekeren, K., Daumann, J., Schnell, T., Schnitker, R., Möller-Hartmann, W., & Gouzoulis-Mayfrank, E. (2008). Correlation of passivity symptoms and dysfunctional visuomotor action monitoring in psychosis. *Brain : a journal of neurology*, *131*(Pt 10), 2783–2797. https://doi.org/10.1093/brain/awn184

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593-1599.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual review of neuroscience*, *23*(1), 473-500.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.*, *57*, 87-115.

Schürhoff, Franck, Andrei Szöke, Alexandre Méary, Frank Bellivier, Frédéric Rouillon, David Pauls, and Marion Leboyer. "Familial aggregation of delusional proneness in schizophrenia and bipolar pedigrees." *American Journal of Psychiatry* 160, no. 7 (2003): 1313-1319.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in neurobiology*, *74*(1), 1-58.

Seeman, P. (2013). Schizophrenia and dopamine receptors. *European Neuropsychopharmacology*, *23*(9), 999-1009.

Serretti, A., & Olgiati, P. (2004). Dimensions of major psychoses: a confirmatory factor analysis of six competing models. Psychiatry research, 127(1-2), 101-109.

Shehzad, Z., Kelly, A. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., ... & Petkova, E. (2009). The resting brain: unconstrained yet reliable. *Cerebral cortex*, *19*(10), 2209-2229.

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *The American journal of psychiatry*, *162*(12), 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Shin, D. J., Lee, T. Y., Jung, W. H., Kim, S. N., Jang, J. H., & Kwon, J. S. (2015). Away from home: the brain of the wandering mind as a model for schizophrenia. *Schizophrenia research*, *165*(1), 83-89.

Shinn, A. K., Baker, J. T., Lewandowski, K. E., Öngür, D., & Cohen, B. M. (2015). Aberrant cerebellar connectivity in motor and association networks in schizophrenia. *Frontiers in human neuroscience*, *9*, 134.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, *106*(31), 13040-13045.

Snyder, S. H., Banerjee, S. P., Yamamura, H. I., & Greenberg, D. (1974). Drugs, neurotransmitters, and schizophrenia. *Science*, *184*(4143), 1243-1253.

So, S. H., Freeman, D., Dunn, G., Kapur, S., Kuipers, E., Bebbington, P., ... & Garety, P. A. (2012). Jumping to conclusions, a lack of belief flexibility and delusional conviction in psychosis: a longitudinal investigation of the structure, frequency, and relatedness of reasoning biases. *Journal of abnormal psychology*, *121*(1), 129.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. Neuroimage, 35(3), 1378-1384.

Sorg, C., Manoliu, A., Neufang, S., Myers, N., Peters, H., Schwerthöffer, D., ... & Förstl, H. (2013). Increased intrinsic brain activity in the striatum reflects symptom dimensions in schizophrenia. *Schizophrenia bulletin*, *39*(2), 387-395.

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self-and external-agency attribution: a brief review and meta-analysis. Brain Structure and Function, 216(2), 151-157

Selten, M., van Bokhoven, H., & Kasri, N. N. (2018). Inhibitory control of the excitatory/inhibitory balance in psychiatric disorders. *F1000Research*, *7*.

Stoodley, C. J., & Schmahmann, J. D. (2010). Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex*, *46*(7), 831-844.

Steel, C., Garety, P. A., Freeman, D., Craig, E., Kuipers, E., Bebbington, P., ... & Dunn, G. (2007). The multidimensional measurement of the positive symptoms of psychosis. *International journal of methods in psychiatric research*, *16*(2), 88-96.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological psychiatry*, *84*(9), 634-643.

Tamminga, C. A., Pearlson, G., Keshavan, M., Sweeney, J., Clementz, B., & Thaker, G. (2014). Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. *Schizophrenia bulletin*, *40*(Suppl 2), S131-S137.

Taylor, S. F. (1996). Cerebral blood flow activation and functional lesions in schizophrenia. *Schizophrenia Research*, *19*(2-3), 129-140.

Tomasi, D., & Volkow, N. D. (2014). Mapping small-world properties through development in the human brain: disruption in schizophrenia. *PloS one*, *9*(4), e96176.

Tost, H., Alam, T., & Meyer-Lindenberg, A. (2010). Dopamine and psychosis: theory, pathomechanisms and intermediate phenotypes. *Neuroscience & Biobehavioral Reviews*, *34*(5), 689-700.

Tritsch, N. X., Oh, W. J., Gu, C., & Sabatini, B. L. (2014). Midbrain dopamine neurons sustain inhibitory transmission using plasma membrane uptake of GABA, not synthesis. *Elife*, *3*, e01936.

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), 1-10.

Turner, D. C., Aitken, M. R., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C., & Fletcher, P. C. (2004). The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cerebral Cortex*, *14*(8), 872-880.

Uhlhaas, P. J., & Mishara, A. L. (2007). Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience. Schizophrenia bulletin, 33(1), 142-156.

Van Os J, Kapur S. Schizophrenia. *Lancet*. 2009;374(9690):635-645. doi:10.1016/S0140-6736(09)60995-8

Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological medicine*, *39*(02), 179-195.

Van Rossum, J. M. (1966). The significance of dopamine-receptor blockade for the mechanism of action of neuroleptic drugs. Archives internationales de pharmacodynamie et de therapie, 160(2), 492.

Verger, A., Horowitz, T., Chawki, M. B., Eusebio, A., Bordonne, M., Azulay, J. P., ... & Guedj, E. (2020). From metabolic connectivity to molecular connectivity: application to dopaminergic pathways. *European Journal of Nuclear Medicine and Molecular Imaging*, *47*(2), 413-424.

Ventura, J., Ered, A., Gretchen-Doorly, D., Subotnik, K. L., Horan, W. P., Hellemann, G. S., & Nuechterlein, K. H. (2015). Theory of mind in the early course of schizophrenia: stability, symptom and neurocognitive correlates, and relationship with functioning. *Psychological medicine*, *45*(10), 2031.

Waltz JA, Gold JM. Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. Schizophrenia research. 2007;93:296–303.

Waltz, J. A., Schweitzer, J. B., Gold, J. M., Kurup, P. K., Ross, T. J., Salmeron, B. J., ... & Stein, E. A. (2009). Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology*, *34*(6), 1567-1577.

Waltz JA, Kasanova Z, Ross TJ, Salmeron BJ, McMahon RP, Gold JM, et al. The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. PloS one. 2013;8:e57257.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43-48.

Weber S, Johnsen E, Kroken RA, et al. Dynamic Functional Connectivity Patterns in Schizophrenia and the Relationship With Hallucinations. *Front Psychiatry*. 2020;11:227. Published 2020 Mar 31. doi:10.3389/fpsyt.2020.00227

Whitfield-Gabrieli, S., Nieto-Castanon, A. (2012). Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity* doi:10.1089/brain.2012.0073.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews neuroscience*, *5*(6), 483-494.

Yung, A. R., & McGorry, P. D. (1996). The prodromal phase of first-episode psychosis: past and current conceptualizations. Schizophrenia bulletin, 22(2), 353-370.

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage*, *200*, 174-190.

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage*, *200*, 12-25.

Zhang, S., Hu, S., Chao, H. H., & Li, C. S. R. (2016). Resting-state functional connectivity of the locus coeruleus in humans: in comparison with the ventral tegmental area/substantia nigra pars compacta and the effects of age. *Cerebral Cortex*, *26*(8), 3413-3427.

Zhang, J., Hughes, L. E., & Rowe, J. B. (2012). Selection and inhibition mechanisms for human voluntary action decisions. *Neuroimage*, *63*(1), 392-402.

Zhang, B., Lin, P., Wang, X., Ongur, D., Ji, X., Situ, W., ... & Wang, X. (2019). Altered Functional Connectivity of Striatum Based on the Integrated Connectivity Model in First-Episode Schizophrenia. *Frontiers in psychiatry*, *10*, 756.

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., & Berns, G. S. (2003). Human striatal response to salient nonrewarding stimuli. *Journal of Neuroscience*, *23*(22), 8092-8097.

Appendices

Appendix A: Prediction Error Task Activation and Functional Connectivity in Psychosis and its Association with Delusion Symptoms

Table A-1. Demographic Differences between Included and Excluded Subjects

	Included PTS	Excluded PTS	P-value	Included HC	Excluded HC	P-value
	(sd)	***(sd)***		***(sd)***	***(sd)***	
N	47	7		15	5	
Male/Female	21/26	3/4	.928	6/9	3/2	.317
Avg. Age (y)	38.8 (11.4)	44.1 (11.7)	.252	36.9 (13.6)	29.8 (9.7)	.625
Avg. Daily CPZ	280.7 (319.8)	458.9 (305.8)	.259			
GAF	54.8 (13.0)	45.8 (8.0)	.109	81.4 (4.0)	82.6 (2.5)	.538
PANSS Delusion (P1)	2.8 (1.8)	3.3 (1.2)	.464			
PANSS Positive	15.6 (6.4)	21.5 (2.8)	.032			
PANSS Negative	12.6 (5.0)	13.0 (2.8)	.840			
PANSS General	28.5 (8.6)	30.5 (3.0)	.579			
PANSS Total	56.7 (17.0)	65.0 (3.6)	.246			
LDPS Delusions (P1)	2.80 (1.1)	3.0 (0.9)	.682			
PDI Total	106.7 (70.7)	126.0 (95.9)	.525	52.4 (43.5)	45.0 (53.6)	.759
PDI Endorse	10.0 (5.4)	9.6 (6.3)	.841	5.4 (4.0)	4.6 (5.0)	.721
PDI Distress	28.3 (20.3)	36.3 (27.4)	.360	12.5 (10.5)	6.8 (7.8)	.287
PDI Preoccupation	30.3 (20.8)	37.4 (30.2)	.434	14.4 (12.2)	15.2 (19.0)	.936
PDI Conviction	33.8 (23.7	40.3 (30.4)	.522	17.2 (14.8)	15.0 (18.4)	.756

Abbreviations: Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, PANSS -Positive and Negative Symptom Scale.

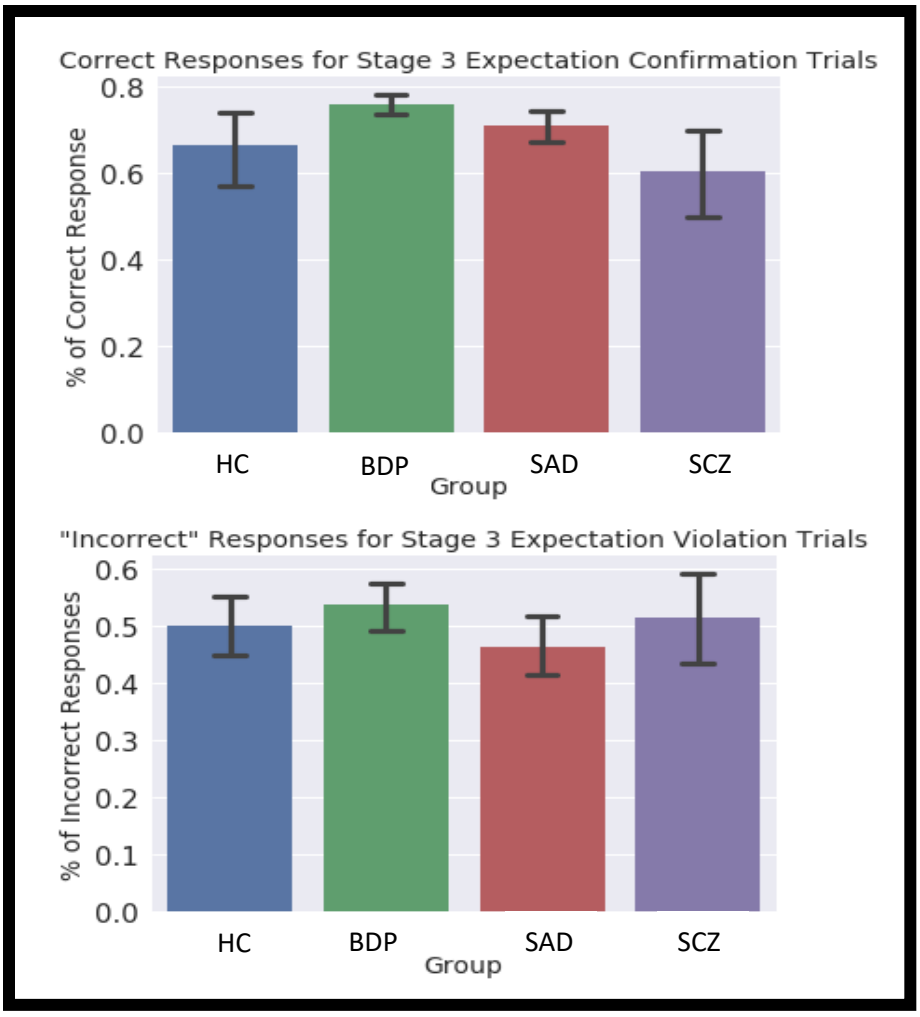

Figure A-1. Accuracy in Stage 3 of Prediction Error Task Across Participant Subgroups

A) Correct Responses during Expectation Confirmation Trials. B) Appropriately "Incorrect" Response during Expectation Violation Trials. Abbreviations: BDP – Bipolar Disorder with Pyschosis, HC- Healthy Control, SAD – Schizoaffective Disorder, SCZ - Schizophrenia

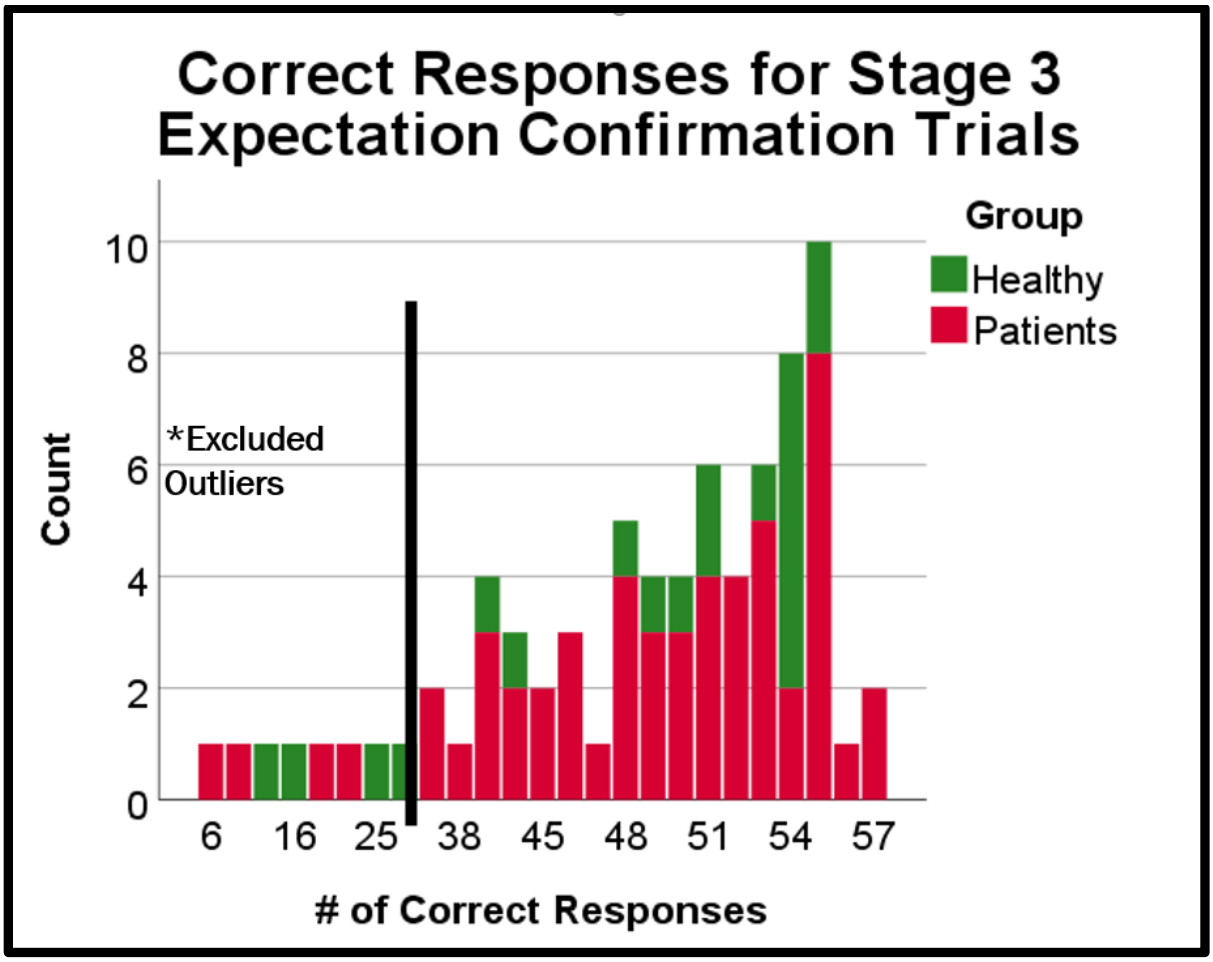

Figure A-2. Histogram of Correct Responses to Expectation Confirmation Trials

Subjects with low task performance shown as outliers above were excluded from the analysis. After behavioral data exclusions remaining subjects were 47 Psychotic Patients – (13) Schizophrenia, (16) Psychotic Bipolar Disorder, (18) Schizoaffective Disorder – and 16 Healthy Controls

Appendix B: Prediction Error Resting State Functional Connectivity and its Association with Delusion Symptoms

Table B-1. Scanning Parameters across BSNIP-1 Study Sites

fMRI	TR (ms)	TE (ms)	Flip angle (degree)	Slices (N)	Matrix (mm)	Voxel Size (mm)	Vendor
Baltimore	2210	30	70	36	64x64	3.4x3.4x3	Siemens TrioTim
Boston	3000	27	60	30	64x64	3.4x3.4x5	GE Signa HDX
Chicago	1775	27	60	29	64x64	3.4x3.4x4	GE Signa HDX
Dallas	1500	27	60	29	64x64	3.4x3.4x4	Philips
Detroit	1570	22	60	29	64x64	3.4x3.4x4	Siemens TrioTim
Hartford	1500	27	70	29	64x64	3.4x3.4x5	Siemens Allegra
sMRI	**TR (ms)**	**TE (ms)**	**Flip angle (degree)**	**Slices (N)**	**Matrix (mm)**	**Voxel Size (mm)**	**Vendor**
Baltimore	2300	2.91	9	160	256x240	1x1x1.2	Siemens TrioTim
Boston	6.98	2.84	8	166	256x256	1x1x1.2	GE Signa HDX
Chicago	6.98	2.84	8	166	256x256	1x1x1.2	GE Signa HDX
Detroit	2300	2.94	9	160	256x240	1x1x1.2	Siemens TrioTim
Dallas	6.6	2.8	8	170	256x256	1x1x1.2	Philips
Hartford	2300	2.91	9	160	256x240	1x1x1.2	Siemens Allegra

.

Table B-2. Demographics and Clinical Characteristics for Included and Excluded Subjects

	Included Patients	Excluded Patients	P-value	Included Healthy	Excluded Healthy	P-value
N	338	39		186	27	
Male/Female	160/178	23 /16	.023	72/114	11/16	.499
Avg. Age (year)	35.9 (12.2)	34.0 (14.3)	.297	37.8 (12.4)	46.7 (9.5)	<.001
Avg. Daily CPZ	391.9 (382.2)	450.1 (365.0)	.375			
PANSS Delusion	2.7 (1.4)	2.5 (1.7)	.461			
PANSS Positive	15.9 (5.3)	15.5 (6.8)	.608			
PANSS Negative	14.6 (5.1)	15.4 (6.2)	.303			
PANSS General	31.9 (8.6)	28.2 (9.1)	.006			
PANSS Total	62.4 (16.5)	58.9 (18.9)	.164			
GAF	52.1 (13.3)	52.6 (14.2)	.790	85.9 (6.8)	81.7 (7.9)	.005
mFDpower (mm)	0.21 (0.13)	0.20 (0.12)	.470	0.17 (0.10)	0.14 (0.09)	.193

Abbreviations: Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, GAF – Global Assessment of Function, PANSS -Positive and Negative Syndrome Scale. Values in parenthesis are standard deviations.

Table B-3. Demographics and Clinical Characteristics for Included and Excluded subjects (Participants with Medication Information)

	Included Patients	Excluded Patients	P-value	Included Healthy	Excluded Healthy	P-value
N	243	39		186	27	
Male/Female	108/135	23 /16	.065	72/114	11/16	.499
Avg. Age (year)	35.4 (11.9)	35.0 (14.0)	.862	37.8 (12.4)	46.7 (9.5)	<.001
Avg. Daily CPZ	385.7 (370.6)	450.1 (365.0)	.314			
PANSS Delusion	2.6 (1.6)	2.5 (1.7)	.651			
PANSS Positive	16.0 (5.2)	15.6 (6.8)	.689			
PANSS Negative	14.6 (5.2)	14.8 (5.3)	.843			
PANSS General	32.3 (8.9)	27.4 (7.8)	.002			
PANSS Total	62.9 (16.6)	57.5 (16.1)	.061			
GAF	51.8 (13.3)	50.8 (13.6)	.667	85.9 (6.8)	81.7 (7.9)	.005
mFDpower (mm)	0.21 (0.13)	0.19 (0.12)	.361	0.17 (0.10)	0.14 (0.09)	.193

Abbreviations: Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, GAF – Global Assessment of Function, PANSS -Positive and Negative Syndrome Scale. Values in parenthesis are standard deviations.

Table B-4. Demographics and Clinical Characteristics for Included Subjects

	SCZ	SAD	BPD	HC	P-value	Post-hoc
N	122	101	115	186		
Male/Female	79/43	44/57	37/78	72/114	<.001	*a, d, e
Avg. Age (year)	34.5 (12.0)	37.8 (12.0)	36.7 (12.5)	37.8 (12.4)	.070	
PANSS Delusion	3.1 (1.5)	3.2 (1.1)	1.9 (1.0)		<.001	*d, f
PANSS Positive	16.8 (5.5)	18.1 (4.8)	13.0 (4.2)		<.001	*d, f
PANSS Negative	16.3 (5.9)	15.3 (4.6)	12.2 (3.7)		<.001	*d, f
PANSS General	31.9 (9.0)	35.2 (8.9)	29.1 (7.7)		.009	*e, f
PANSS Total	65.0 (17.2)	68.6 (15.8)	54.4 (12.8)		<.001	*d, f
GAF	48.8 (12.3)	47.2 (11.4)	59.9 (12.3)	85.9 (6.8)	<.001	* a, b, c, d, f
mFDpower (mm)	0.20 (0.13)	.025 (0.14)	.020 (0.11)	0.17 (0.10)	<.001	*b, e, f

Abbreviations: BPD – Bipolar Disorder w/ Psychosis, CPZ – Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, GAF – Global Assessment of Function, HC – Healthy Control, PANSS -Positive and Negative Syndrome Scale, SAD – Schizoaffective Disorder, SCZ – Schizophrenia
Post-Hoc (Bonferroni-corrected): SCZ vs HC[a]; SAD vs HC[b]; BPD vs HC[c]; SCZ vs BPD[d]; SCZ vs SAD[e]; SAD vs BPD[f]. Values in parenthesis are standard deviations.

Table B-5. Demographics and Clinical Characteristics for Included Subjects (Participants with Medication Information)

	SCZ	SAD	BDP	HC	P-value	Post-hoc
N	89	73	81	186		
Male/Female	32/57	29/44	22/59	72/114	<.001	*a, d, e
Avg. Age (year)	33.5 (11.4)	37.6 (11.7)	35.5 (12.4)	37.8 (12.4)	.031	*a
Avg. Daily CPZ	438.0 (375.5)	451.5 (387.9)	269.0 (322.9)		.002	*d, f
PANSS Delusions	3.1 (1.5)	3.2 (1.1)	1.9 (1.0)		<.001	*d, f
PANSS Positive	16.9 (5.5)	18.4 (4.3)	12.9 (4.1)		<.001	*d, f
PANSS Negative	15.8 (6.1)	15.9 (4.4)	12.2 (3.7)		<.001	*d, f
PANSS General	31.7 (9.0)	36.2 (8.6)	29.2 (7.7)		<.001	*d, f
PANSS Total	64.4 (17.9)	70.6 (14.8)	54.4 (12.4)		<.001	*d, e, f
GAF	49.1 (12.7)	46.7 (10.4)	59.4 (13.0)	85.9 (6.8)	<.001	* a, b, c, d, f
mFDpower (mm)	0.18 (0.11)	0.25 (0.15)	0.60 (0.41)	0.17 (0.10)	<.001	*b, d, f

Abbreviations: BDP – Bipolar Disorder w/ Psychosis, CPZ – Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, GAF – Global Assessment of Function, HC – Healthy Control, PANSS -Positive and Negative Syndrome Scale, SAD – Schizoaffective Disorder, SCZ – Schizophrenia
Post-Hoc (Bonferroni-corrected): SCZ vs HC[a]; SAD vs HC[b]; BDP vs HC[c]; SCZ vs BDP[d]; SCZ vs SAD[e]; SAD vs BDP[f]. Values in parenthesis are standard deviations.

Table B-6. Significant D-PE Connectivity found in the Combined group (top), and Differences (bottom) between Healthy Controls and Patients with Medication Information

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE	Type of Connectivity (+/-)
Healthy Controls and Patients with Medication Information					
R DLPFC	3708	00, 50, -14	Frontal Pole- Medial Prefrontal Cortex – Superior Frontal Gyrus – Paracingulate Gyrus – Anterior Cingulate Cortex – Subcallosal Cortex	<.000001	-
	2504	00, -52, 16	Precuneus-Posterior Cingulate Cortex- L Hippocampus- Para-Hippocampal Cortex- posterior Temporal Fusiform Cortex	<.000001	-
	2174	34, 34, 26	R Frontal Pole – Middle Frontal Gyrus	<.000001	+
	1615	14, 10, 64	R Suprior Frontal Gyrus – Paracingulate Gyrus – Anterior Cingulate Cortex - Supplementary Mortor Area	<.000001	+
	1282	50, 16, -08	R Insula – Inferior Frontal Gyrus - Frontal Operculum -Temporal Pole -Frontal Orbital Cortex	<.000001	+
	989	-38, 38, 32	L Frontal Pole – Middle Frontal Gyrus	<.000001	+
	773	-42, 26, -28	L Temporal Pole – Frontal Orbital Cortex	.000002	-
	738	62, -36, 40	R Supramarginal Gyrus – Angular Gyrus	.000003	+
	406	-58, -40, 50	L Supramarginal Gyrus	.000457	+
	375	06, 08, -04	Thalamus- Nucleus Accumbens	.000787	-
	335	18, -10, 36	Posterior Cingulate Cortex	.001622	+
	333	24, -20, -22	L Hippocampus – Para Hippocampal Cortex	.001683	-
	258	22, 48, -16	R Frontal Pole	.007110	+
	254	-40, -64, 28	L Superior Lateral Occipital	.007705	-

Table B-6. Significant D-PE Connectivity found in the Combined group (top), and Differences (bottom) between Healthy Controls and Patients with Medication Information, Continued

R VLPFC	5800	54, 18 24	R Frontal Pole – Middle Frontal Gyrus – Inferior Frontal Gyrus – Frontal Operculum – Frontal Orbital Cortex - Insula – Precentral Gyrus	<.000001	+
	3204	-06, 46, -06	Frontal Pole – Anterior Cingulate Cortex – Paracingulate Gyrus – Subcallosal Gyrus – Medial Prefrontal Cortex – Caudate – Nucleus Accumbens	<.000001	-
	2671	-02, 48, 18	Precuneus – Posterior Cingulate Cortex – Intercalcarine Cuneal Cortex – Cuneal Cortex – Lingual Gyrus	<.000001	-
	1957	38, -46, 48	R Supramarginal Gyrus – Superior Posterior Lateral Cortex – Angular Gyrus – Superior Lateral Occipital Cortex – Posterior Central Gyrus	<.000001	+
	1279	-48, 12, 26	L Middle Frontal Gyrus – Frontal Pole – Inferior Frontal Gyrus – Precentral Gyrus	<.000001	+
	757	-50, -36, 44	L Superior Posterior Lateral Cortex – Supramarginal Gyrus	<.000001	+
	614	58, -46, -08	R Inferior Temporal Gyrus - Middle Temporal Gyrus	.000004	+
	405	-16, 36, 58	L Frontal Pole – Superior Frontal Gyrus	.000698	-
	381	06, 28, 48	R Superior Frontal Gyrus- Paracingulate Gyrus	.001042	+
	176	-24, -74, -50	L Cerebellum lobules 7b/2/1/8	.049252	+

Table B-6. Significant D-PE Connectivity found in the Combined group (top), and Differences (bottom) between Healthy Controls and Patients with Medication Information, Continued

L Caudate	5029	-14, 14, 04	Putamen – Caudate – Thalamus – Pallidum – Nucleus Accumbens – Amygdala - L Hippocampus	<.000001	+
	176	-08, -86, 32	Cuneal – Precuneus Cortex	.038687	-
R Caudate	5656	16, 14, 04	Putamen – Caudate – Thalamus – Pallidum – Accumbens – Amygdala – R Insula Cortex – R Frontal Operculum	< .000001	+
	2992	00, -76, 32	Precuneus – Posterior Cingulate Cortex – Cuneal Cortex	<.000001	-
	382	04, 16, 56	Superior Frontal Gyrus	.000702	+
	228	24, -36, -06	L Frontal Operculum – Central Operculum- Insula Cortex	.013198	+
	188	26, -42, -04	R Hippocampus – posterior Para Hippocampal Cortex	.031037	-
L Midbrain	4333	-10, -24, -10	Brainstem – Hippocampus – Thalamus – Amygdala - posterior Para – Hippocampal Cortex – Cerebellum	<.000001	+
	691	-04, 32, 04	Medial Prefrontal Cortex – Frontal Pole – Anterior Cingulate Cortex – Paracingulate Gyrus	.000003	+
Healthy Controls vs Patients with Medication Information					
R DLPFC	286	08, -38, 72	R Postcentral Gyrus – Precentral Gyrus	.004094	-
R Caudate	253	-08, -40, 00	Posterior Cingulate Cortex – Cerebellum	.007912	-

Abbreviations: BPD – Bipolar Disorder, DLPFC – Dorsolateral Prefrontal Cortex, HC – Healthy Controls, L – Left, FWE – p-value Family Wise Error, R – Right, SCZ – Schizophrenia, SAD – Schizoaffective Disorder, VLPFC – Ventrolateral Prefrontal Cortex

Table B-7. D-PE Connectivity Associated with Delusion Severity (Adjusted for Medication)

	Cluster Size ($8mm^3$ voxels)	Peak Voxel x, y, z	Location of Cluster	pFWE	Type of Connectivity (+/-)
Connectivity Associated with Delusion Severity in Patients (Diagnosis Interaction with Medication Covariate)					
R VLPFC	353	-02, -14, 74	L Precentral Gyrus	.015218	-
SCZ vs BDP	423	00, -16, 74	Precentral Gyrus	.000456	-
R DLPFC	441	-10, -34, -16	Brainstem – Cerebellum – Para Hippocampal Cortex	.000117	-
SCZ vs BDP	715	-12, -32, -14	Brainstem – Cerebellum – Para Hippocampal Cortex	.000003	-
SAD vs BDP	280	-12, -32, -12	Brainstem – Cerebellum –	.004636	-

Abbreviations: BPD – Bipolar Disorder, DLPFC – Dorsolateral Prefrontal Cortex, L – Left , FWE – p-value Family Wise Error, R – Right, SCZ – Schizophrenia, SAD – Schizoaffective Disorder, VLPFC – Ventrolateral Prefrontal Cortex

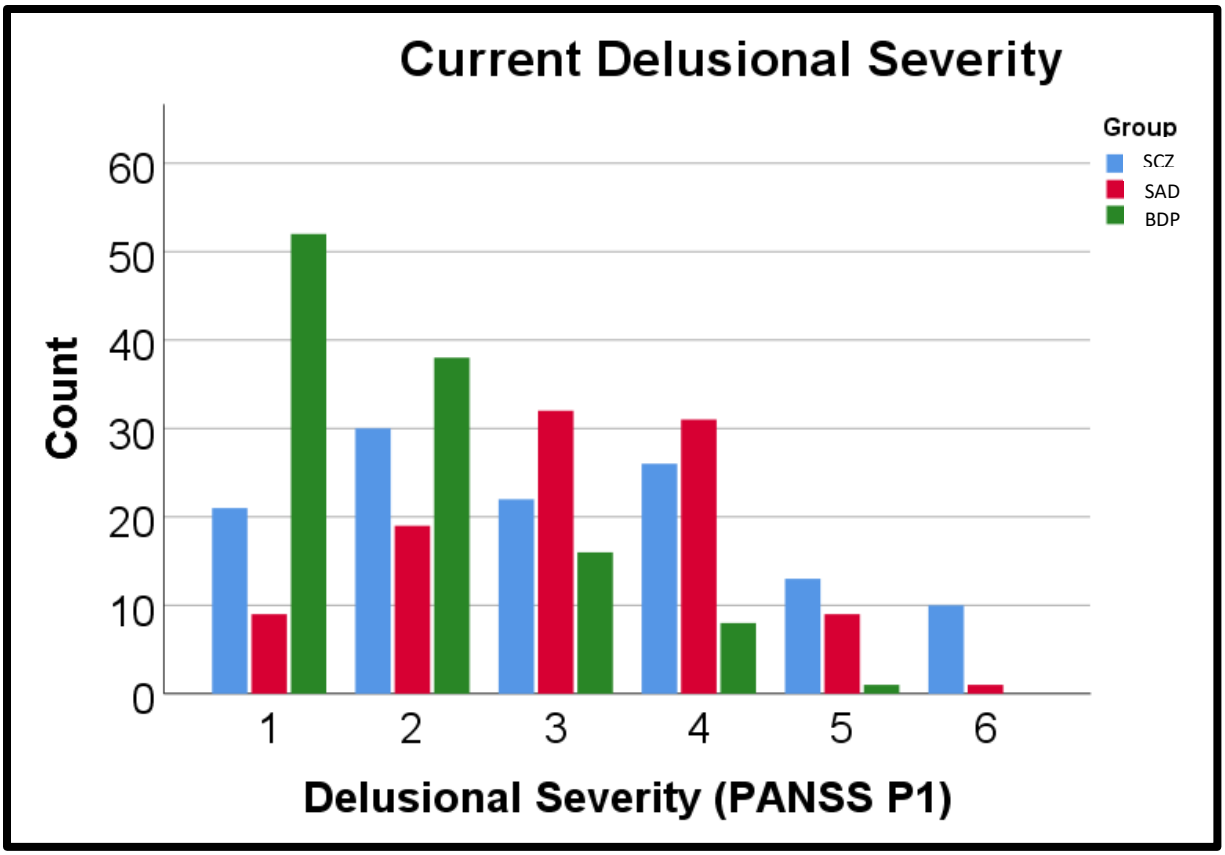

Figure B-1. Distribution of Delusion Severity across Diagnoses

Current delusion severity was measured by P1 measure on the Positive and Negative Syndrome Scale (PANSS). Patients were represented across the severity spectrum with delusion severity ranging from no present symptoms (1) to severe delusion symptoms (6). Abbreviations: BDP – Bipolar Disorder with Pyschosis, SAD – Schizoaffective Disorder, SCZ - Schizophrenia

In analyses of average daily CPZ equivalent dosage information, delusion severity was minimally correlated with dose for the whole sample, with variation among the diagnostic subgroups (Spearman's rho = 0.149, p<.020 two tailed for whole sample; Subgroups correlations were: BDP - 0.215, p=.054; SADP - -0.002, p=.984; SZP - 0.210, p=.047.).

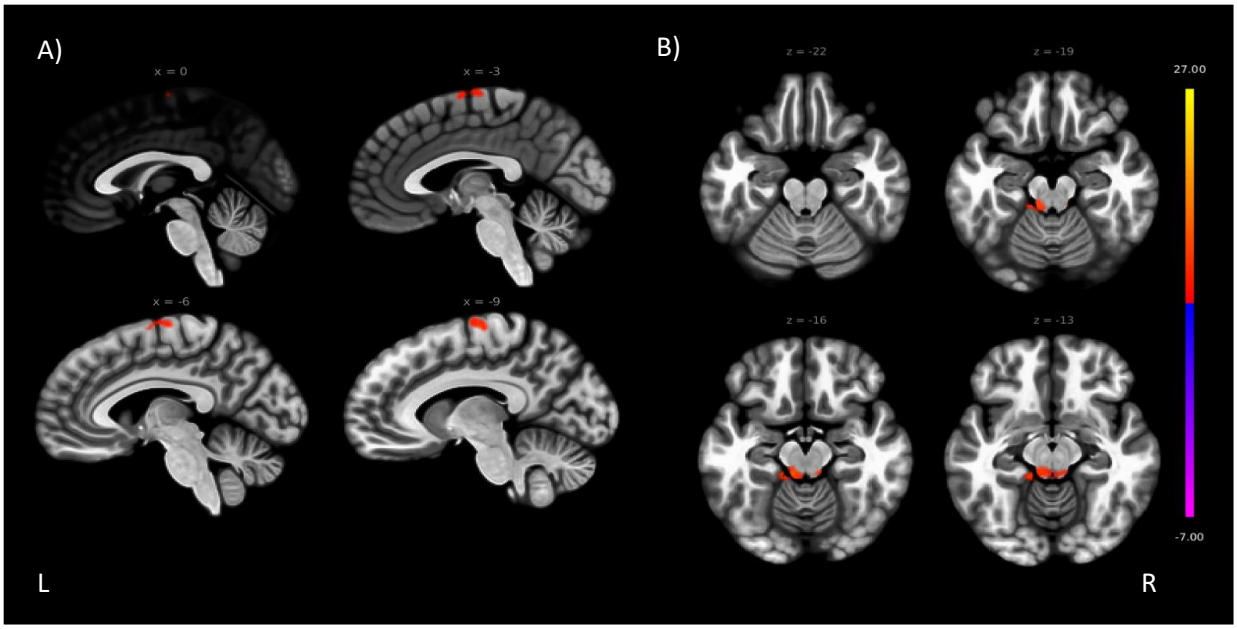

Figure B-2. Exploratory Interaction of Diagnosis with D-PE Connectivity and Patient Delusion Severity

A) A significant diagnosis interaction was observed in the right ventrolateral prefrontal cortex with the cerebellum. Decreased connectivity within schizophrenia subjects was associated with increased delusion symptoms. This result was trending significance when controlling for medication. B) A trending significant diagnosis interaction was also observed in the right dorsolateral prefrontal cortex to the midbrain. Decreased connectivity within schizophrenia and schizoaffective subjects was associated with increased delusion symptoms. This result was significant when controlling for medication. Results overlaid upon an MNI-152 T1 template shown as significant z-statistics. Regions of positive connectivity show greater significance from red-to-yellow and regions of negative connectivity show greater significance from blue-to-purple.

Appendix C Prediction Error Related Effective Connectivity in Psychosis and its Association with Delusion Symptoms

Table C-1. Demographic Differences between Included and Excluded Subjects

	Included Patients *(sd)*	**Excluded Patients** *(sd)*	**P-value**	**Included Healthy** *(sd)*	**Excluded Healthy** *(sd)*	**P-value**
N	324	67		182	31	
Male/Female	155/169	39 /28	.042	70/112	13/18	.133
Avg. Age (y)	35.8 (12.1)	35.1 (14.4)	.862	37.8 (12.3)	45.7 (10.5)	<.001
Avg. Daily CPZ	389.1 (371.8)	423.9 (362.5)	.314			
PANSS Delusion (P1)	2.7 (1.4)	2.6 (1.6)	.651			
PANSS Positive	15.9 (5.4)	15.6 (6.2)	.689			
PANSS Negative	14.7 (5.1)	15.1 (6.0)	.843			
PANSS General	31.9 (8.9)	29.2 (9.2)	.002			
PANSS Total	60.7 (16.5)	59.8 (18.3)	.061			
mFDpower (mm)	0.21 (0.13)	0.22 (0.13)	.361	0.17 (0.10)	0.15 (0.09)	.193

Abbreviations: CPZ – Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, PANSS -Positive and Negative Syndrome Scale

Table C-2. Demographics and Clinical Characteristics for Included Subjects (Participants with Medication Information)

	Patients	Healthy Controls	P-value
	(sd)	***(sd)***	
N	237	182	
Male/Female	108/129	70/112	0.145
Avg. Age (y)	35.3 (11.8)	37.8 (12.3)	0.039
Avg. Daily CPZ	389.1 (371.8)		
PANSS Delusion (P1)	2.7 (1.4)		
PANSS Positive	16.0 (5.3)		
PANSS Negative	14.6 (5.2)		
PANSS General	32.1 (8.9)		
PANSS Total	62.7 (16.6)		
mFDpower (mm)	0.21 (0.13)	0.17 (0.10)	0.001

Abbreviations: CPZ – Chlorpromazine Equivalents, mFDpower – mean Framewise Displacement power, PANSS -Positive and Negative Syndrome Scale

Table C-3. Summary of Effective Connectivity Results in Association with Patient Status (Participants with Medication Information)

Connection →	Parameter Estimate (typical is 0.1 Hz)	Connection Valence: + excitatory - inhibitory	Connectivity: ↑Increased ↓Decreased
Patient Status Association			
rCaud to rCaud	-0.172		↓
rVLPFC to rCaud	-0.117		↓
rVLPFC to rVLPFC	-0.265		↓
rDLPFC to rDLPFC	-0.133		↓
Midbr to rDLPFC	-0.128		↓
Group Mean Effective Connectivity			
rCaud to lCaud	0.172	+	
lCaud to rCaud	0.213	+	
lCaud to lCaud	-0.252	-	
rDLPFC to rCaud	-0.083	-	
rVLPFC to rVLPFC	-0.342	-	
rDLPFC to rDLPFC	-0.255	-	
Midbr to Midbr	-0.207	-	

N= 419 Subjects – 182 Healthy Controls and 237 Psychosis Patients on Medication

Table C-4. Summary of Effective Connectivity Associated with Delusion Severity

(Patients with Medication Information)

Connection →	Parameter Estimate (typical is 0.1 Hz)	Connection Valence: + excitatory - inhibitory	Connectivity: ↑Increased ↓Decreased
Delusion Association			
lCaud to rDLPFC	0.036		↑
rVLPFC to rCaud	-0.033		↓
rVLPFC to rVLPFC	-0.064		↓
rDLPFC to rDLPFC	-0.091		↓
Midbr to rCaud	-0.029		↓
Group Mean Effective Connectivity			
rCaud to rCaud	-0.109	-	
rCaud to lCaud	0.163	+	
lCaud to rCaud	0.163	+	
lCaud to lCaud	-0.236	-	
lCaud to rVLPFC	-0.137	-	
rVLPFC to rVLPFC	-0.273	-	
rDLPFC to rCaud	0.105	+	
rDLPFC to rDLPFC	-0.181	-	
Midbr to lCaud	-0.179	-	
Midbr to Midbr	-0.209	-	

N= 237 Psychosis Patients on Medication

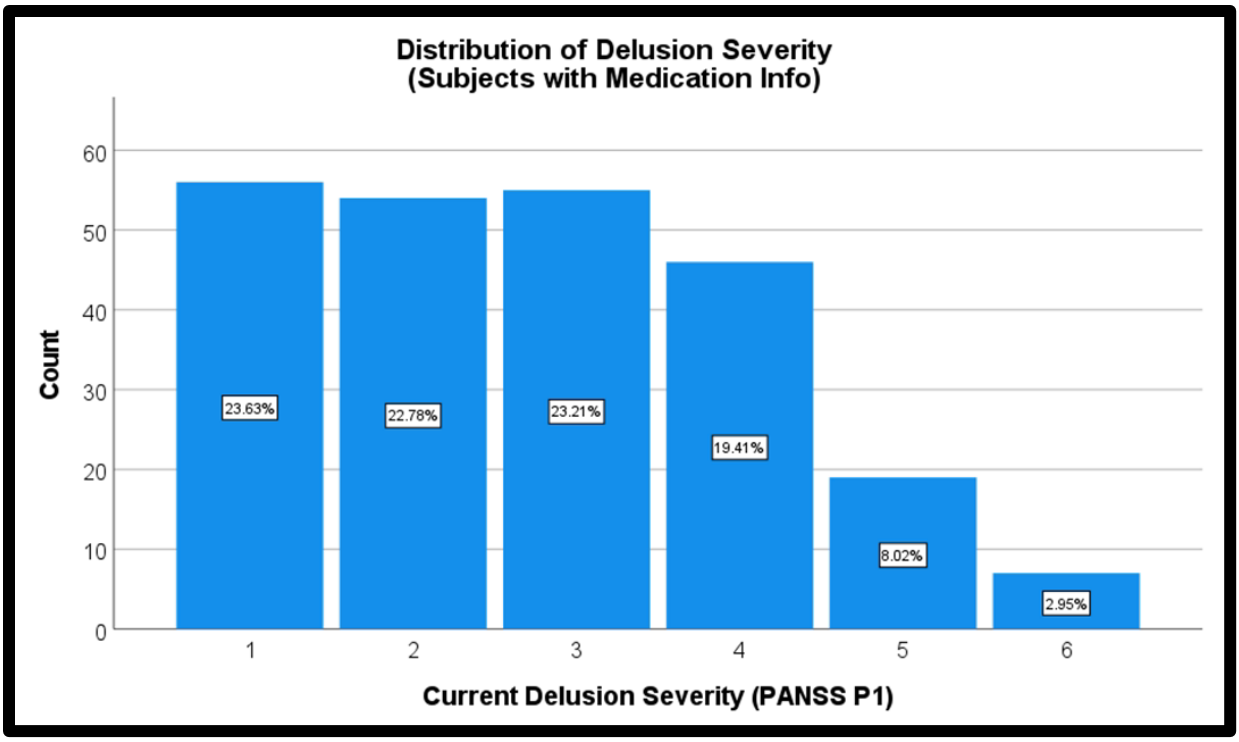

Figure C-1. Distribution of Delusion Severity in Patients with Medication Information

Current delusion severity was measured by P1 measure on the Positive and Negative Symptom Scale (PANSS). Patients were represented across the severity spectrum with delusion severity ranging from no present symptoms (1) to severe symptoms (6).

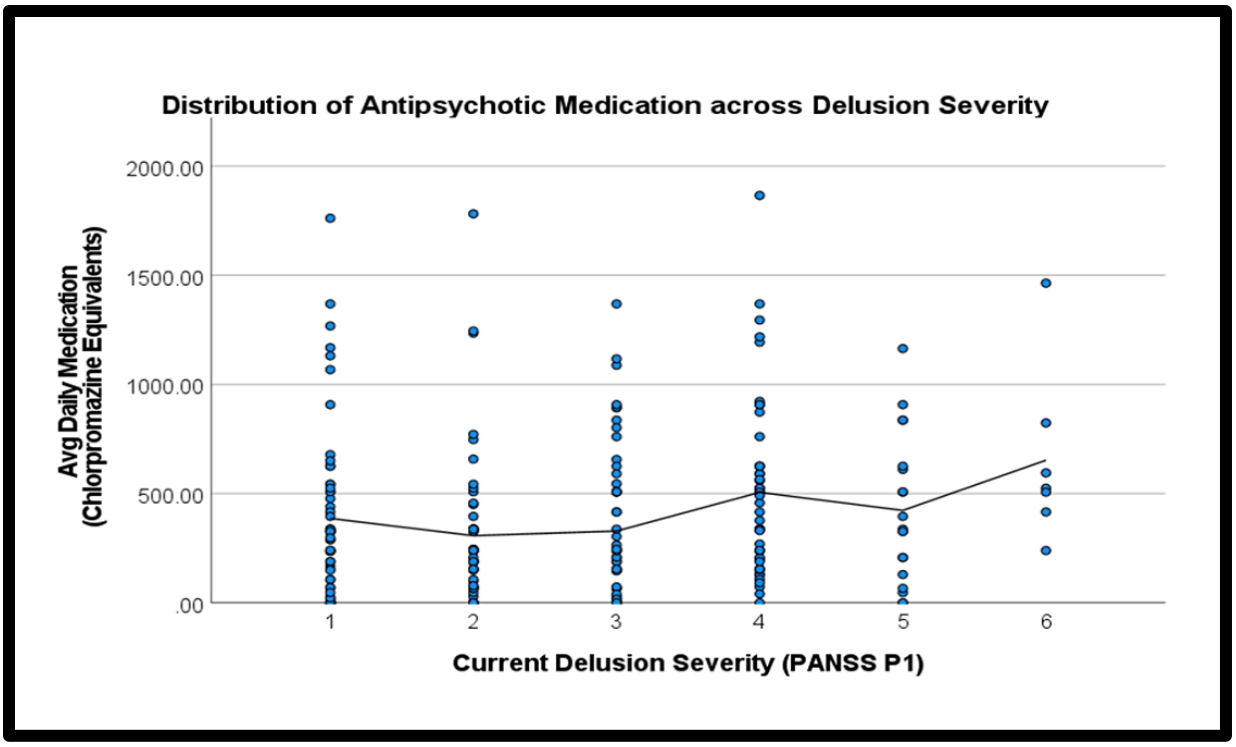

Figure C-2. Distribution of Antipsychotic Medication Usage across Delusion Severity

A small correlation was observed between patient's average daily antipsychotic usage and delusion severity.

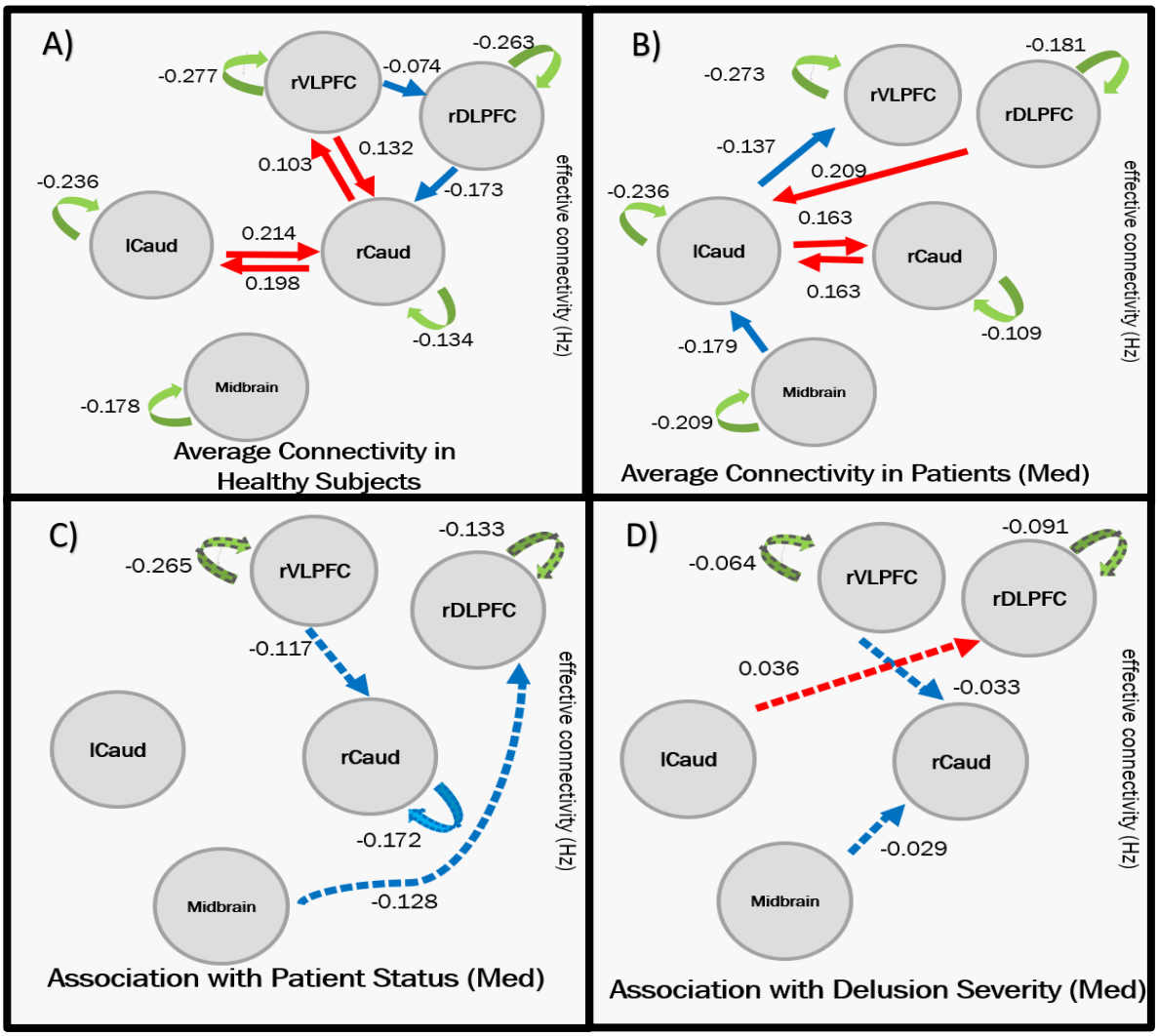

Figure C-3. Effective Connectivity Strengths in Medication Adjusted Analysis

Between node connections represent the rate of change in neural response in one region due to activity from the connected regions measured in hertz (Hz). Connections are depicted as red arrows/positive values are excitatory and blue arrows/negative values are inhibitory. Self-connections represent self-inhibition in each region or sensitivity to input measured in unitless log scaling parameter (multiplying the default value of -0.5Hz). Connections are depicted as green arrows starting and ending at the same node - negative values imply disinhibition (positive would indicate stronger self-inhibition, though none were positive in this analysis).
A) Average Effective Connectivity in Healthy Subjects
B) Average Effective Connectivity in Psychosis Subjects
C) Changes in Effectivity Connectivity associated with Patient Status (dashed connections) relative to mean effectivity connectivity across both patients and healthy controls (bold connections)
D) Changes in Effectivity Connectivity associated with Delusion Severity (dashed connections) relative to mean effectivity connectivity across patients (bold connections)