

Functional Magnetic Resonance Imaging of Goal Maintenance in Schizophrenia:
Activation, Functional Connectivity, and Reliability

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Andrew Bogdan Poppe

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Angus W. MacDonald (Advisor)

May, 2016

Acknowledgements

I would like to thank my advisor, Angus MacDonald, III. Your constant enthusiasm, patience, curiosity, wisdom, and fearlessness were inspirations to me as a scientist-in-training; however, your skepticism and methodological rigor were just as important as they taught me to be a careful, questioning scientist. You have been dedicated and unselfish as an advisor, and I hope I can live up to the example you have set. I would also like to thank all of my friends and colleagues in the TRiCAM Laboratory and the Psychology department. I probably could have gotten through this without you, but it would not have been nearly as fun. Most of all, I'd like to thank my mom and dad for always being there for me and for supporting my decision to pursue a Ph.D. I could never have done any of this without your love, support, and sacrifices my entire life. I love you both.

Abstract

Cognitive deficits are some of the most debilitating and difficult to treat symptoms of schizophrenia. Goal maintenance is a facet of cognitive control that has been shown to be impaired in schizophrenia patients as well as their unaffected first-degree relatives.

Previous fMRI activation studies found less activation in dorsolateral prefrontal cortex (dlPFC) in schizophrenia patients compared with healthy controls during the completion of a goal maintenance task. This dissertation consisted of a series of studies employing a large, multisite retest dataset of schizophrenia patients and healthy control subjects.

These studies sought to replicate previous activation findings using a newer goal maintenance task, to use group independent component analysis (ICA) to determine if schizophrenia patients also exhibited dysfunctional functional connectivity or functional network connectivity (FNC) compared with healthy controls during the performance of that task, and to evaluate the test-retest reliability of each of these metrics, directly compare them, and assess the influence of subject group and data collection site on reliability estimates. It replicated previous activation study findings of reduced dlPFC activity during goal maintenance. It additionally found that the temporal association between a frontoparietal executive control network and a salience network was stronger in healthy controls than in schizophrenia patients and that the strength of this relationship predicted performance on the goal maintenance task. It also found that the task-modulation of the relationship between left- and right-lateralized executive control networks was stronger in healthy controls than in schizophrenia patients and that the

strength of this task-modulation predicted goal maintenance task performance in healthy controls. Finally, reliability estimates found that ICA and tonic FNC had acceptable overall reliability and that they minimized site-related variance in reliability compared with dynamic FNC and general linear model. These results indicate that ICA and tonic FNC may provide better tools for group contrast fMRI studies examining schizophrenia, especially those that incorporate a multisite design.

Table of Contents

List of Tables	v
List of Figures	vi
Chapter 1	1
Chapter 2	37
Chapter 3	59
Chapter 4	83
Chapter 5	120
Bibliography	126
Appendix 1	151
Appendix 2	164

List of Tables

Table 2.1 Demographic and Clinical Characteristics.....	53
Table 3.1. Components Significantly Correlated with Time Course of DPX B Cues Across Groups.....	76
Table 3.2. Component Connections with Groups Differences in Tonic Functional Network Connectivity.....	77
Table 4.1. Subject Demographics.....	106
Table 4.2. Reliability of Analysis Methods.....	107
Table 4.3. tFNC Reliability by Magnitude of tFNC.....	108
Supplementary Table 1.1. Clusters of Significant Activation in fMRI Analysis.....	156
Supplementary Table 2.1. Sample Demographics.....	167

List of Figures

Figure 1.1. Mixing of Variables Increases Normality.....	35
Figure 1.2. Orthogonality Versus Independence.....	36
Figure 2.1. DPX Behavioral Results.....	54
Figure 2.2. Behavioral Efficiency Analysis.....	55
Figure 2.3. Whole brain fMRI GLM results.....	56
Figure 2.4. fMRI Efficiency Analysis, B Cues – A Cues.....	57
Figure 3.1. Functional network connectivity during the DPX task in SZ and HC.....	78
Figure 3.2. Group differences in tonic and dynamic functional network connectivity....	80
Figure 3.3. Functional Network Connectivity Predicts Task Performance.....	82
Figure 4.1. GLM Activation Spatial Maps.....	109
Figure 4.2. Average Reliability of Analysis Methods.....	110
Figure 4.3. Average Reliability Separated by CNTRACS Site.....	111
Figure 4.4. Spatial Overlap of ICA Results.....	112
Figure 4.5. Spatial Maps of ICA and GLM Reliability.....	113
Figure 4.6. Relationship between ICA Results and DPX B Cue Time Course.....	115
Figure 4.7. Average Reliability of ICA Separated by Component Type.....	117
Figure 4.8. Relationship between tFNC Magnitude and Reliability.....	118
Supplementary Figure 1.1. fMRI Efficiency Analysis.....	159
Supplementary Figure 1.2. fMRI Efficiency Analysis, B Cues – A Cues.....	160

	vii
Supplementary Figure 1.3. fMRI Efficiency Analysis.....	162
Supplementary Figure 1.4. Confirmatory fMRI GLM results.....	163
Supplementary Figure 2.1. Demonstration of dynamic functional network Connectivity.....	168
Supplementary Figure 2.2. Task-related component maps.....	170
Supplementary Figure 2.3. Task modulation of dynamic functional network connectivity between ECN components.....	172
Supplementary Figure 2.4. DPX Performance Predicted by dFNC in Schizophrenia Patients.....	174

Chapter 1: Background on Functional Connectivity, Reliability, and Goal Maintenance

This chapter is intended to review background information about functional connectivity measures of functional magnetic resonance (fMRI) data, reliability theory, reliability considerations of fMRI methods, and how these concepts apply to studies of schizophrenia patients (SZ). It will review the disconnectivity hypothesis of schizophrenia and how functional connectivity methods have been brought to bear on that question. It will explain the theoretical underpinnings of independent component analysis (ICA) and review how it can be applied to questions about the neural underpinnings of cognitive deficits in SZ. It will lay the groundwork for the three subsequent chapters that describe the three fMRI studies that compose the body of this dissertation.

Functional Connectivity

Schizophrenia as a disconnectivity syndrome. There have been various theories about the nature of schizophrenia and its causes, especially with regard to specific levels of analysis (e.g., neurotransmitters, genes, social/developmental precursors, pre- or perinatal insults; Ritsner & Gottesman, 2011). One theory regarding the neural underpinnings of the symptomatology in schizophrenia is that schizophrenia is a syndrome of impaired or otherwise abnormal connectivity between brain regions (Friston & Frith, 1995). As opposed to other theories that specify discrete brain regions that show pathology, theories about disconnectivity hold that there are widespread impairments in

the coordination of neural networks. That is not to say that localized pathology does not exist; however, pathophysiology in discrete brain regions may not be able to offer a broad account of the various impairments and symptoms seen in patients with schizophrenia (Andreasen et al., 1999). Other evidence from neuroanatomical studies show micro-pathology of neural architecture in the brains of people with schizophrenia (Karbassforoushan & Woodward, 2012), which also involve altered neurotransmission (Lewis, Pierri, Volk, Melchitzky, & Woo, 1999). There are different methods of addressing and defining connectivity, such as diffusion tensor imaging for identifying white matter tracts. The current paper will focus on functional connectivity as measured by functional magnetic resonance imaging (fMRI).

Functional connectivity as means of assessing disconnectivity. Functional connectivity refers to the relationship, or dependence, of the time courses of neuronal activity between different areas of the brain (Friston, Frith, Liddle, & Frackowiak, 1993). As opposed to neuronal activation studies, which aim to find discrete areas of cortex that respond to certain stimuli, functional connectivity studies intend to elucidate the ways in which disparate brain regions activate together. Functional connectivity should be distinguished from effective connectivity, which aims to determine if, and to what degree, one brain area influences or controls another (Friston, 2011). Functional connectivity only measures the temporal association between brain areas without discerning causality between the areas. Biswal, Yetkin, Haughton, and Hyde (1995) were the first to apply functional connectivity measures to resting state fMRI data, showing that activity in the sensorimotor region of one hemisphere during rest correlated with

activity in the opposite hemisphere. Since then, large numbers of studies have measured functional connectivity both during task and during rest. Usually, studies are most interested in low-frequency oscillations in the neural activity patterns, as higher-frequency fluctuations are often associated with physiological signals of non-interest, such as cardiac activity and respiration (van den Heuvel & Hulshoff Pol, 2010). Researchers have identified a series of networks, starting with the “default mode network,” which routinely appear in studies of functional connectivity. The default mode network is so named because it consistently shows up in resting-state studies and typically shows negative correlations with task timelines. It is typically comprised of medial prefrontal cortex (PFC), posterior cingulate, precuneus, posterior parietal, and parts of the temporal lobe (Raichle et al., 2001). Other networks such as the frontal-parietal “executive” networks confirmed that functional connectivity measures could be used to discern task-relevant networks in addition to resting state networks.

The attribute of functional connectivity measures that make them attractive for studying schizophrenia is that they may be able to measure the disconnectivity underlying the symptomatology if the disconnectivity hypothesis is correct. Because functional connectivity aims to measure the association between brain regions, it can determine if schizophrenia patients’ brain networks act differently or are functionally organized differently from healthy controls.

Types of functional connectivity measures in fMRI. There are a few methods of measuring functional connectivity in fMRI data, but the methods can be broken down for the most part into model-based methods and model-free methods (Li, Guo, Nie, Li, &

Liu, 2009). The model-based methods make some assumptions about the nature of the functional connectivity patterns, usually by defining the areas that can be included in the analysis beforehand. Model-free methods do not make that kind of assumption, although they do make other assumptions.

The first type of functional connectivity analysis method I will discuss is **seed-based correlation**. This is the method Biswal et al. used in their initial fMRI functional connectivity study. Typically, this type of model-based functional connectivity analysis first defines a region of interest based on some *a priori* hypothesis or information (Margulies et al., 2010). The time series of that region is then extracted, and the researchers then quantify a measure of association between that extracted time series and the time series of all the other voxels or all the other regions of interest in the brain. Oftentimes this quantification of association is simply the Pearson product-moment correlation. What results is a map of the functional connectivity between the seed region of interest and the rest of the brain (van den Heuvel & Hulshoff Pol, 2010). There are variations in the methods researchers use to extract the time series of the region of interest as well as in the methods for defining association, but the overall methods of defining a seed and identifying the connectivity with that seed are constant within this method. One drawback of this method is that bias is introduced into the analysis by the choice of seed region.

The second type of analysis is **independent component analysis (ICA)**. This model-free method identifies the underlying groups of voxels that activate coherently in fMRI data. A more complete explanation of ICA will follow in Section 2 of the current paper. The benefits of ICA over seed-based methods are mainly that it requires no seed

region or model time course of interest (as opposed to seed-based correlation and activation contrast methods such as a general linear model). ICA identifies components across the brain that cohere temporally without introducing bias from the selected seed region or regions.

The third type of functional connectivity analysis is **graph theoretical analysis**. Graph theoretical methods are not actually separate from seed-based methods or ICA, because they can be applied to the results of both of those methods. In essence, graph theory seeks to identify a graph made up of nodes and edges (van den Heuvel & Hulshoff Pol, 2010). The nodes represent different regions of interest in the brain, and the edges represent the functional connectivity between them. After this graph is identified, metrics can be calculated on the graph, such as how efficiently information travels around the graph and topological features of the graph. The graph can provide information about the structure of information processing in people's brains. The nodes in these analyses can be determined a priori by an anatomical atlas, the results of previous studies, or a set of independent components.

Graph theoretical analyses typically find local and global descriptions of the graph, local having to do with the nodes of the graph and global having to do with the graph as a whole. Some of the terminology used in graph theory is distance, the minimum number of edges connecting two nodes; path, a series of connected nodes; and degree, the number of edges connecting to a node (Li et al., 2009). Some measures that are built on those terms are clustering coefficient, which is a local measure of whether nodes that are connected to a node are also connected to themselves; local and global efficiency

describe the ease of information travel throughout a network; and average path length, which is a global measure of how distributed the network is (Bullmore & Sporns, 2009).

As mentioned, the current paper's central focus is on ICA, the reliability of the method, and its application to schizophrenia research. The choice of this topic was made because of ICA's attractive features as an analysis tool and because there has been little synthesis of the ICA literature to date in schizophrenia research using fMRI. What follows is a detailed description of the method, followed by reviews of reliability studies using the method and schizophrenia studies using the method.

Independent Component Analysis

An explanation of independent component analysis. ICA is a member of a group of methods called blind source separation methods. This term refers to methods which aim to elucidate the underlying signals in a mixture of data (Hyvärinen & Oja, 1997). These methods are 'blind' because they find those sources without the use of a priori information about the signals or how they are mixed. As will be shown, ICA can be applied to many different problems.

One example of a situation in which ICA is an appropriate method is the cocktail party problem (Brown, Yamada, & Sejnowski, 2001; Stone, 2002). This famous example occurs when there is a group of N people in a room, each speaking to a fellow person. The resulting mixture of voice frequencies and amplitudes over time represents the observed data. The goal of this problem is to separate this mixture of voices into a group of individual voices, one for each person. To do this, one must record this room using N or more microphones in order to arrive at no less than N distinct mixtures of those voices.

Once these mixtures are obtained, ICA may be employed to separate the mixture into the constituent parts.

ICA, as its name suggests, assumes that the underlying source signals in a mixture are statistically independent of each other (to be discussed further below) (Stone, 2002). This is a reasonable assumption in most cases, such as the amplitude of voices in the cocktail party. Knowing the amplitude/frequency of one voice provides little information about other voices in the room. If we make that assumption, then we can look for unrelated sources of variation within the mixture, and that is just what ICA does.

Additionally, the source signals are assumed to be non-normally distributed (or non-Gaussian)(Langlois, Chartier, & Gosselin, 2010). Again, this assumption is tenable in most applications. The reasons for this assumption will be discussed below.

Last, because of the way ICA attempts to find the independent sources underlying the mixture of sources, it is assumed that the matrix that represents the way the sources are mixed is invertible. This means that there must exist as many mixtures of the source signals as there are source signals. This will be explained in more detail shortly.

Because the goal of ICA is to identify underlying sources of a mixed data set, it resembles the more common methods of factor analysis and principle component analysis (PCA). The assumptions of the ICA model just discussed make it different from both these methods. Both factor analysis and PCA find sources that are uncorrelated. Although similar to independence, uncorrelatedness does not provide as strict a degree of unrelatedness as statistical independence does (Stone, 2002). This is because signals that are uncorrelated may still be related nonlinearly. With factor analysis, this results in factors that may be rotated in infinitely many ways to provide separate sets of factors that

are equally statistically significant. ICA results in components that are uniquely defined, in that they may not be rotated to provide equivalent components. Any rotation will result in a set of components that are no longer statistically independent (Stone).

One way of thinking about linear orthogonality versus statistical independence is to imagine the linear versus nonlinear associations between signals (De Lathauwer, De Moor, & Vandewalle, 2000). Orthogonality between signals means that there is zero linear association between the signals (the Pearson's correlation between them would equal zero). Independence means that there is no linear *or nonlinear* association between the signals. This idea is depicted in figure 1.2. Another way of thinking about it is that the expected value of the product of two signals will equal the product of the expected values of the individual signals if the signals are orthogonal (Langlois et al., 2010). This can be written as:

$$E[x_1 * x_2] = E[x_1] * E[x_2]$$

where $E[\]$ is the expected value. If the variables are statistically independent, the following will be true:

$$E[g_1(x_1) * g_2(x_2)] = E[g_1(x_1)] * E[g_2(x_2)]$$

where g_1 and g_2 can be any functions.

In practice, attempting to perform a PCA or factor analysis on a mixture of signals fails to produce those source signals (Stone, 2002). What results is simply a set of linearly orthogonal mixtures.

The discussion of ICA requires talking about vectors and matrices of data. A vector here refers to a collection of values arranged in a particular way, such as the vector \mathbf{s} being composed of

$$s_1 \quad s_2 \quad s_3 \quad \dots \quad s_n$$

where s_1 refers to the first data point and n refers to the number of data points. Here, s_1 could refer to the amplitude of a voice in the first millisecond of a recording, which would make n equal to the number of milliseconds in the recording. However, as will be demonstrated later, s_1 could also refer to the darkness or lightness of the first pixel in a photographic image, which would make n equal to the number of pixels in that image (Bell & Sejnowski, 1995).

Matrices are made up of multiple vectors. A matrix of multiple source signals would look like this:

$$\begin{array}{cccccc} s_1^1 & s_2^1 & s_3^1 & \dots & s_n^1 \\ s_1^2 & s_2^2 & s_3^2 & \dots & s_n^2 \\ s_1^3 & s_2^3 & s_3^3 & \dots & s_n^3 \\ \dots & \dots & \dots & \dots & \dots \\ s_1^m & s_2^m & s_3^m & \dots & s_n^m \end{array}$$

where n is the number of data points in a source signal and m is the number of source signals. Therefore, s_n^m refers to the n^{th} data point of the m^{th} source signal.

If \mathbf{s} is a vector of values representing one source signal, then \mathbf{S} is the matrix of all source signals, with each signal representing one column. Likewise, if \mathbf{x} is a vector of values representing one mixture of those source signals, then \mathbf{X} is the matrix of all the observed mixtures of those signals. Therefore, \mathbf{X} represents the observed data matrix. The underlying assumption of source separation is that the columns of \mathbf{X} are in actuality mixtures (for our purposes linear mixtures) of the columns of \mathbf{S} . We can denote the “mixing” matrix that creates \mathbf{X} as \mathbf{W} . Therefore, we can construct this simple equation:

$$\mathbf{X} = \mathbf{W}\mathbf{S}$$

Note that neither \mathbf{W} nor \mathbf{S} is known in this equation. It is evident that in order to arrive at the matrix of source signals \mathbf{S} , we must first determine what the mixing matrix \mathbf{W} is. Then, we may take the inverse of \mathbf{W} in order to “unmix” the observed data \mathbf{X} back into \mathbf{S} .

$$\mathbf{S} = \mathbf{W}^{-1}\mathbf{X}$$

Conceptually, to determine an unmixing matrix, we must identify some aspect of source signals that is not present (or is less present) in mixtures of those sources (Lee, Girolami, & Sejnowski, 1999). Once a measure of this aspect is identified, it then remains to find an unmixing matrix that maximizes that measure. Typically, non-Gaussianity is used as this measure. This is because the central limit theorem states that any mixture of independent variables will tend to be more normally distributed than either of the constituent parts (Hyvärinen & Oja, 2000). Thus, by finding signals that are least normally distributed, we can identify likely candidates for non-mixtures. We assume that these non-mixtures represent the original source signals.

This aspect of mixtures can be demonstrated using histograms of the respective variables and the mixture. If two signals A and B are represented by histograms, then it becomes clear when the histogram of their sum is displayed that that mixture is more normally distributed than either of the original sources. Please refer to figure 1.1.

This fact is key to the principle of ICA. Because we assume $\mathbf{X} = \mathbf{W}\mathbf{S}$, every column of \mathbf{X} is a linear combination of the columns of \mathbf{S} . Therefore, each is more Gaussian than any of the columns of \mathbf{S} . In fact, if we estimate the vector \mathbf{w}^{-1} (in order to estimate one signal in \mathbf{S}) to maximize the non-Gaussianity of $\mathbf{w}^{-1}\mathbf{X}$, it will be least

Gaussian when it *equals* one of the columns of \mathbf{S} , the original source signals. Therefore, $\mathbf{w}^{-1}\mathbf{X}$ results in one independent component (Yang & Amari, 1997).

This process has two drawbacks. The first is that each solution results in two sets of independent components, one positive and one negative. It is impossible for an ICA algorithm to distinguish between these two solutions. Therefore, it is sometimes necessary to change the sign of certain components. In practice, it is obvious when such an error occurs. An example is when isolating components of a mixed image, one component might appear negatively valued (Stone, 2002).

The second drawback has to do with the scaling of the resulting components. Most ICA algorithms require that the mixtures be scaled to have a mean of zero and a variance of one. The resulting components thus lose the information of absolute magnitude. Again, in practice this fact does not prove to be troublesome as the magnitude is either arbitrary or can be inferred by applying the inverse of the whitening matrix to the appropriate component (Stone, 2002).

There are different methods of measuring, and thus maximizing, how non-Gaussian a signal is. The two most popular methods employ measures of the kurtosis and of the entropy of the source signals (Langlois et al., 2010). We will focus on entropy, specifically negentropy, as a means of maximizing non-Gaussianity, as a demonstration.

Entropy is a measure of how much information is given by observing a random variable (Hyvarinen & Oja, 2000). Variables with high entropy are more random and unpredictable.

Entropy is measured as:

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$$

where $f()$ is the probability density function for that variable.

For a given variance structure, a Gaussian variable has the largest entropy. Therefore, by comparing the entropy of a variable to a Gaussian variable with the same variance, we may obtain a measure of the degree of non-Gaussianity. Specifically, we can use negentropy because it is always non-negative. It is:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{Gaussian}}) - H(\mathbf{y})$$

where $\mathbf{y}_{\text{Gaussian}}$ is a Gaussian random variable with the same variance structure as \mathbf{y} and $H()$ is the entropy of the variable.

However, negentropy itself is difficult to calculate because it requires knowledge of the probability density function for the variables. There are methods of approximating negentropy that perform well and are much less complicated. Hyvarinen and Oja (2000) proposed using the following approximation:

$$J(\mathbf{y}) \propto [E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\}]^2$$

where $G()$ is any non-quadratic function and \mathbf{v} is a standardized Gaussian variable.

Typically, the hyperbolic tangent is used as $G()$.

FastICA algorithm. The algorithm employed in the demonstration below is the fastICA algorithm (Hyvarinen & Oja, 2000). It consists of the following steps:

1. Initialize \mathbf{w}_i (typically as a random variable)
2. $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$

3. $\mathbf{w} = \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$
4. If not converged, go to step 2.

In this algorithm, to converge means that the dot product of the old and new values of \mathbf{w} is close to 1. This depends on the data being pre-whitened, which is part of a series of preprocessing steps to be discussed next.

Preprocessing in ICA. The first step in preprocessing is to demean the data by subtracting the column means from the columns. This results in a matrix \mathbf{X} whose columns all have means equal to 0. This step facilitates easier computation of the ICA algorithm (Langlois et al., 2010).

The next step is to whiten the data. Whitening consists of linearly transforming the matrix \mathbf{X} such that the columns are all orthogonal and all have variances of one. That way, the covariance matrix of \mathbf{X} equals the identity matrix. One method of whitening is to first perform an eigenvalue decomposition of the covariance matrix of \mathbf{X} such that the results are $\mathbf{E}\mathbf{D}\mathbf{E}^T$, where \mathbf{E} is a matrix of eigenvectors and \mathbf{D} is the diagonal matrix of eigenvalues. Next, to whiten, we do:

$$\mathbf{X}_{new} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{X}$$

This transformation forces the mixing matrix to be orthogonal, making it easier to estimate due to the fact that orthogonal matrices contain approximately half the number of parameters to estimate compared with a random matrix (Hyvarinen & Oja, 2000).

It is also sometimes useful at this stage to examine the eigenvalues obtained while whitening in order to determine if dimension reduction is possible. This can greatly

reduce computational costs in the case of large datasets, such as those encountered in fMRI studies.

Independent Component Analysis of fMRI Data

ICA was first employed with fMRI data by McKeown and others in 1998 (McKeown et al., 1998). In this seminal paper, the authors showed [1] that ICA can be successfully used to detect task-related components in fMRI data; [2] that although the task-related components that were derived using ICA shared the bulk of their mass with areas identified using more traditional fMRI analyses of those tasks, the ICA-derived components included spatially disparate regions that had not been associated with those tasks previously; and [3] that ICA outperformed PCA with regard to estimating the spatial extent and time course of components in that data set.

There are some particulars of their study that serve to illustrate the challenges and considerations of applying the ICA procedure to fMRI data. The first is that McKeown chose to use spatial ICA (sICA) as opposed to temporal ICA (tICA). This means that the components that result from the analysis represent groups of voxels which are maximally spatially independent from other components, but which share a common time course. Another way of looking at it is that ICA decomposes the mixed fMRI data set into a set of images (the spatial maps of the components) that are linearly mixed in different ratios over the full time course of the scan (the relative contribution of a given image at each time point makes up its time course) (Calhoun, Adali, Pearlson, & Pekar, 2001). The alternative (tICA) derives a set of time courses, each with an associated spatial map, that linearly mix to form the whole time course (Calhoun, Adali, Pearlson, et al.). Although

McKeown argued for the use of sICA for fMRI data based on the fact that the spatial patterns sought for both the signal components and the noise components in fMRI data are typically sparse and localized (McKeown, 2003), the choice of whether to use sICA, tICA, or some other variation has been up to the individual researcher. However, sICA is the most common ICA method employed with fMRI data, partially due to the fact that there are many more observations in space (voxels) than in time. Some researchers have attempted to combine sICA and tICA methods to some extent (Calhoun, Eichele, & Pearlson, 2009; Seifritz et al., 2002; Stone, Porrill, Porter, & Wilkinson, 2002) or attempted to circumvent the problem by using 3-dimensional tensors as opposed to matrices (Beckmann & Smith, 2005); however, sICA is still the most widely used method with fMRI data.

The second aspect of the McKeown study that illustrates a difficulty with applying ICA to fMRI data is that the authors performed ICA on each subject's functional data individually. This aspect of the study does not become problematic until one wants to draw conclusions from the group's components as a whole or to compare two separate groups. If each subject has her or his own mixing matrix calculated, then each subject will have a unique set of spatial maps and time courses. Additionally, the components that result from ICA are not ordered, further complicating the issue. Unlike a GLM analysis in which each regressor's time course is defined by the researcher *a priori*, thus creating a natural means of comparing subjects, there is no natural way of deciding which components to compare between subjects/groups. Researchers were quick to suggest solutions to this problem, and they did so in various ways (Calhoun et al., 2009).

The first type of solution still employs individual ICAs, but subsequently a strategy is used to group the components based on similarity. The strategy is usually one of template matching (Calhoun, Adali, McGinty, et al., 2001) or a clustering algorithm (Esposito et al., 2005). Template matching involves assessing the spatial similarity of the components resulting from an ICA with a template, or map, of some regions of interest. Clustering techniques find groups, or clusters, of components that are most spatially similar to each other and dissimilar from other clusters. These strategies allow for good individualized decompositions for each subject, but they have been criticized because there is no guarantee that the components that are grouped together actually represent the same thing in different subjects (Margulies et al., 2010). With regard to template-matching, the approach assumes that components of interest will actually match a set of templates and that those templates are appropriate (Margulies et al.).

The second type of solution involves grouping the data from all subjects together and performing an ICA on that grouped data. This general method was first proposed by Calhoun and others (2001), and variations on that theme have arisen since then. In general, the method consists of concatenating subjects' 4-dimensional functional data (3 space dimensions X time) in the time dimension into a single data set. Then a single ICA is performed on the whole data set. This process results in a set of group component spatial maps and time courses. Calhoun's method then involves back-reconstruction, which puts the group components back into subject space. By keeping the subject-specific aspects of the components in reference to the group components, this method allows between-group interpretations more naturally than with a template-matching procedure. Because the concatenated group data set typically is quite large, performing an

ICA on it is computationally demanding. Typically, a series of PCA data reduction steps is performed prior to the ICA step in order to make computation possible. These PCA steps introduce potential problems into the analysis, which will be discussed later in this paper.

Another form of group ICA is referred to as probabilistic ICA (pICA), and was introduced by Beckmann and Smith (2004). Whereas the typical model of group ICA enforces a deterministic decomposition of the fMRI functional data (the matrix multiplication of the unmixing matrix with the source matrix perfectly recreates the original data), pICA allows for probabilistic noise in the mixture. This makes a direct reconstruction of the original data impossible from the results of the pICA, but it may produce more robust group estimates (Beckmann & Smith, 2004). pICA has been implemented in the MELODIC software within the FSL package (<http://www.fmrib.ox.ac.uk/fsl/>).

Typically, pICA is followed by a procedure known as dual regression (also known as spatio-temporal regression; Filippini et al., 2009). Dual regression is similar to back-reconstruction in that its aim is to compute the subject-specific features of the group level components. However, whereas back-reconstruction is more or less “undoing” the stages of the group ICA that produced the group level components by employing the unmixing matrix and the intermediate PCA steps to derive subject level data, dual regression does not use information from the group analysis itself to compute subject specific components. Instead, it consists of first performing a spatial regression with the group level component spatial maps as regressors and the subjects’ functional data as the predicted variable. The betas that result represent the individual time courses for that

subject. Then, a second, temporal regression is performed using the time courses calculated in the previous regression, again in the subjects' functional data. The result of this second regression provides a set of spatial maps representing how closely related each voxel's time course in the functional data is with the computed component time course.

These two group ICA methods of "noise-free" ICA plus back-reconstruction (implemented in the GIFT software package, <http://icatb.sourceforge.net/>) and pICA plus dual regression are the most commonly used methods for ICA analysis of fMRI data. However, there are other methods such as tensor ICA (Beckmann & Smith, 2005) and approaches that would in effect perform a group ICA on each group of subjects individually as opposed to all in the same group (Celone et al., 2006).

The third aspect of the McKeown study (1998) that brings up an issue with ICA of fMRI data (although it is not specific to fMRI) is the fact that the authors extracted 144 components from each of the subjects' data (one for each time point in the data). The ICA algorithms typically require a number of components to be pre-selected in order to be extracted, and this number, known as dimensionality, is arbitrary. The choice of dimensionality is nontrivial, as demonstrated by Ma et al. (2007), who found that the ICA results varied significantly when the dimensionality is much less than the number of time points in the data (as is typical in ICA studies). There have been attempts to use techniques to estimate the number of components that should be extracted by the ICA algorithm, such as minimum description length criterion (MDL; Rissanen, 1978) and Laplace approximation, among others (Calhoun & Adali, 2012). These information theory-derived tools eliminate the need for the researcher to decide on an arbitrary

number of components, but it remains to be seen if these tools result in the “best” decomposition of fMRI data. In fact, the question of what is best changes with regard to the research question, as some researchers intentionally use high-dimensionality ICA to break up components spatially into smaller pieces (Kiviniemi et al., 2009).

The last aspect of the McKeown study that illuminates a consideration when doing ICAs with fMRI data is at what level to threshold the spatial maps that result from the ICA. The spatial maps that result from an ICA contain values that can be interpreted as each voxel’s contribution to the time course. McKeown z-transformed these values and set a threshold of $|z| > 2$, such that voxels with z values less than 2 and greater than negative 2 were zeroed. This process of applying a transformation (typically one that imposes a Gaussian distribution on the values) is very commonly done (Li et al., 2009; Zhao et al., 2004) and has been shown to inflate the type 1 error rate (Ma et al., 2007). Additionally the decision of at which point to cut is an arbitrary one. Using pICA, it is possible to estimate a null distribution based on the noise estimates of the analysis, but even then an arbitrary level of significance is required, where $p < .05$ (typically corrected for multiple comparisons and cluster thresholded) is the usual threshold. It remains to be seen what effect the chosen threshold has on the validity of ICA results in fMRI, but the choice does affect the reliability of the analysis (Poppe et al., 2013).

This section aimed to provide a summary of some of the challenges inherent in applying ICA to fMRI studies (particularly group studies) as well as many of the common solutions that have been proposed and implemented to address these challenges. The reader should take away from this section the idea that there is no singular method known as ICA when applied to fMRI data, and conclusions drawn about the validity and

reliability of ICA analyses must take into account the specific methodology used in each study.

Reliability of fMRI Analyses

Importance of reliability. Reliability is the property of tests that deals with how consistently a measurement is being made. Scientists can have no real confidence in the results of tests that cannot provide consistent results. Validity depends on it. Additionally, the extent to which two measures correlate also depends on the reliabilities of the two measures. If either of the measures has a low reliability, it sets an upper bound on how highly the two measures can be associated. In the field of neuroimaging, reliability is just as important. There have been relatively few investigations of reliability in fMRI, however, perhaps because it is commonly assumed to be sufficient because fMRI measures a physical property. However, as we shall see, this assumption is misplaced.

There are also multiple ways of thinking about reliability. Probably the most common way of describing reliability is to say that a test provides the same score on the same subject at two time points. This is test-retest reliability (TRT reliability). It is easily interpreted and is easily calculated, relatively. For fMRI research, TRT reliability is the primary form of reliability that has been studied thus far.

Reliability considerations in fMRI. There are many sources of noise in fMRI research that can disrupt its TRT reliability. These come from sources at different levels of the analysis. The deepest levels of noise result from hardware inconsistencies and failures, such as b0 field inhomogeneity and thermal or other fluctuations in the scanner

room. The effects on reliability of these hardware problems has largely been ameliorated by the years of strict specification by scanner manufacturers, and so the sources of poor reliability in fMRI research largely stem from other sources, such as signal to noise ratio of the data due to scanner settings.

The data collected by fMRI is an indication of how much oxygenated blood is in any one area of the brain at a given time. The individual cells in the brain and surrounding tissues are summarized by volumetric pixels (voxels) typically ranging between 0.5^3 mm^3 to 4.5^3 mm^3 in volume. In essence, we get one "score" for each voxel at every timepoint of the scan. If a certain voxel is seen to have an increase in its score following a given stimulus (and which follows a plausible physiological response schedule), we may be inclined to believe that neurons in that voxel required more oxygenated blood because they increased their activities in response to a stimulus. However, there are always sources of noise that impede our ability to make those claims, because the increase in signal in a given voxel between two time points may be due to chance or to some other process rather than neurons' hemodynamic response.

Scan protocol settings such as the voxel size, the length of each acquisition, the echo time, and the angle of acquisition all influence the signal to noise ratio of the data. This SNR is an indication of how sure we are that the score we have for a voxel is indicative of oxygenated blood rather than to some source of noise in the data.

Classical test theory (CTT) can be used with fMRI data. Any observed score, such as the scores in each of our voxels after a scan, are made up of some true score and any error associated with it. In CTT the errors are thought of as random, such that the mean of the errors in a given data set is zero with some variance. Voxels with more noise

associated with them, then, will be less reliable than other voxels, because the errors are not systematic and so are not stable. Averaging over many scores should give a better indication of the true score because the random errors should in effect cancel each other out. This has implications for scanner settings, because changing the voxel size by 1mm^3 greatly changes the number of voxels in the data. This specific change has at least two effects. First, there are fewer total voxels if the size gets larger, but each of those voxels contains more cells. So, for a given volume of brain, averaging over that volume is less effective for reducing voxels' measurement error, because there are fewer voxels to contribute. However, each of those voxels should be contributing less specific error than their smaller volume counterparts.

In TRT reliability work that has been conducted thus far with regard to fMRI, the majority has been conducted with activation contrast studies. This means that the analysis method is to regress the time course of whatever stimulus was presented to the subjects onto those subjects' fMRI data. Each voxel gets another score. This time, the score represents how closely related to the task timeline the voxel's temporal pattern of activation was. If a voxel activated whenever a particular stimulus was presented, it will have a high score for that stimulus. Then, typically clusters of voxels are found whose relationship with a particular stimulus is greater than or less than either their typical (average) signal or than their relationship to some other signal. If these voxels' difference in relationship between the two stimuli (or between stimulus and mean) is more pronounced than would be expected by chance, then we say they are significantly related to the stimulus. This comparison of the relationship of voxel time course and stimulus

time course to the mean or to other stimulus time course is known as an activation contrast method.

When scientists investigate reliability in fMRI, they are typically interested in how consistent the resulting scores following statistical analysis are, rather than raw scores. This is akin to a researcher being interested in the reliability of the accuracy of a subject tested with a multi-item achievement test (accuracy being a summary statistic) rather than the consistency of any one item. Thus, researchers in fMRI typically examine the reliability of their statistical results after an activation contrast procedure. The main question of this method is, “Do the same areas of brain activate significantly to the same task at different time points?”

Methods of measuring fMRI reliability. There are different choices for what a researcher might be interested in when assessing reliability. It should be noted here that the term “reliability” in this section is used rather broadly to mean general consistency of a measure. In subsequent chapters, reliability will refer to the specific property of the consistent measurement of individual differences such that individuals who score highly at one time point also score highly at a second time point.

In activation contrast studies, there are three main categories. The first is whether the location of activation remains constant from time 1 to time 2. This consideration is usually measured by a spatial overlap statistic, such as the Dice coefficient (Dice, 1945), which measures the spatial overlap between two sets of data. It is defined as:

$$SI_D = \frac{2 \times V_{overlap}}{(V1 + V2)}$$

where SI_D is the Dice Similarity Index, V_1 is the number of voxels above threshold in the first cluster, V_2 is the number of voxels above threshold in the second cluster, and $V_{overlap}$ is how many voxels are overlapping between them. This coefficient can vary between 0 (no overlap) and 1 (perfect overlap). The similar Jaccard coefficient (Jaccard, 1901; Levandowsky & Winter, 1971) measures the same idea, and is a mathematical transformation of the Dice coefficient. Both coefficients are used in fMRI reliability studies to assess spatial overlap of significant clusters of voxels.

A stricter form of reliability is whether the magnitudes of voxel activations within a given cluster is reliable over time (Bennett & Miller, 2010). In order to measure this, researchers typically use an intra-class correlation coefficient (ICC)(Shrout & Fleiss, 1979). This metric is measured in different ways, as defined by Shrout and Fleiss, but commonly it is defined as:

$$ICC = \frac{\sigma_{Between}^2}{\sigma_{Between}^2 + \sigma_{within}^2}$$

where $\sigma_{Between}^2$ is the variance in scores between voxels, and σ_{within}^2 is the variance within voxels. This shows how reliability is measured, in that if voxels vary much more between themselves than from one time point to another, the ICC and the reliability are high. The ICC typically ranges from 0 to 1, although other values are possible.

Furthermore, there are a few ways to summarize ICCs when measuring reliability in fMRI data, each attempting to report the central tendency of the distribution of all ICCs in the voxel data (Bennett & Miller, 2010). One caveat for interpretation of the ICC between groups is that if one group is more heterogeneous than the other, or has more

heterogeneous voxel activity, it can have higher ICC values without actually having more consistent results (Weir, 2005).

A problem with both of the previous methods of assessing fMRI reliability is that they depend on the threshold used with the spatial maps (i.e., the determination of what is significant and what is not). In order to eliminate that problem, one could assess the ICC for all voxels in the brain, regardless of whether they are super-threshold. This method is the strictest in that it measures whether all voxels act the same at both time points regardless of whether they respond to the stimuli of the fMRI task or not. These three methods are used in activation contrast studies of reliability, but as will be shown, they are also used in the studies performed on functional connectivity metrics. Additionally, other metrics are used to assess reliability, but these three are the most common.

Reliability in activation contrast studies. Bennet and Miller (2010) performed a literature review in which they summarized the results of activation contrast fMRI studies that looked at TRT reliability. They grouped their results according to whether the individual studies used spatial overlap measures, ICC, or some other reliability metric. They reviewed 63 papers total. With regard to spatial overlap studies, the metrics were converted into Dice coefficients and the minimum, mean, and maximum values were reported if available. The average of each of those three values was also reported. They found that the average of the mean spatial overlap in the reviewed studies was .48, mean minimum was .31, and the mean maximum was .67. Likewise, they did the same with ICC studies, and found that the average mean was .50, the mean minimum was .17, and

the mean maximum was .75. They did not summarize the studies using other reliability metrics.

Although there is no rule for what constitutes an acceptable reliability based either on Dice coefficients or ICC values, a minimum value of .4 ICC has been proposed (Cicchetti & Sparrow, 1981). Based on that criterion, activation contrast studies appear to have acceptable reliability on the whole. It is necessary to establish that, because it allows the comparison between functional connectivity reliability and activation contrast reliability. The interpretation of functional connectivity reliability thus will not be impeded by questions about fMRI reliability generally, as there is a precedent value for an alternative analysis of fMRI data. However, the “acceptability” of a measure’s reliability when that measure is used in group contrast studies depends on the size of the effect in question as well as the sample size of the study. As noted by Button and colleagues (2013) noted, unreliable studies in neuroscience (whatever the source of the unreliability) are wasteful and can increase type 2 errors as well as exaggerate the size of true effects.

Considerations for reliability in ICA studies. There are a number of factors that could potentially influence the TRT reliability of ICA results and that are specific to the manner in which ICA is performed. Some of these factors are subject concatenation order, component number (model order), the type of ICA algorithm used, the patient status of the subjects, and how the subjects are engaged during the scan (i.e., task versus resting state).

The first factor is subject concatenation order (SCO). This refers to the order in which subjects are placed into a temporal concatenation ICA. In actuality, SCO becomes a problem because of the PCA steps used prior to ICA for the purposes of data reduction. Depending on the size of the dataset, typically either two or three PCA steps are used to reduce fMRI data. If two steps are used, the first step typically reduces data within an individual subject. The second step then concatenates all the reduced data from all subjects and reduces that further. This process does not introduce effects caused by SCO. However, if three PCA steps are used to reduce data prior to ICA, which is the typical method, the second step forms subgroups of subjects and reduces the data of those subgroups. The third PCA step then acts on the concatenated reduced data from those subgroups. Because of the nature of this process, an individual subject or a group of subjects may have undue, or at least differential, influence on the final results depending on the subgroup in which she or he is placed. If the subjects were ordered differently, the results of the second PCA step, and therefore the results of the final ICA would be different. This has been demonstrated by (Zhang et al., 2010), who discussed SCO and the number of effective subject orders that are possible. There are fewer effective subject orders than total different subject orders, because the order of subjects within a subgroup at the second PCA step does not affect the results of the second PCA step. Zhang et al. showed that the repeatability of ICA results was affected by SCO, and they proposed subject order independent ICA (SOI-ICA). This method consists of performing a number of ICA analyses on the same set of data with varying effective subject orders. The results of these analyses are then combined to arrive at some central tendency of the resulting components, and thus the underlying true signals. A similar method has been proposed

(Wisner, Atluri, Lim, & MacDonald, 2013) to correct for SCO, which consists of performing many ICAs on the same set of data with varying subject orders and then performing a final, meta-ICA on the results of those ICAs. A question left open by Wisner and others was how many subject orders are necessary to achieve acceptable reliability, and this question was answered by Poppe et al. (2013). They found that 25 subject orders were sufficient in both a resting-state sample and a task-based sample to provide adequate reproducibility of group spatial components.

The second factor that may influence the TRT reliability of ICA results is the number of components (model order) that are specified to be derived in the ICA algorithm. Model order has been examined previously with regard to its effects on group difference metrics between patients with seasonal affective disorder and healthy controls (Abou Elseoud et al., 2011), as well as with regard to the robustness and potential validity of resulting components (Kiviniemi et al., 2009). Abou Elseoud found that the ability to detect group differences differed with different model orders, and the optimum model order was 70 components. Above 70, the effect stabilized such that there was no added benefit. This result may suggest that the reliability of the analysis is what is affected by the model order, with optimum reliability being achieved at 70 components. Kiviniemi et al. found that using a model order of 70 not only resulted in robust components, but these components represented a functional segmentation of the fMRI data. Poppe et al. (2013) examined the effect of model order on the spatial reproducibility of ICA components, and it was found that higher model orders produced the highest reproducibility scores in both task-based data and resting state data 70 and 60 components respectively.

Some researchers have attempted to ascertain what the optimal number of components might be for a given data set by using various algorithms. Some of the methods currently in use are Minimum Description Length criteria, AIC, and Laplace approximation (Li, Adali, & Calhoun, 2007). However, the results of these algorithms have not been adequately assessed with regard to their effect on TRT reliability.

The third factor that might influence reliability is how the spatial maps resulting from ICAs are thresholded. When using noisy ICA, it is possible to derive a Z score for a given voxel's association with the timecourse of a component, and therefore there is some justification for using a threshold of about 1.96. However, that value is based on an ultimately arbitrary 'significance' idea that has little relevance for the spatial maps resulting from ICA.

The fourth factor refers to the ICA algorithm itself. The two most commonly used methods of performing group ICA are "noisy" ICA and "noise-free" ICA. There have been some comparisons made between these two types (Erhardt et al., 2011), but it remains to be seen how the ICA algorithm type affects TRT reliability.

The fifth factor that may influence TRT reliability in ICA is the form of back-reconstruction used to arrive at individual time courses and spatial maps for each subject. The two most popular methods are referred to as dual regression and back reconstruction after GICA. A study recently compared these two methods of back-reconstruction (Erhardt et al., 2011), but again it remains to be seen what effect on reliability this choice has.

Another factor that could influence the TRT reliability of ICA results has to do with the characteristics of the subjects themselves and how those characteristics are dealt

with in the analysis. For instance, control subjects' data are likely more reliably measured to begin with than a group such as people with cocaine addiction or schizophrenia, and the results of analyses on controls' data may also be more reliable. There is evidence that the TRT reliability of schizophrenia subjects is lower than that of healthy controls despite the same reliability between groups in the cognitive task performed in the scanner (Manoach et al., 2001). This finding could have implications for interpreting ICA results, because apparent group differences could arise either because a functional network is less active on average in schizophrenia patients relative to healthy controls or because that network is less reliably measures in patients relative to controls.

The last factor I will discuss that could influence the TRT of group ICA results is the activity that subjects were engaged in at the scan time. The majority of FC studies in general are conducted using resting state scans. It is conceivable that the results of these studies are less reliable than the results of studies in which subjects performed a task in the scanner at both time points. The 'resting state' is typically an unconstrained time during which a subject may think about whatever pleases her or him. Although there may be some similarity both in what a given subject thinks about at two time points and in what two different subjects think about at the same time point, this similarity may be less than the similarity would have been had they been engaged in a structured cognitive task. A counterpoint may be that people may engage different cognitive or neural networks when performing the same task. Considering that the resting state has been disproportionately represented in the reliability literature of ICA, the reliability of ICA during a cognitive task is important to establish. The cognitive task that the current dissertation involves is a measure of the executive function known as goal maintenance.

Goal Maintenance

Cognitive deficits are debilitating for schizophrenia patients (Heinrichs, 2005), and these deficits may represent biomarkers for schizophrenia (Snitz, MacDonald, & Carter, 2006). One cognitive process currently being studied as a potential imaging biomarker for treatment response in schizophrenia is goal maintenance. Goal maintenance refers to an executive process that encompasses the representation of task goals and rules, the maintenance and manipulation of information relevant to those goals, and the application of that information to guide behavior (Henderson et al., 2012). Both schizophrenia patients (Jones, Sponheim, & MacDonald, 2010; Servan-Schreiber, Cohen, & Steingard, 1996) and their nonpsychotic first-degree relatives (MacDonald, Pogue-Geile, Johnson, & Carter, 2003) have demonstrated specific deficits on tasks measuring goal maintenance.

The establishment of goal maintenance as an imaging biomarker first requires that the reliability of activation patterns observed in fMRI data during the execution of a goal maintenance task must be established. The goal maintenance task to be used in this study is the dot pattern expectancy task (DPX). The establishment of DPX imaging as a reliable method will allow researchers to use the DPX results to define brain regions that represent reliable biomarkers of schizophrenia itself. It will give researchers a tool for assessing the specificity of activation differences during the DPX task in these brain regions to schizophrenia compared with other psychiatric illnesses. The identification of these brain regions that are differentially active during the DPX will allow the search for specialized pharmacological interventions for cognitive deficits to have a better-defined

target. Likewise, cognitively derived imaging biomarkers may serve as measures of progress in cognitive remediation or medication treatment protocols. To these ends, this dissertation will address the following seven specific aims:

Specific Aims

- (1) The dot pattern expectancy task (DPX) has never been used in an fMRI activation study that included schizophrenia patients. Therefore I will determine if a GLM analysis of the DPX replicates activation differences between SZ and HC observed using other goal maintenance measures. (Chapter 2)
- (2) There is ample evidence of disrupted functional connectivity in SZ, and the interaction of intrinsic connectivity networks has been hypothesized to subserve cognitive control (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008). Functional network connectivity (both tonic and dynamic; tFNC and dFNC, respectively; Jafri, Pearlson, Stevens, & Calhoun, 2008; Sakoğlu et al., 2010) is well suited for measuring the temporal relationships among brain networks. The second aim will be to find relationships among ICA-derived brain networks that underpin goal maintenance ability in SZ and HC. (Chapter 3)
- (3) The dot pattern expectancy task (DPX) has been shown to be a reliable measure of goal maintenance in schizophrenia patients and healthy controls (Jones et al., 2010). It remains to be seen if the activation differences observed in schizophrenia patients compared with healthy controls during the execution of the DPX are also reliable. A test-retest design will be employed to test the reproducibility and reliability of the results of a GLM analysis of the fMRI data

collected during the DPX task in both schizophrenia patients and healthy controls.

(Chapter 4)

- (4) There is some evidence that the results of ICA analyses are reproducible and reliable, but a direct comparison between ICA and the more traditional GLM analysis method has not been done. This study will perform both types of analyses on a test-retest sample of schizophrenia patients and healthy controls, and it will directly compare the results. (Chapter 4)
- (5) No reliability studies currently exist of either tFNC or dFNC. The fifth aim of this dissertation will be to assess the test retest reliability of tFNC and dFNC in HC and SZ. (Chapter 4)
- (6) Schizophrenia patients routinely show differential activation and functional connectivity in many of the same brain regions, such as dorsolateral prefrontal cortex (dlPFC), the default mode network (DMN), and posterior parietal lobes. There is evidence that schizophrenia patients have reduced reliability of activation-based analyses in “cognitive” brain areas as opposed to those areas more associated with primary sensory regions (Manoach et al., 2001). If this result holds true using ICA, then it leads to the theoretic possibility that differences found between schizophrenia patients and healthy controls in these areas (to the exclusion of other areas) may represent a measurement confound causing the ability to detect group differences in those areas to be greater than in other areas. This study will attempt to discern any main effects of group and component on reliability as well as any group by component interactions. (Chapter 4)

(7) Some previous research has found that collecting fMRI data from different sites reduces overall reliability if measures are not taken to correct for the effect, whereas other studies have not found that to be the case. Those studies used activation analysis methods. It is possible that ICA is not as sensitive to site effects, because it is possible that any effect due to site will encompass one of the independent components that result from the analysis. This study will examine whether such site effects on reliability present a problem in a multisite reliability study of schizophrenia patients and healthy controls. Additionally, it will compare such site effects in both GLM and ICA analysis methods. (Chapter 4)

Figure 1.1. Mixing of Variables Increases Normality

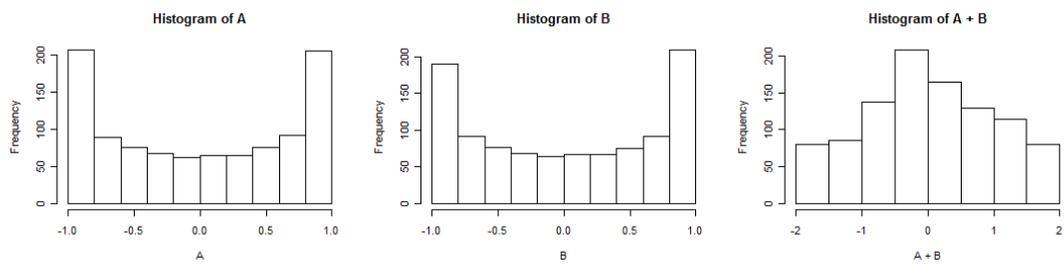
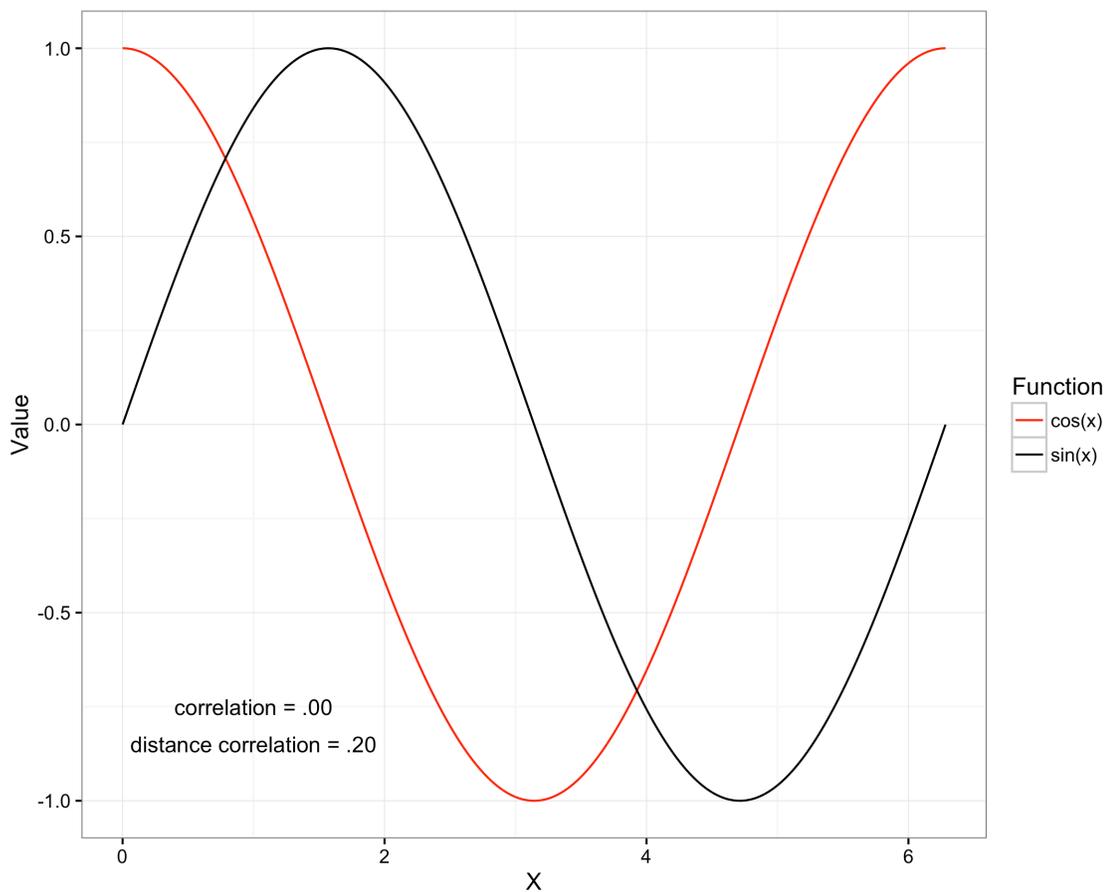


Figure 1.2. Orthogonality Versus Independence



Here, the two curves represent the values of sine and cosine from 0 to 2π . The linear dependence, as measured by Pearson's Product Moment Correlation is 0, indicating orthogonality. However, the distance correlation (Székely, Rizzo, & Bakirov, 2007), a measure of statistical dependence, between the two is non-zero.

Chapter 2: Reduced Frontoparietal Activity in Schizophrenia is Linked to a Specific Deficit in Goal Maintenance: a Multi-site Functional Imaging Study

Foreword: This chapter was written in collaboration with Angus W. MacDonald, who edited and revised versions of the manuscript. Additionally, Deanna M. Barch, Cameron S. Carter, James M. Gold, J. Daniel Ragland, and Steven M. Silverstein provided edits and some text revisions on final versions of the manuscript. The text of this chapter has been published in *Schizophrenia Bulletin* and is being reproduced here and as Appendix 1 with permission from Oxford University Press. The full citation of the published work is as follows:

Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., & MacDonald, A. W. (2016). Reduced Frontoparietal Activity in Schizophrenia Is Linked to a Specific Deficit in Goal Maintenance: A Multisite Functional Imaging Study. *Schizophrenia Bulletin*. In press.

Abstract

Patients with schizophrenia (SZ) previously demonstrated specific deficits in an executive function known as goal maintenance, associated with reduced middle frontal gyrus (MFG) activity. This study aimed to validate a new tool – the Dot Pattern Expectancy (DPX) task – developed to facilitate multi-site imaging studies of goal maintenance deficits in schizophrenia or other disorders. Additionally, it sought to arrive at recommendations for scan length for future studies using the DPX. Forty-seven SZ and

56 healthy controls (HC) performed the DPX in 3-Tesla functional magnetic resonance imaging (fMRI) scanners at five sites. Group differences in DPX-related activity were examined with whole brain voxelwise analyses. SZs showed the hypothesized specific performance deficits with as little as one block of data. Reduced activity in SZ compared with HC was observed in bilateral frontal pole/middle frontal gyrus as well as left posterior parietal lobe. Efficiency analyses found significant group differences in activity using 18 minutes of scan data but not 12 minutes. Several behavioral and imaging findings from the goal maintenance literature were robustly replicated despite the use of different scanners at different sites. We did not replicate a previous correlation with disorganization symptoms among patients. Results were consistent with an executive/attention network dysfunction in the higher levels of a cascading executive system responsible for goal maintenance. Finally, efficiency analyses found that 18 minutes of scanning during the DPX task is sufficient to detect group differences with a similar sample size.

Introduction

Cognitive deficits represent a debilitating and difficult to treat facet of schizophrenia, and they involve many aspects of cognition including memory, attention/concentration, and executive functioning (Gold & Harvey, 1993). Although these deficits remain largely unaffected by traditional psychotherapeutic and pharmacological interventions, recent initiatives in both these domains hold promise for effective treatments (Choi, Wykes, & Kurtz, 2013; Wykes, Huddy, Cellard, McGurk, & Czobor, 2011). Therapeutic efforts depend on accurate and reliable measures of deficits

in specific cognitive functions to chart treatment-related changes (Carter & Barch, 2007). The Cognitive Neuroscience Test Reliability and Clinical applications for Serious mental illness (CNTRaCS) Consortium was organized to develop and evaluate novel cognitive neuroscience-based measures of cognitive deficits in schizophrenia that tap specific brain-based mechanisms (Carter & Barch, 2007; Gold et al., n.d.).

In addition to the cognitive domains of visual integration (Silverstein et al., 2015) and relational encoding and retrieval (Ragland et al., 2015), CNTRaCS sought a valid measure for goal maintenance, which reflects the ability to retain and utilize relevant contextual information while pursuing a novel goal. For example, goal maintenance is required to overcome one's habitual route home from work given an errand that must be completed on the way. It is more than remembering the errand; rather, it is keeping the context of the errand in mind to alter the overlearned habit. Deficits in goal maintenance can impede life functioning in multiple functional domains, including employment, education, socializing, and recreation because it is required to complete tasks that necessitate responses to be modified based on differing contexts. Specific deficits in goal maintenance have been observed in SZ (Jones et al., 2010; Servan-Schreiber et al., 1996) and their unaffected relatives (Delawalla, Csernansky, & Barch, 2008; MacDonald et al., 2003).

To measure goal maintenance, Cohen and Servan-Schreiber (Servan-Schreiber et al., 1996) modified the traditional AX-CPT paradigm by changing the expectancy of AX pairings. The Dot Pattern Expectancy task further modified this paradigm by using dot patterns instead of letters, thereby enabling parametric manipulation of item difficulty by varying the similarity of target and non-target stimuli (MacDonald, Goghari, et al., 2005).

The DPX also addresses the issue that overlearned representation of letters might reduce the sensitivity of the letter-based expectancy AX-CPT. Moreover, the DPX has been shown to reliably measure goal maintenance (Jones et al., 2010; Strauss et al., 2013) and has been optimized for use with SZ by reducing the length of the task while maintaining its reliability (Henderson et al., 2012). Maximizing the efficiency and reliability of DPX also enhances its treatment utility, as shorter measures are less cumbersome to administer and less prone to participant fatigue.

Previous studies demonstrated that HC activated middle frontal gyrus (MFG) on trials of the expectancy AX task that required goal maintenance (Barch et al., 2001; MacDonald & Carter, 2003). Activation differences have been observed in this region when comparing HC with SZ (Barch et al., 2001; MacDonald, Carter, et al., 2005) and their unaffected relatives (Delawalla et al., 2008). Regions within MFG have been theorized to instantiate premotor representations based on external contextual cues accompanying stimuli (Koechlin, Ody, & Kouneiher, 2003), so hypoactivation in this region may indicate impairment in that ability. One previous report of DPX neuroimaging findings exists in HC (Lopez-Garcia et al., 2015), which showed activation of the same brain regions when completing the DPX as when performing the expectancy AX. However, the DPX has never been used to examine brain activation in SZ.

Using tasks such as the DPX to examine cognition and brain activation changes to treatment response would be facilitated if the task could be used successfully across many sites. Early studies of multisite fMRI (Casey et al., 1998; Ojemann et al., 1998; Vlieger, Lavini, Majoie, & den Heeten, 2003) found good reproducibility between sites, as did later multisite studies that included SZ (Ford et al., 2009; Schneider et al., 2007).

Multisite imaging allows for greater sample sizes and more power to detect group differences and treatment effects. Thus, the current study involved five CNTRaCS sites and included standardized imaging protocols to reduce between-site differences.

The use of imaging tasks to study treatment changes is also facilitated by having short and efficient protocols. Thus, to produce a more efficient and reliable measure of goal maintenance for evaluating treatment success, we sought to quantify a minimum length of fMRI scan capable of detecting group differences in goal maintenance. This is a practical question for future studies, as shorter scans may reduce participant fatigue, lower per subject costs, and allow for larger sample sizes. Therefore, the aims of the present study were threefold. First, we wished to replicate the finding of a specific deficit in goal maintenance in SZ compared with HC using this newly optimized DPX paradigm. Second, we wished to determine whether this task could be successfully implemented in a multisite context. Third, we wished to establish a recommended scan length to observe activation differences in groups of this size.

Methods

Subjects. Data were collected across five CNTRaCS sites. A complete methodology for the current study can be found in appendix 1. A complete subject recruitment protocol has been previously published (Henderson et al., 2012), and the current sample has been previously described (Ragland et al., 2015; Silverstein et al., 2015). The final sample consisted of 103 subjects (56 HCs, 47 SZs). There were no significant differences between included and excluded controls or patients on demographic, behavioral, or symptom indices (p 's > .08). The final groups were

demographically similar on age, and they did not differ on any measured demographic variable with the exception of education (see table 2.1). Subject groups did not differ on average relative or absolute head movement after removing subjects with excessive movement (both p 's > .45).

Dot pattern expectancy task and analysis. The Dot Pattern Expectancy task (DPX) has been described previously (Henderson et al., 2012; Jones et al., 2010). The task was performed in four blocks by each subject, with each trial consisting of a cue dot pattern followed by a probe dot pattern. One dot pattern was identified as a valid cue ('A' cue), and another as a valid probe ('X' probe). All other cues were invalid ('B' cues), and all other probes were invalid ('Y' probes). Besides the valid 'AX' target trials, three other possible combinations of cues and probes ('AY', 'BX', and 'BY') made up three distinct non-target trial types enabling the identification of a specific deficit in a subject's ability to maintain goal-relevant information throughout a trial. Each block of the DPX task consisted of 40 trials: twenty-four AX (60%), six AY (15%), six BX trials (15%), and four BY (10%).

For the DPX behavioral data, we employed two primary analyses. Groups were first compared using an independent samples t test on d' -context (Servan-Schreiber et al., 1996), a measure of general impairment on the DPX task. To establish a specific deficit, we fit a mixed effects logistic regression within a hierarchical model. Accuracy data were predicted using a small number of variables, with the minimum being the 'group' variable. Additional variables were added to the model, such as 'trial type (i.e., AX, AY, BX, BY)' and 'site membership' (i.e., which CNTRaCS site produced the data). Each

model was assessed using the Akaike Information Criterion (AIC) to determine the simplest model that predicted the data as well as or better than any other. Once a model was chosen, main effects and interactions of the variables were evaluated with a particular emphasis on ‘BX’ trial type and the comparison of ‘BX’ and ‘AY’ trials.

fMRI data acquisition and preprocessing. Three CNTRaCS sites used Siemens Trio 3 Tesla scanners (Minnesota, Washington University, UC Davis), one site used a Siemens Allegra 3 Tesla system (Rutgers), and the fifth site employed a Phillips 3 Tesla scanner system (MPRC). Scanning details can be found in the appendix 1. The scan session included the collection of four, 180-volume scans during four blocks of the DPX task. Quality control “phantom” scans were also collected on each scanner at the time of each subject’s data collection.

Preprocessing using FMRIB Software Library (FSL v. 4.1.8)(Smith et al., 2004) included motion correction (Jenkinson, Bannister, Brady, & Smith, 2002), brain extraction (Smith, 2002), prewhitening (Smith et al., 2004), high-pass temporal filtering with sigma of 100 s; B_0 field unwarping, spatial smoothing with a 5 mm FWHM Gaussian kernel, and spatial normalization and linear registration (Jenkinson & Smith, 2001) to the MNI 152 standard brain. Subjects with poor data quality were removed from the analysis (see appendix 1 for details).

General linear model. Following preprocessing, functional data were analyzed with a general linear model approach using the fMRI Expert Analysis Tool (FEAT) within the FSL software library. The following events from correct trials were modeled

for each subject: “A” Cues, “B” Cues, “AX” Probes, “AY” Probes, “BX” Probes, and “BY” Probes. Cue Errors and Probe Errors were also modeled, although they were not used in further analyses. Variation in the neural response was accounted for using the default FMRIB Linear Optimal Basis Set (FLOBS)(Woolrich, Behrens, & Smith, 2004) included with FSL.

Whole-brain analyses were performed at the group level in a voxelwise GLM analysis within FEAT. The primary contrast of interest at this group level was a comparison of SZ with HC on the lower level contrast of B Cue activation minus A Cue activation, although within-group analyses were also conducted to determine typical activation patterns for each group. Based on goal maintenance literature (Blackman et al., n.d.; Cohen & Servan-Schreiber, 1992; Lesh et al., 2013; Lopez-Garcia et al., 2015; MacDonald, Cohen, Stenger, & Carter, 2000; MacDonald & Carter, 2003; MacDonald, Carter, et al., 2005; Niendam et al., 2014), the contrast of B cues with A cues was chosen because B trials require the ability to maintain goal-relevant information to overcome the prepotent “target” response in the event of an X probe. Site membership as well as estimates of the data smoothness, signal-to-fluctuation-noise ratio (SFNR), average relative movement, and average absolute movement were also included as explanatory variables in the analysis. Their inclusion was intended to assess the effect of, and control for cross-site differences. A threshold of $z > 3.09$ and whole-brain corrected cluster extent threshold of $p < .05$ were employed for all group-level tests.

Scan length analysis. Additionally, we analyzed the functional data in a stepwise fashion to determine how long a scan must be to detect group differences in brain activation patterns. Data were analyzed with two and three scans (12 and 18 minutes, respectively), and the results of these were compared with the full, four-scan data analysis (24 minutes). Qualitative analysis of the results was used to indicate whether the effects seen in the full data analysis were present in the reduced data analyses. Additionally, quantitative analyses using the Dice coefficient (Dice, 1945) were conducted to measure the extent of overlap.

Results

Behavioral results. To investigate sensitivity to context on the DPX, we calculated the signal-detection metric d' -context and compared HCs to SZs using an independent samples t -test. Figure 2.1a shows that HCs had significantly higher d' -context scores ($M = 3.38, SD = 0.77$) than SZs did ($M = 2.80, SD = 0.97; t(86.99) = 3.30, p = .001$). The mixed-effects logistic regression included group and trial type as fixed explanatory variables and subject as a random variable. As displayed in Figure 2.1b, the interaction of group and trial type showed that compared with AY trials, HCs were significantly more accurate on BX trials than SZs ($z = -3.16, p = .002$).

To evaluate the association between psychiatric symptoms and performance on the DPX, we performed correlation tests between d' -context scores and BPRS positive ($M = 9.6, SD = 5.3$), negative ($M = 7.2, SD = 2.3$), and disorganization ($M = 4.9, SD = 1.7$) subscale scores, as well as total BPRS scores ($M = 40.3, SD = 10.1$) for SZs. No significant correlations were observed in this sample (all tests: $|r| < .19, p > .21$).

Behavioral efficiency analysis results. The mixed effects logistic regression model used with the whole dataset was also applied in a stepwise manner with one, two, and three blocks of the DPX behavioral data. As illustrated in Figure 2.2, the significant group by trial-type effect (BX compared with AY trials) was present regardless of the number of blocks used (one block: $z = -3.8, p < 0.001$; two blocks: $z = -3.8, p < 0.001$; three blocks: $z = -3.4, p < 0.001$; four blocks: $z = -3.2, p < 0.002$).

fMRI results. The fMRI analyses indicated there were no significant effects of site as measured by group level F tests of contrasts including CNTRaCS site as an explanatory variable. Additionally, no significant activation was observed in contrasts that included SFNR, smoothness, and relative or absolute movement estimates as explanatory variables.

The exploratory whole-brain analyses yielded significant differences in activation in the contrast of “B” cues with “A” cues. SZ displayed activation in left MFG and bilateral lateral occipital lobes, whereas HC activated in various regions of the cortex (peak in right lateral occipital lobe). When comparing groups, HC activated more compared to SZ in right MFG/frontal pole, left posterior parietal lobe, and left MFG/frontal pole, as displayed in Figure 2.3. No significant correlations were observed between activation and BPRS subscale scores (all tests: $|r| < .21, p > .16$). Full statistical results are presented in supplementary table 1.1.

fMRI efficiency analysis. In an effort to determine the minimum necessary scan length to detect experimental effects in groups of this size on the DPX task, we conducted a series of step-wise analyses. In the “B” cues minus “A” cues contrast (illustrated in Figures 4 and S1), the results showed remarkable consistency across all three analysis conditions (Dice coefficients of 0.72, 0.73, and 0.89). However, in the HC > SZ contrast illustrated in supplementary figure 1.2, there were more varied results. There were no significant group differences in the 12-minute analysis. In the 18-minute analysis, there was some overlap in group differences in activation in the left frontal lobe with the 24-minute analysis; however, the 18-minute results included left parietal activation not seen in the 24-minute analysis (Dice coefficient of 0.40).

To rule out the possibility that the efficiency results were driven by changes as the study progressed as opposed to the amount of data in the analysis, we performed a sliding window analysis. Specifically, we analyzed two blocks at a time according to the following chunks: first and second, second and third, third and fourth. Again, the significant activation associated with the “B” cues greater than “A” cues contrast retained its consistency and had Dice coefficients greater than .78 among them. The maps of HC and SZ activation to this contrast are presented in supplementary figure 1.3.

Discussion

The Dot Pattern Expectancy task (DPX) was previously established as a reliable measure of goal maintenance that can discern specific deficits in SZ compared with HC (Henderson et al., 2012; Jones et al., 2010; Strauss et al., 2013). The current study is the first to employ fMRI and the DPX to determine the neural underpinnings of goal

maintenance deficits in SZ compared with HC and was able to do so across five sites with different scanners. The study replicated previous behavioral findings (MacDonald, Carter, et al., 2005) of a specific deficit in goal maintenance in SZ compared with HC and found the deficit to be consistent across blocks. The study also replicated the findings from previous research with drug-naïve patients performing the expectancy AX task (MacDonald, Carter, et al., 2005) who showed reduced activity in middle frontal cortex during goal maintenance. The current study showed lower activity in SZ compared with HC in an executive/attention network consisting of bilateral frontal pole/MFG as well as left posterior parietal lobe. Finally, a recommended scan length was estimated for future studies employing samples of a comparable size.

In terms of behavior, SZ demonstrated the hypothesized specific deficit in goal maintenance previously observed on the DPX (Henderson et al., 2012; Jones et al., 2010) and the expectancy AX (MacDonald et al., 2003; MacDonald, Carter, et al., 2005; Servan-Schreiber et al., 1996). This group difference was first observed after only 6 minutes (40 trials) and it remained throughout the administration (24 minutes; 160 trials total). This result suggests the DPX is sufficiently sensitive for efficient studies of goal maintenance in this population. Further, the analysis suggests this is not merely the result of differential effects of learning or fatigue, allowing us to examine the efficiency of the imaging analysis with knowledge that the behavioral effect was consistent over time. For reference, a previous imaging study of expectancy AX in schizophrenia patients (MacDonald, Carter, et al., 2005) had a task duration of 40 minutes as compared to 24 minutes for the longest analyses conducted in the current study.

FMRI analyses showed that HC had larger activation differences between “B” cues and “A” cues compared with SZ in bilateral MFG. These results agree with previous literature with regard to the MFG’s importance in successfully utilizing goal maintenance and SZs’ deficits in that region. Hypofrontality in left (MacDonald & Carter, 2003) and right (MacDonald, Carter, et al., 2005; Perlstein, Dixit, Carter, Noll, & Cohen, 2003) MFG has been reported in SZ during the performance of the expectancy AX task.

In addition to group differences in bilateral MFG, differences in superior/posterior parietal lobes were also detected. There are several possible mechanisms that could lead to the increased parietal activation during B cues compared with A cues. The dorsal visual stream has been associated with processing spatial relationships (including dot patterns; Vogels, Sary, Dupont, & Orban, 2002), and therefore HC’s increased activation to B cues may reflect increased processing related to identifying or categorizing the group of “B” patterns, that are less familiar, compared to patients. These regions are also associated with visuomotor control (Milner & Goodale, 2012), and an alternative explanation is that B cue-related activity may reflect the preparation needed to inhibit a prepotent response, activity that is reduced in schizophrenia. A third explanation, which is not entirely distinct, is that this difference in posterior parietal activation reflects its role in a more general executive functioning network (Blackman et al., n.d.; Minzenberg, Laird, Thelen, Carter, & Glahn, 2009). This region’s covariation with dorsolateral prefrontal cortex reflects the demands for the representation and maintenance of contextual goals. The relative contributions of visual cognitive, response preparation, and goal maintenance functions to the parietal activation observed during the DPX are in need of further clarification. Together with MFG, these regions are consistent with a

visuospatial reasoning and attentional control network (Dosenbach et al., 2007; Laird et al., 2011). This network is thought to be integral to top-down control and managing responses given changing demands. Left MFG and posterior parietal cortex compose an executive network has been associated with language tasks, as well as working and explicit memory tasks. SZ have demonstrated disrupted functional (Deserno, Sterzer, Wüstenberg, Heinz, & Schlagenhaut, 2012) and white matter connectivity (Wang et al., 2013) in this network, and these changes were associated with deficits in working memory and performance IQ, respectively. Functional dysconnectivity in a frontoparietal network has previously been observed in SZ and their healthy first-degree relatives while performing the expectancy AX task (Poppe, Carter, Minzenberg, & MacDonald, 2015). Although the present study does not identify a network per se, it does implicate the same regions as being deficiently activated by SZ.

We also observed group differences in bilateral frontal pole, a region that has been shown to underpin the maintenance, monitoring, and processing of subgoals during a working memory task (Braver & Bongiolatti, 2002; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999). The frontal pole is theorized to perform this action as one facet of a cascading executive system whereby information from the environment provides contextual cues that are interpreted and acted upon to achieve some goal (Koechlin & Summerfield, 2007).

The results observed in the present study were robust to site effects. No significant effects of site were observed in the fMRI analyses. This finding highlights the practicality of combining imaging data from multiple sites, thereby allowing for larger sample sizes in fMRI patient studies.

The efficiency analyses undertaken to establish a recommended scan length for future fMRI studies of the DPX task with similar sample sizes found that 18 minutes of scanning is required to observe group differences. However, the within-group task activation patterns seen in the full 24 minutes could be observed almost undiminished in half that time. It was further determined that these effects are unlikely due to changes in the scanning session over time, as there were few if any differences among a set of three sliding window analyses employing two blocks of data each.

Limitations. Previous research (Lesh et al., 2013; A. W. MacDonald, Carter, et al., 2005) demonstrated an association between BOLD activation in dlPFC/MFG and disorganization symptoms in SZ, but no such association was observed in the current study. However, other studies of the AX-CPT in schizophrenia either did not observe a correlation between dlPFC activity and disorganization symptoms (MacDonald & Carter, 2003; Niendam et al., 2014; Yoon et al., 2008) or did not report any such correlations (Barch et al., 2001; Holmes et al., 2005; Lesh et al., 2015; Perlstein et al., 2003). Likewise, no association between BOLD activation and accuracy on the DPX task was observed.

Conclusions and future directions. The present study, the first imaging study of the DPX task in schizophrenia, provides support for the task as a cross-site probe of goal maintenance-related activity of the MFG and other related executive control regions. It replicated previous studies showing a specific deficit in goal maintenance in schizophrenia. The imaging analyses replicated previous findings of MFG hypoactivation

and also found significantly less parietal activation in addition to frontal regions when comparing SZ with HC on context-intensive trials of the DPX task. This was also a multi-site study that incorporated data from five sites and different scanners and found no significant effects of site in the imaging analyses. The behavioral and imaging efficiency analyses showed that the DPX is an efficient tool for assessing goal maintenance ability in imaging studies. Of interest for future research are questions about the reliability of the BOLD activations and group differences observed in the current study, as well as whether there are functional connectivity differences that may explain performance on the DPX task. Also of import is whether there is plasticity in the regions or networks that underlie goal maintenance and whether such regions and networks may act as targets for training or pharmacological treatment in the future. Treatments such as cognitive remediation (Ramsay & Macdonald, 2015) or cognitive enhancing medication (Nikiforuk, Hołuj, Kos, & Popik, 2016) might be capable of ameliorating impaired network activity underlying goal maintenance deficits in SZ, which could then improve goal maintenance ability and functional outcomes. We hope this work is useful in future endeavors to answer those questions.

Table 2.1 Demographic and Clinical Characteristics.

	Group		Test
	Patients	Controls	
N	47	56	
Mean Age (yrs)	35.6 (12.1)	34.8 (11.9)	$t(101) = 0.33$
% Male	74.5	75	$\chi^2(1) = 0.00$
% Caucasian	55.3	62.5	$\chi^2(1) = 0.29$
% Right-Handed	85.1	83.9	$\chi^2(1) = 0.00$
Mean Education (yrs)	13.9 (2.0)	15.3 (2.6)	$t(101) = -2.89^*$
Premorbid Functioning ^a	36.3 (9.6)	37.5 (10.6)	$t(99) = 0.58$
Mean Parental Education (yrs)	14.0 (2.5)	13.8 (2.7)	$t(92) = 0.33$
BPRS Total	40.3 (10.1)	n/a	
Positive Symptoms ^b	9.6 (5.3)	n/a	
Negative Symptoms ^c	7.2 (2.3)	n/a	
Disorganization ^d	5.0 (1.7)	n/a	
Antipsychotic Meds			
Typical/Atypical/None	2/44/1	n/a	

Note: BPRS refers to the Brief Psychiatric Rating Scale. Parenthetical numbers following means represent standard deviations. Asterisks following test statistics represent $p < .05$.

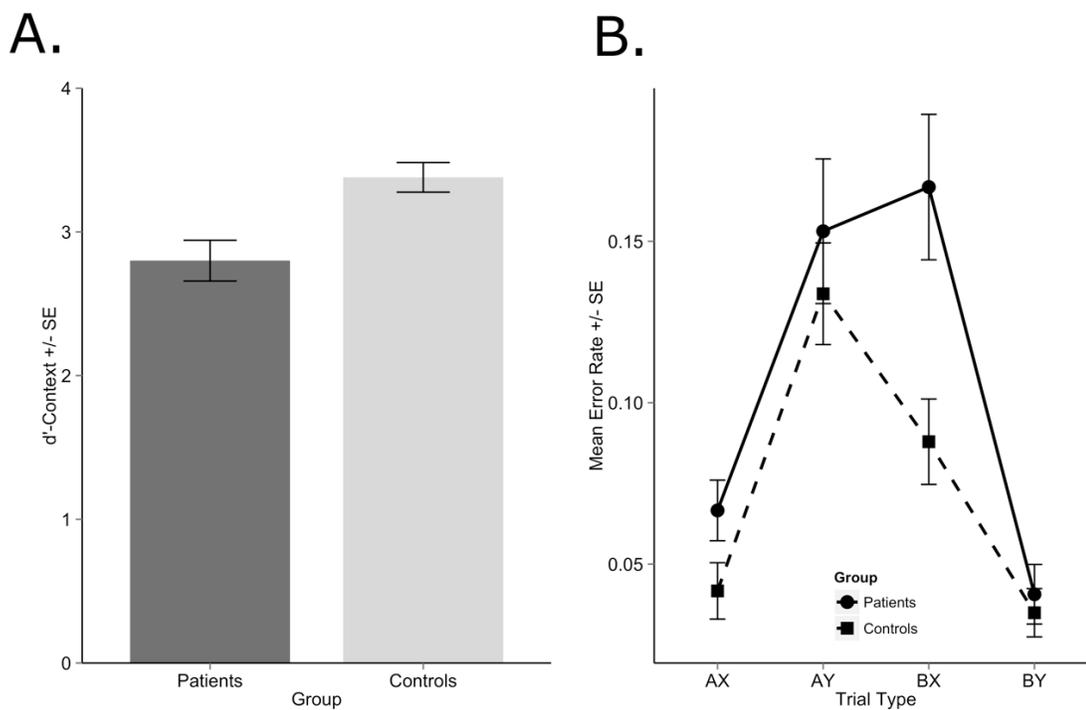
^a Wechsler Test of Adult Reading (Wechsler, 2001)

^a BPRS items 8, 9, 10, and 11.

^b BPRS items 13, 16, 17, and 18.

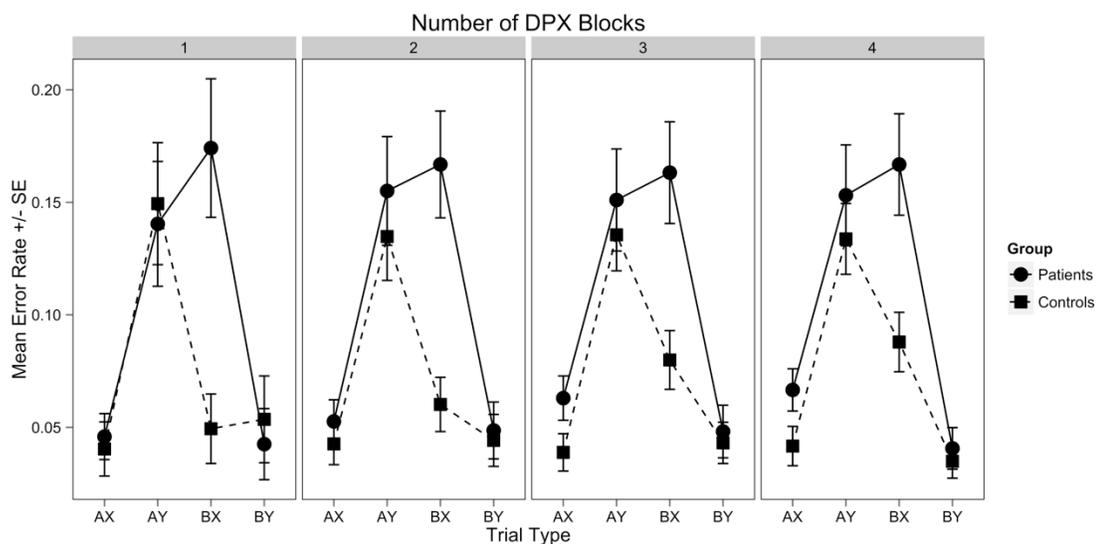
^c BPRS items 12, 14, 15, and 24.

Figure 2.1. DPX Behavioral Results.



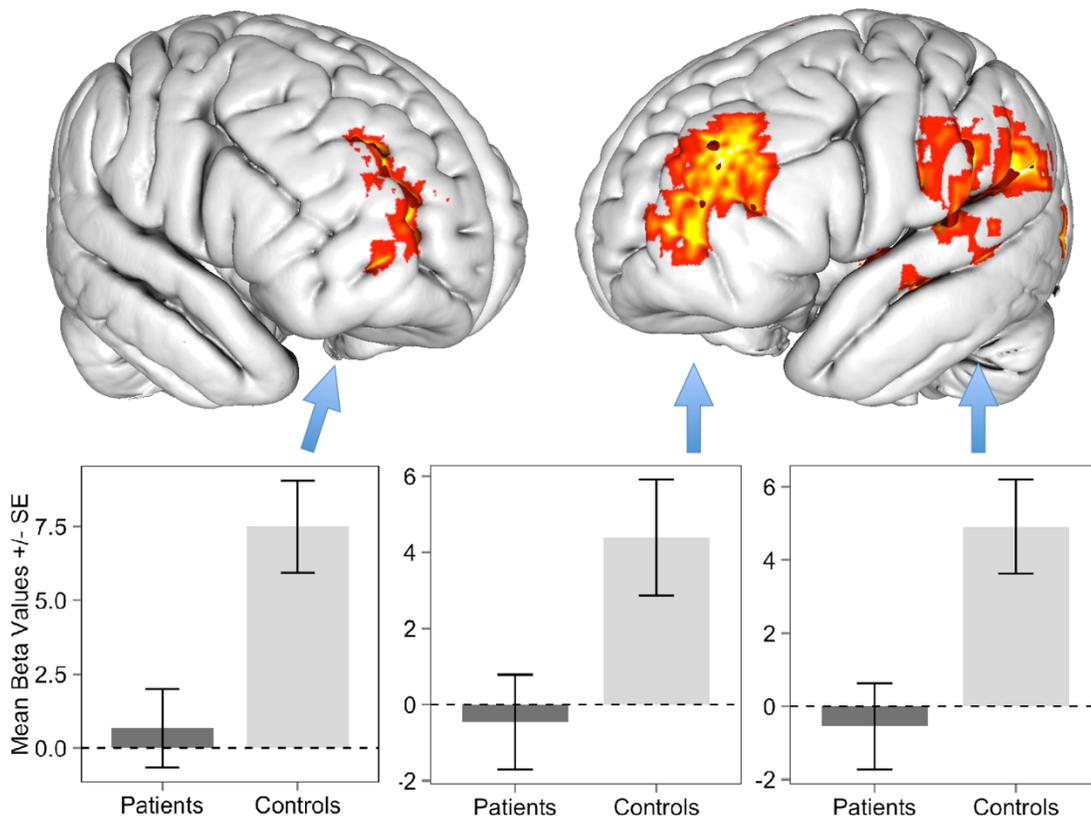
A. d' -context scores on the DPX task. B. Error rates separated by group, trial type. Error bars represent standard error of the mean.

Figure 2.2. Behavioral Efficiency Analysis.



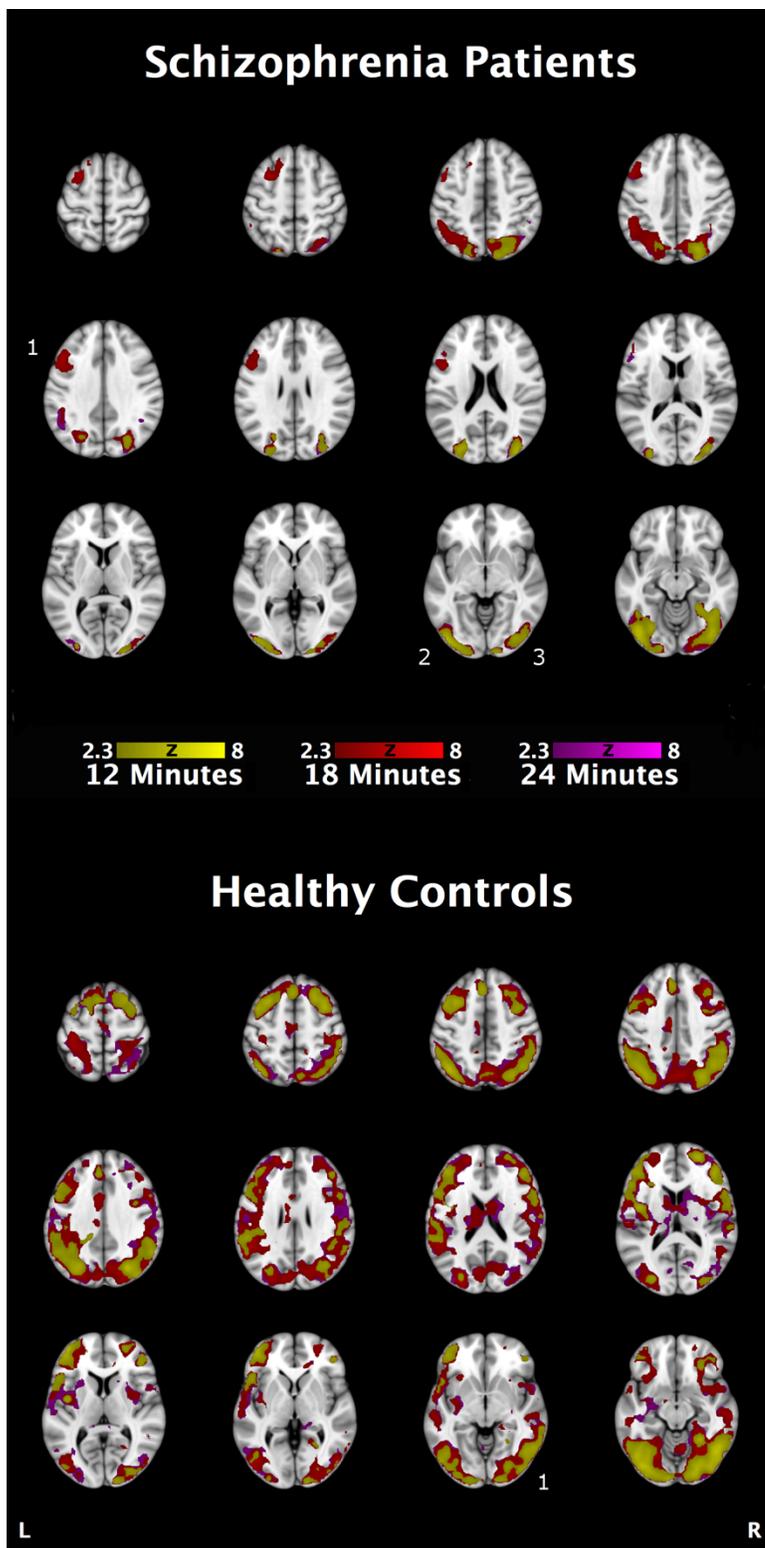
Error rates on the DPX task calculated from four amounts of data. The first figure represents only the first block (40 trials) of data, the second represents the first and second blocks (80 trials), *et cetera*.

Figure 2.3. Whole brain fMRI GLM results.



Beta values represent B Cues - A Cues contrast. Regions with greater activation in HC than SZ (3 clusters). Cluster 1 had a peak voxel Z score of 4.28, volume = 11,056 mm³, and MNI coordinates (x,y,z) of 26, 54, 14. Cluster 2 had a peak voxel Z score of 4.43, volume = 30,664 mm³, and coordinates of -40, 4, 12. Cluster 3 had a peak voxel Z score of 4.31, volume = 23,408 mm³, and coordinates of -50, -40, 22.

Figure 2.4. fMRI Efficiency Analysis, B Cues – A Cues.



Top portion displays SZ activation (3 clusters). Cluster 1 had a peak voxel Z score of 4.39, volume = 10,152 mm³, and MNI coordinates (x,y,z) of -50, 12, 42. Cluster 2 had a peak voxel Z score of 6.09, volume = 39,408 mm³, and coordinates of -44, -82, -10. Cluster 3 had a peak voxel Z score of 5.91, volume = 48,856 mm³, and coordinates of 38, -76, -16. The bottom displays HC activation(1 cluster). That cluster had a peak voxel Z score of 6.78, volume = 495,912 mm³, and coordinates of 46, -66, -14. The scan lengths that define the color of activation are cumulative, such that 18 minute data includes 12 minute data, and 24 minute data includes the previous two.

Chapter 3: Tonic and Dynamic Functional Network Connectivity of Salience and Executive Networks: Relation to Goal Maintenance Deficits in Schizophrenia

Foreword: This chapter was written in collaboration with Angus W. MacDonald and Deanna M. Barch, who provided edits and revisions to earlier drafts. The data collection methods in this chapter are largely redundant with those outlined in Chapter 2. Methods specific to this chapter can be found under within this chapter as well as in appendices 1 and 2.

Abstract

Cognitive control deficits in schizophrenia patients (SZ) may be caused, at least in part, by neural dysconnectivity. Recently, the salience (SN) and executive control (ECN) networks have been proposed to facilitate cognitive control through complementary mechanisms, and SZ have shown abnormal connectivity in those networks. Some evidence suggests coordination between the ECN and SN is critical for cognitive control, in addition to functioning of the networks individually. This study is the first to examine the dynamic relationship between brain networks and their relationship with goal maintenance in SZ and healthy controls (HC). Fifty-four HC and 46 SZ performed the dot pattern expectancy (DPX), a goal maintenance task, while functional magnetic resonance images were acquired. Data underwent group independent component analysis followed by tonic (tFNC) and dynamic (dFNC) functional network connectivity to examine stationary and dynamic temporal relationships between networks. Groups differed on five tFNC connections, and tFNC strength between SN and right ECN

predicted DPX performance in SZ. dFNC between left and right ECN components was significantly related to the DPX task demands. Groups differed in connectivity strength here (HC>SZ), which predicted DPX performance in HC. These tFNC findings suggest SZ fails to signal salient task information from the SN to the ECN. Likewise, the dynamic coordination between left and right ECN during task completion predicted goal maintenance in HC and was deficient in SZ, which argues for the importance of the transitory interaction between multiple networks for successful cognitive control in HC and SZ.

Introduction

Schizophrenia patients (SZ) exhibit deficits in various cognitive abilities, and these deficits represent some of the most debilitating and difficult to treat symptoms (Choi et al., 2013; Gold & Harvey, 1993). Schizophrenia is hypothesized to be a disorder of dysconnectivity, and this disrupted communication amongst brain regions is thought to cause the cognitive deficits observed in the disorder (see Zhou, Fan, Qiu, & Jiang, 2015 for review). One construct encompassing many of the cognitive domains affected by schizophrenia is cognitive control, which includes several related control functions including directing attention to relevant stimuli, ignoring irrelevant events, retrieving important information from memory, and maintaining and utilizing that information to achieve some goal (Suchy, 2009).

Two large-scale networks have been theorized to underpin cognitive control: the salience (SN) and central executive (ECN) networks (Menon, 2011; Seeley et al., 2007). The SN is made up of regions including anterior insula, dorsal anterior cingulate cortex,

anterior prefrontal cortex, medial superior frontal cortex, and frontal operculum, and it is thought to detect and orient to meaningful stimuli and to mediate the flow of salient information to networks involved with attentional processing (Menon & Uddin, 2010). The ECN is composed of middle frontal gyrus (MFG) and lateral parietal cortices (Dosenbach et al., 2008), and it is theorized to update attention and exert top down control of behavior in response to feedback.

ECN and SN have long been associated with working memory and cognitive control (Nowrangi, Lyketsos, Rao, & Munro, 2014), and previous studies found that the strength of functional connectivity within them predicts performance and trait measures (Seeley et al., 2007; Song et al., 2008). Structures within the ECN and SN have been hypothesized to form a cognitive control loop that guides behavior during a task (Miller & Cohen, 2001), and SN structures play a critical role in activating the ECN and recruiting networks based on task demands (Sridharan, Levitin, & Menon, 2008).

Patients with schizophrenia have demonstrated several deficits in these networks in previous research. Grey matter volume deficits in SN structures correlate with severity of reality distortion symptoms in SZ (L Palaniyappan, Mallikarjun, Joseph, White, & Liddle, 2011), and resting state functional connectivity between SN and ECN is reduced in SZ (Manoliu et al., 2014; Moran et al., 2013; Lena Palaniyappan, Simmonite, White, Liddle, & Liddle, 2013). Although the interplay between ECN and SN may be related to individual differences in cognitive control and to cognitive control deficits in SZ, few studies have examined this relationship during the completion of cognitive tasks or assessed how the relationship between these networks changes with task demands, such as goal maintenance.

Goal maintenance refers to the ability to maintain, manipulate, and utilize context- and goal-relevant information during the performance of a task and has been shown to be impaired in SZ (Jones et al., 2010; Servan-Schreiber et al., 1996) and their unaffected first-degree relatives (Delawalla et al., 2008; MacDonald et al., 2003). Activation studies of healthy controls (HC) during goal maintenance tasks have found bilateral MFG activation (Barch et al., 2001; MacDonald & Carter, 2003). SZ have demonstrated reduced activity in left (MacDonald & Carter, 2003) and right (MacDonald, Carter, et al., 2005; Perlstein et al., 2003) MFG compared with HC during goal maintenance in activation studies. An independent component analysis (ICA) of goal maintenance found the time course of a left MFG and left posterior parietal lobe component was less related to the task demands in SZ than in HC (Poppe et al., 2015). These extant findings shed light on dysfunction within several regions involved with the ECN and SN, but not how these networks interact. In addition to FC, which finds brain networks based on temporal similarity, functional network connectivity (FNC) measures how those networks are themselves temporally related. Tonic FNC (tFNC; Jafri *et al.*, 2008) quantifies the stationary temporal relationship between networks across an entire scan, whereas dynamic FNC (dFNC; Sakoğlu et al., 2010) measures how the relationship between networks changes over time.

The tools of tFNC and dFNC are suited to answer questions about the ECN and SN and their involvement with goal maintenance deficits because they can measure both how networks are related across an entire scan (tFNC) and how that relation updates throughout a scan (dFNC). Both tFNC and dFNC have been employed in SZ imaging studies, although often these studies scan subjects at rest (Damaraju et al., 2014; He et al.,

2015; Magnuson et al., 2015; Meda et al., 2012). To our knowledge, no study has examined how networks interact in schizophrenia during goal maintenance. The present study goals were to assess tFNC and dFNC of networks derived using group ICA in SZ and HC during the Dot Pattern Expectancy task (DPX; MacDonald, Goghari, et al., 2005), a measure of goal maintenance. It is expected that SZ will exhibit deficient tFNC and dFNC involving ECN and SN components and that the extent of dysconnectivity will predict impairment on the DPX task.

Method

Subjects. The data for the current study were collected under the protocol of the Cognitive Neuroscience Test Reliability and Clinical applications for Schizophrenia consortium (CNTRACS), which includes five data collection sites: University of California at Davis, Maryland Psychiatric Research Center at the University of Maryland (MPRC), Rutgers University, University of Minnesota, and Washington University in St. Louis. The protocol for subject recruitment has been detailed previously (Henderson et al., 2012), and the current sample is the same as that used in previous studies (Ragland et al., 2015; Silverstein et al., 2015). The final sample consisted of 100 subjects (54 HC, 46 SZ). As displayed in supplementary table 1.1, groups did not differ on any demographic variable with the exception of education. Groups were not significantly different with regard to absolute or relative head movement after removing subjects (all $p > .21$). Specific inclusion/exclusion details for the current sample can be found in appendix 1.

Dot pattern expectancy task. Subjects performed four blocks of 40 trials of the Dot Pattern Expectancy task (DPX). Each trial was composed of cue-probe pairs of 500 ms duration with inter-stimulus intervals ranging from 2500 to 3500ms and inter-trial intervals ranging from 2500 to 12500ms. A particular dot pattern was designated as a valid cue (“A”) and all other cues were invalid (“B”). Likewise, one dot pattern was designated as a target probe (“X”) and all other probes were non-target (“Y”). Subjects were instructed to respond with a “target” button press in the event that an “A” cue was followed by an “X” probe and to respond with a “non-target” button press in all other circumstances. The three other possible combinations of cues and probes (“A-Y”, “B-X”, and “B-Y”) enabled the identification of a specific deficit in a subject’s ability to maintain goal-relevant information throughout a trial. Twenty-four trials (60%) were A-X trials, six trials (15%) were A-Y trials, six trials (15%) were B-X trials, and four trials (10%) were B-Y trials.

fMRI data acquisition and preprocessing. Scanners used for data acquisition consisted of three Siemens Trio 3 Tesla scanners (Minnesota, Washington University, UC Davis), one Siemens Allegra 3 Tesla (Rutgers), and one Phillips 3 Tesla scanner (MPRC). A standard acquisition and preprocessing procedure was employed. Details can be found in appendices 1 and 2.

Independent component analysis and functional network connectivity. Following preprocessing, functional data were analyzed with a meta-ICA pipeline described in detail in previous studies (Poppe et al., 2013; Wisner et al., 2013). Briefly, the approach

consists of running probabilistic group ICA multiple times within the Multivariate Exploratory Linear Optimized Decomposition into Independent Components tool (MELODIC; Christian F. Beckmann & Smith, 2004), each time randomizing the subject order in which the data are entered into MELODIC. A final ICA is then performed with the results of the previous step used as input data. The choice for the number of repetitions of the lower level ICAs (25) as well as the number of components specified at each stage (60) was based on previous research demonstrating optimization using these parameters on various metrics (Poppe et al.).

The spatial maps resulting from the meta-ICA were then used in dual regression (Beckmann, Mackay, Filippini, & Smith, 2009) to produce subject-specific spatial maps and time courses for each of the 60 components. Artfactual and noise components were next identified by visual inspection and removed from subsequent analyses. These noise components represented head motion, white matter, and ventricular activity. Twenty-five components were removed, leaving 35 components in the following analyses.

Task-relation of independent components. First, correlations were calculated between the subject-specific time courses for each component produced by dual regression and the time course of B cues from the DPX task for each subject. This resulted in 35 correlation coefficients per subject. These correlation coefficients were then Fisher- z transformed (Fisher, 1915), and one-sample t tests were performed to determine which correlation values were significantly different from zero. The p values of these t tests were corrected for multiple comparisons by multiplying each p value by the number of tests (35). For all significant comparisons, independent samples t tests were performed between subject

groups to determine if groups differed in the extent to which the component correlated with the DPX task time course. The p values from these t tests were corrected by multiplying each value by the number of tests.

Tonic functional network connectivity. The tFNC analysis was based on previous work (Jafri et al., 2008). The process involved calculating the maximal constrained lagged correlation between each pair of component time courses for each subject. First, time courses were interpolated using a cubic spline (Forsythe, Moler, & Malcolm, 1977) by a factor of four, allowing for sub-TR temporal precision. Next, Pearson product moment correlation coefficients were computed between each pair of components at each “level” of lag between -5 and 5 seconds. This resulted in 21 correlations for each pair of components, and the largest of those correlations in absolute magnitude was selected as the final correlation value between those two components for that subject. Correlation values were then transformed using a Fisher z transform to allow statistical procedures to be performed across subjects.

To determine which pairs of components were significantly correlated, one-sample t tests were performed for each component pair across groups. The p values for these tests were corrected by multiplying by the number of component pairs (35 choose 2 combinations, or 595), with p values that remained less than .05 being considered associated with significant differences. To detect group differences on the magnitude of correlation between component time courses, independent samples t tests were performed between HC and SZ for each pair of components. The p value from each of these tests

was recorded and multiplied by 595 to correct for multiple comparisons. Those p values that remained less than .05 were determined to signify significant group differences.

To determine if tFNC predicted performance on the DPX task, tFNC values of any connection determined to be significantly different between groups were correlated with a measure of task performance called d' Context (Servan-Schreiber et al., 1996). Correlation coefficients were also calculated between these tFNC values and psychiatric symptom measures in SZ. We next assessed whether tFNC connections with group differences were significantly correlated with psychotic symptom measures in SZ.

Dynamic functional network connectivity. dFNC, based on the work of Sakoğlu and colleagues (2010), is a method of evaluating how the relationship between any two components changes over the course of a scanning session. First, for each subject and each component pair, the optimal lag value obtained during tFNC is used to pre-lag the components relative to one another. Next, a sliding window approach is used to calculate correlation coefficients along the length of the scan. The width of this window was set at 15 seconds, and the window moved 4 seconds at each step. This procedure is explained in supplementary figure 2.1. The dFNC analyses results in a set of correlations (one for each 15 second sliding window) between each pair of networks identified in the initial ICA.

To determine if variation in these dFNC values across the sliding windows of the task were related to the DPX task, a task demands metric was produced. First, the boxcar function of B cue onsets from the DPX task was convolved with a double gamma hemodynamic response function. Next, a moving average of this B-cue timeseries was computed with the same parameters as the sliding window utilized in dFNC. This B-Cue

task demand time-series for each subject was then correlated with each component pair's dFNC values. This resulted in 595 correlation values for each subject, each indicating the extent to which each component pair's dFNC values were associated with that subject's B-Cue task demands. One sample t tests were conducted across subjects for each of these 595 component pairs, and the p values from these t tests were multiplied by 595 to correct for multiple comparisons. Any tests whose p values remained less than .05 were considered statistically significant. Independent samples t tests were conducted between subject groups for any component pair that was determined to be statistically significant in the previous step. The p values from these tests were multiplied by the number of independent samples t tests conducted to correct for multiple comparisons.

In order to ensure that results were not spurious, a resampling technique was used to produce a null distribution of p values. This distribution was created by randomizing the task strength function that was matched with each set of dFNC values, such that one subject's dFNC values were correlated with another subject's task demands values. This process was repeated 10,000 times. The obtained p value from the independent samples t test described above was then compared to this distribution of p values.

We next assessed if performance on the DPX task was related to the extent to which dFNC was correlated with B-Cue task demands. First, the correlation coefficients between dFNC and task demands were transformed with a Fisher z transform and then correlated with the d' context values described previously. Last, we determined if the extent to which dFNC correlated with task demands was itself correlated with BPRS symptom measures in SZ.

Results

Task-relation results. As presented in table 3.1 and in supplementary figure 2.2, 19 components were found to be significantly linearly related to the DPX B-cue task time line across subject groups. Groups did not differ significantly on any of these components after multiple comparison correction (all corrected $p > .05$).

Tonic functional network connectivity results. Of all component connections, 556 (93%) had tFNC values that were significantly different from zero, as presented in figure 3.1. Due to the high proportion of significant connections, all connections were tested for group differences in tFNC magnitude. As shown in table 3.2 and figure 3.2, five connections were found to be significantly different between HC and SZ. In all cases, the FNC values were positive and HC had higher magnitude FNC than SZ did.

Each of the connections that showed group differences was tested to determine if the strength of the connection was correlated with d' Context, a signal detection measure of task performance on the DPX task. These correlations were conducted in each subject group separately, resulting in ten total correlation tests. Of the five connections, only the connection between a right ECN component and a SN component was significantly correlated with d' Context after correcting for multiple comparisons, and this relationship only existed in SZ, as displayed in figure 3.3A. All other correlations were non-significant (all corrected p 's $> .05$). No connection was significantly related to BPRS symptom measures (all p 's $> .05$).

Dynamic functional network connectivity. There was only one connection that was significantly related to the DPX task demands metric after multiple comparisons correction. As shown by the green connection in figure 3.2, this connection involved components 9 and 34, which are left and right ECN components involving bilateral MFG and bilateral posterior parietal lobes (Brodmann areas 6, 7, 8, and 40). The average correlation between these components' dFNC and DPX task demands was 0.03 with a standard deviation of 0.08 ($t(99) = 4.07, p \text{ (corrected)} < .05$). A resampling procedure showed that the obtained r value was smaller than 9993 out of 10,000 values in the null distribution, corresponding to a p value of .0007. As shown in supplementary figure 3.3, HC ($M = .05, SD = .08$) had significantly higher correlations in this connection than did SZ ($M = .01, SD = .08; t(98) = 2.28, p = .012$). As displayed in figure 3.3B, task performance was predicted by the extent to which the dFNC of this connection correlated with the DPX task time course for HC but not for SZ. The dFNC between the ECN components did not correlate with any BPRS symptom measures in SZ (all p 's $> .05$).

Discussion

The current study utilized group ICA as well as tonic and dynamic functional network connectivity to probe the neural underpinnings of goal maintenance performance in schizophrenia. This represents the first study to examine the dynamic interrelationships between brain networks and assess their task-relatedness in schizophrenia during the performance of the DPX task. Measuring the associations between networks and how those associations changed over the course of the scan allowed an examination of how the ECN and SN operated during a goal maintenance task. The results of the current

study shed new light on how ECN and SN operate in HC and how that operation is disrupted in SZ.

Task-relatedness. Nineteen components' time courses were found to be significantly related to the DPX task timeline in both groups, including left and right ECN components; however, groups did not differ in the extent of task-relatedness of any of these components. These results differ from those of Poppe and colleagues (2015), who found that the extent to which a right ECN component correlated with a goal maintenance task's timeline was significantly reduced in SZ relative to HC. Although SZ in the current study did exhibit a lower correlation between right ECN and the DPX task timeline than HC, the difference in correlation was non-significant. This discrepancy may owe to differences in task characteristics between the expectancy AX task and the DPX task, in that the dot patterns in the DPX engaged the more spatial right hemisphere than did the letters in the AX task. Additionally, analysis procedures differed somewhat between the two studies, the largest difference being the utilization of separate analysis software employing different ICA algorithms. Considering the FNC results in the current study, the lack of any one component's significant group difference in task-relatedness may demonstrate that temporal relationships between networks are more important to task performance than the activity of any singular region or network. This idea may also help to explain the seemingly contradictory findings of hyper- and hypofrontality in SZ activation studies.

Tonic functional network connectivity. Five connections between eight components were found to be significantly different between subject groups. In each case, the strength of the connection was higher in HC than in SZ. Of note, a connection between right ECN and a SN component was found to be significantly stronger in HC than SZ. Recent findings have highlighted the importance of communication between ECN and SN. Cohen and colleagues (2014) found that functional connectivity between ECN and SN increased during high cognitive control conditions of a working memory task. White and others (2010) found a strong direct connection between ECN and SN in both SZ and HC but did not find group differences. Wallis and colleagues (2015) found the SN to be transiently active during specific portions of a working memory task, which led them to conclude that the SN is more dynamic in its role in ongoing cognitive control. The results of the current study suggest that SZ may have deficits in representing salient stimuli and conveying that information to ECN for processing.

Although previous studies have theorized that the SN (especially insula) is responsible for switching between the antagonistic ECN and DMN (Menon & Uddin, 2010; Sridharan et al., 2008), the results of the present study do not support that view. For example, the tFNC connection between SN and posterior DMN (components 40 and 4 respectively) was one of the few that was not significantly different from zero across groups.

The unilateral ECN connection with SN may derive from the theorized specificity of functioning of the right prefrontal cortex compared with left. There is evidence for a process-based organizational structure for the prefrontal cortex in addition to the domain-based structure of primarily verbal functions associated with left hemisphere and spatial

functions associated with right hemisphere (Ambrosini & Vallesi, 2015; Stuss, 2011). In this model, the left PFC has a task-setting role in which it is transiently engaged in forming and selecting task rules and disregarding irrelevant stimuli, whereas the right PFC has a monitoring role in which it is maintaining representations important to the task demands. Given this role for right PFC, the association between right ECN and SN may indicate a continuous conveyance of salient information to the right ECN for processing.

Dynamic functional network connectivity. The dFNC analysis determined that the coordination between left and right ECN components was significantly related to the DPX task timeline, and the strength of this correlation was significantly lower in SZ than in HC. This suggests that SZ had difficulty coordinating the activity of the two CENs to meet task demands. As indicated by the task accuracy results, the root of this difficulty may lie in the dysconnectivity between right ECN and SN.

Previous studies have found associations between right and left ECN. Assaf and colleagues (2010) found that the strength of tFNC between left and right ECN predicted performance on a semantic memory task in healthy subjects, while Arbabshirani and Calhoun (2011) found an increase in the variability of tFNC in SZ compared with HC. However, as the current study demonstrates, the overall correlation between left and right CENs may not be critical for cognitive control, but rather it is the transient coordination of the two networks when needed by task demands.

FNC correlated with task accuracy. The strength of the connection in the current study between SN and right ECN was correlated with performance on the DPX

task in SZ but not in HC. Meanwhile, the extent to which the dFNC between left and right CENs correlated with the DPX task timeline predicted task performance in HC, but not in SZ. This set of results implies that SZ have difficulty both communicating salient information to ECN and also processing that information within ECN. The lack of significant dFNC task performance prediction may mean that SZ's dysfunctional tFNC between ECN and SN so disrupted their goal maintenance capability that the effect of dFNC between ECN components was suppressed. In other words, one speculative hypothesis is that adequate stable communication of salience from SN to ECN is a necessary but not sufficient prerequisite for accurate DPX performance, and if that threshold is passed then performance depends on the dynamic cooperation of left and right CENs during task completion. Thus, if at least some individuals with schizophrenia had impairments in the connectivity between SN and ECN, it may have precluded the ability to see the contribution of connectivity between left and right ECN to performance.

Conclusions and future directions. The current study is the first to examine the dynamic associations between brain networks derived using group ICA of fMRI data collected from SZ and HC during the performance of the DPX task. It was found that SZ possessed significantly lower tFNC between right ECN and SN and that those values predicted task performance on the DPX task in SZ but not HC. The dFNC analysis revealed that the extent to which dFNC between left and right ECN was correlated with the DPX task time course was significantly lower in SZ than in HC. This correlation predicted task performance in HC but not SZ. These results suggest that both stable connectivity between ECN and SN as well as dynamic cooperation between left and right

ECN contribute to top down cognitive control during goal maintenance. Of interest for further research are questions of the reliability of these FNC results. Additionally, a future direction may be to make the tFNC between right ECN and SN a target for training or pharmacological treatment.

Table 3.1. Components Significantly Correlated with Time Course of DPX B Cues Across Groups.

Comp.	Regions	Brodmann Areas	Mean Correlation	<i>t</i>	<i>p</i> Value	Adjusted <i>p</i> Value
28	R,L Occipital	17,18	0.11	14.42	4.5*10 ⁻²⁶	1.6*10 ⁻²⁴
13	R,L Sup Parietal, Lateral Occipital	7,40	0.10	15.55	2.5*10 ⁻²⁸	8.9*10 ⁻²⁷
24	R,L Occipital, Lingual G	18,19	0.08	12.22	1.7*10 ⁻²¹	6.1*10 ⁻²⁰
18	R,L Superior Occipital	7,19	0.06	10.28	2.7*10 ⁻¹⁷	9.6*10 ⁻¹⁶
34†	R Middle FG, Angular/Posterior Supramarginal G	6,8,40	0.06	7.94	3.2*10 ⁻¹²	1.1*10 ⁻¹⁰
51	L Inferior/Middle Temporal Lobe	20,21,37	0.04	6.47	3.8*10 ⁻⁰⁹	1.3*10 ⁻⁰⁷
3	L Postcentral G	2,3,40	0.04	4.97	2.9*10 ⁻⁰⁶	1.0*10 ⁻⁰⁴
56	R Inferior/Middle Temporal Lobe	20,21,37	0.04	6.13	1.8*10 ⁻⁰⁸	6.4*10 ⁻⁰⁷
9†	L Middle FG, Angular/Posterior Supramarginal G	6,8,40	0.03	4.85	4.7*10 ⁻⁰⁶	1.6*10 ⁻⁰⁴
43	R,L Frontal Pole	10	0.02	4.80	5.6*10 ⁻⁰⁶	2.0*10 ⁻⁰⁴
5	R,L SFG	6	0.02	3.58	5.4*10 ⁻⁰⁴	1.9*10 ⁻⁰²
22	R,L Middle Temporal G	21,22,39	-0.05	-8.18	9.7*10 ⁻¹³	3.4*10 ⁻¹¹
17	Anterior Cingulate Cortex, Paracingulate	9,10	-0.04	-8.90	2.7*10 ⁻¹⁴	9.5*10 ⁻¹³
44	R,L Temporal Pole	21,38,47	-0.03	-6.33	7.1*10 ⁻⁰⁹	2.5*10 ⁻⁰⁷
52	R,L Frontal Pole/Sup Frontal	8	-0.03	-5.75	1.0*10 ⁻⁰⁷	3.5*10 ⁻⁰⁶
25	Anterior Cingulate Cortex, Paracingulate	10,24,32	-0.03	-5.44	3.9*10 ⁻⁰⁷	1.4*10 ⁻⁰⁵
7	R,L Precentral/Postcentral G	3,5,6	-0.02	-3.72	3.3*10 ⁻⁰⁴	1.1*10 ⁻⁰²
27	Posterior Cingulate Cortex	23,24,31	-0.02	-3.50	6.9*10 ⁻⁰⁴	2.4*10 ⁻⁰²
50	Frontal Medial Cortex	10,32	-0.02	-3.75	3.0*10 ⁻⁰⁴	1.0*10 ⁻⁰²

Note: DPX, dot pattern expectancy task; R, right; L, left; S, superior; F, frontal; G gyrus; Comp, component. *P* values correspond with one-sample *t* tests suggesting these correlations are significantly different from zero. Adjusted *p* values have undergone Bonferroni correction for multiple comparisons. Components in the top half of the table are positively correlated with the DPX task; components in the bottom half are negatively task-related.

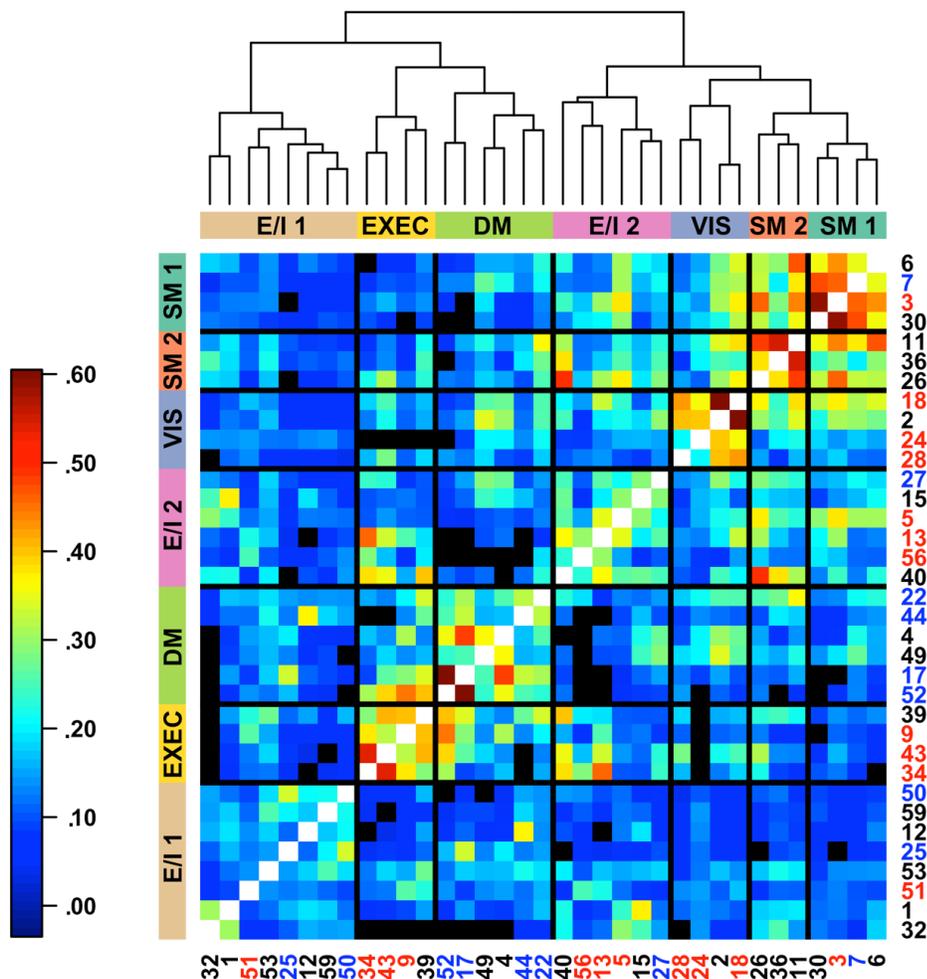
†, Hypothesized to be task-related *a priori*

Table 3.2. Component Connections with Groups Differences in Tonic Functional Network Connectivity

Comp.	Region	BA	tFNC Value		<i>t</i>	<i>p</i>	Adj. <i>p</i>
			HC	SZ			
1			0.40 (0.17)	0.25 (0.13)	4.82	5.2*10 ⁻⁶	0.0031
2	Intracalcarine C, Lingual G	18					
3	Postcentral G	2,3,40					
2			0.46 (0.17)	0.33 (0.15)	4.11	8.2*10 ⁻⁵	0.049
3	Postcentral G	2,3,40					
18	Occipital	18,19					
3			0.44 (0.18)	0.29 (0.15)	4.41	2.7*10 ⁻⁵	0.016
6	Precentral G	3,4,6					
18	Occipital	18,19					
4			0.37 (0.15)	0.19 (0.17)	5.74	1.1*10 ⁻⁷	6.4*10 ⁻⁵
22	Superior Temporal G	21,22,39					
36	Insula	13					
5			0.47 (0.16)	0.34 (0.14)	4.16	6.9*10 ⁻⁵	0.041
34	Middle Frontal G, Angular/Posterior Supramarginal G	6,8,40					
40	Insula, Frontal Operculum, Superior Frontal G	13,8					

Note: C, cortex; G, gyrus; BA, Brodmann area; HC, healthy control; SZ schizophrenia patient; tFNC, tonic functional network connectivity; Adj. *p*, *p* values adjusted for multiple comparisons; Comp., component.

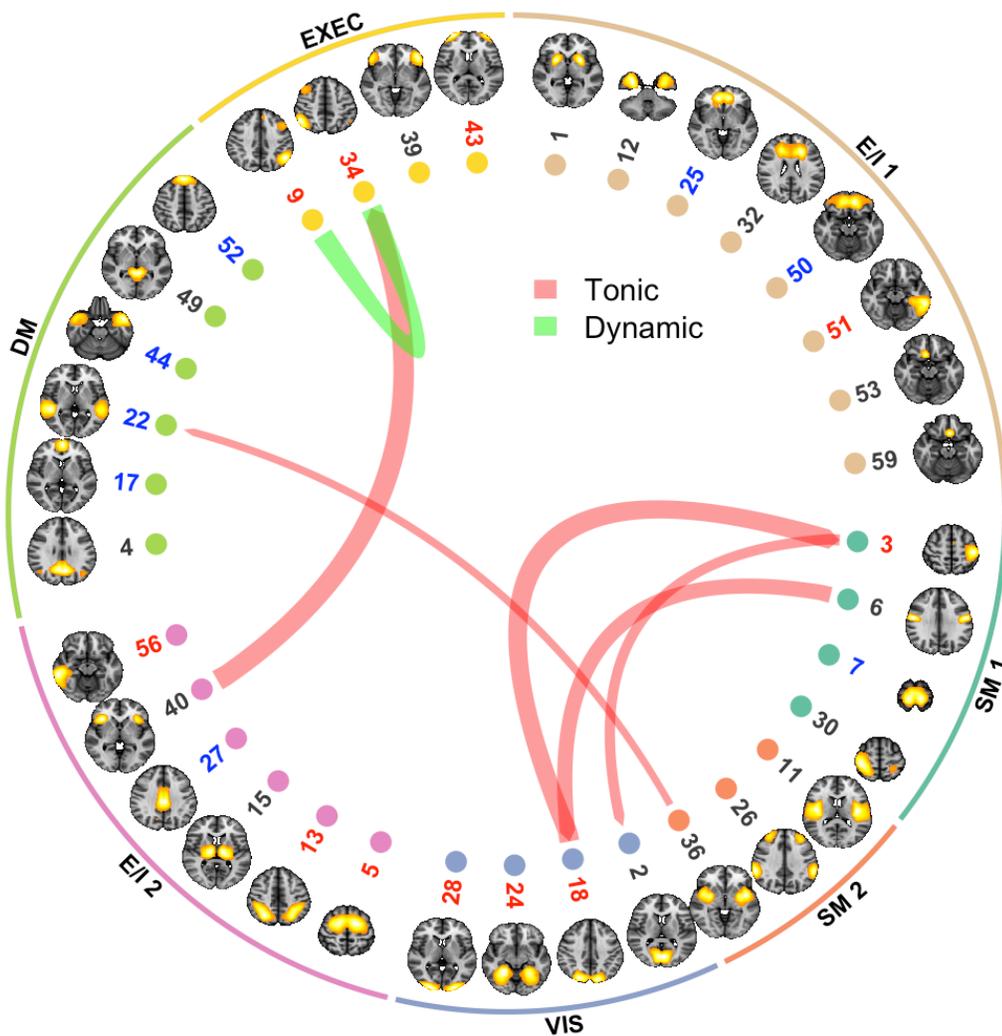
Figure 3.1. Functional network connectivity during the DPX task in SZ and HC.



The heatmap represents the pairwise strength of the connection between component time courses, with warmer colors representing stronger positive connections. Black cells represent those connections whose strength was not significantly different from zero. The dendrogram above the plot was produced based on the tFNC between components with height of the dendrogram representing dissimilarity, and the colored bar beneath it represents groupings produced from that dendrogram. The network groupings, which roughly correspond with those of Ray and colleagues (Ray et al., 2013), are abbreviated

as follows: E/I 1, Emotion/Interoception 1; E/I 2, Emotion/Interoception 2; SM 1, Primary Sensorimotor; SM 2, Secondary Sensorimotor; VIS, Visual; EXEC, Cognitive/Executive; DM, Default Mode. The colors of the component numbers indicate whether each component's time course was positively (red), negatively (blue), or uncorrelated with the DPX task time course. SZ and HC were combined in this analysis because its intention was to identify which connections would be examined for group differences.

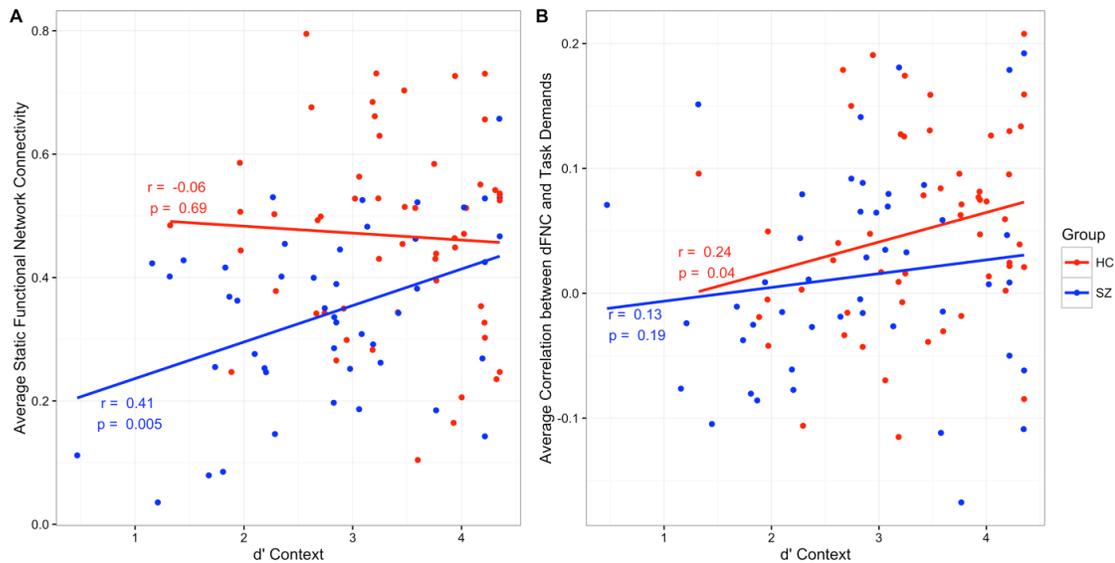
Figure 3.2. Group differences in tonic and dynamic functional network connectivity.



Red arrows represent pairs of components between which tFNC differed significantly between groups. The arrows flow from the preceding component to the following component. Width of the red arrows represents the strength of the correlation, and all correlations are positive in magnitude. The green connection represents the pair of components whose dFNC significantly correlated with the DPX task time course for B Cues between groups. Groups also differed significantly in the strength of this

correlation, with healthy controls showing stronger positive correlation. The colors of the component numbers indicate whether each component's time course was positively (red), negatively (blue), or uncorrelated with the DPX task time course. The colored sections containing groups of components refer to the groupings produced in figure 3.1. Maps are presented in radiological view, so images are reversed left to right.

Figure 3.3. Functional Network Connectivity Predicts Task Performance



A. tFNC between right ECN and SN predicts task performance in SZ. The degree to which the right central executive network correlated with the salience network predicted DPX task performance in schizophrenia patients as measured by d' Context. **B.** Task modulation of left and right ECN dynamic FNC predicts task performance in HC. The extent that HC were able to coordinate the left and right ECN with the time course of B cues on the DPX task itself predicted task performance in HC but not SZ.

Chapter 4: Test-Retest Reliability of GLM, ICA, tFNC, and dFNC in Schizophrenia

Patients and Healthy Controls: a Multi-site Imaging Study

Foreword: This chapter was written in collaboration with Angus W. MacDonald, who provided edits and revisions to earlier drafts. The data collection methods in this chapter are largely redundant with those outlined in Chapter 2 as well as in appendices 1 and 2. Methods specific to this chapter can be found within this chapter's method section.

Introduction

There has been an increase in the last decade of the number of fMRI studies employing ICA to examine group differences between schizophrenia patients (SZ) and healthy controls (HC). There is also a growing trend of utilizing multisite data collection methodology to increase sample sizes in those studies. An important consideration when designing such a study is the reliability of the measure used to compare subject groups. It was the goal of the present study to assess and compare the reliability of four fMRI data analysis methods and to ascertain how susceptible each method is to data collection site effects and subject group effects.

The general linear model (GLM) is a common method of ascertaining which clusters of voxels in the brain show activity that fits a hypothesized pattern, usually the time course of a task performed in the MR scanner. In a review of GLM reliability studies, Bennett and Miller (2010) found that the average reliability was in the "fair" range (about .50; Cicchetti & Sparrow, 1981; Fleiss, 1986), but there was a significant range of reliability estimates in the studies they reviewed. Some possible factors

influencing the reliability of GLM analyses in their review were the complexity of the task performed by subjects (i.e., finger tapping versus a working memory task), the sample size, whether the design was block or event related, the retest interval, and details of the analytic method such as which statistic from the GLM is chosen. What remains to be seen is how the reliability of GLM compares with that of other analysis techniques in the same sample and how those reliability estimates are affected by multiple data collection.

One such alternative data analysis technique is ICA, which finds patterns of voxels that share a similar time course of activity. There have been comparatively fewer studies of the reliability of ICA, and there have been especially few studies of task-based ICA. The studies that have used the intraclass correlation (ICC) to assess reliability of ICA have generally found fair reliability across components whether at rest (Blautzik et al., 2012; Guo et al., 2012; Wisner et al., 2013; X.-N. N. Zuo et al., 2010) or during a task (Poppe et al., 2013). However, no study has examined how the reliability of ICA changes across data collection sites or in patients with schizophrenia.

Two additional analysis techniques for which no reliability studies exist to our knowledge are tonic (tFNC; Jafri et al., 2008) and dynamic (dFNC; Sakoğlu et al., 2010) functional network connectivity. These techniques examine the extent to which the time courses of components resulting from ICA, or any other clustering technique, are related. tFNC examines this relationship between pairs of components across the entire scan and results in a single value for each component pair representing the strength of their connection. dFNC calculates how the temporal relationship between component pairs changes over the course of a scan by calculating correlation coefficients within a sliding

window. This results in a series of estimates of connection strength between component pairs along the course of an fMRI scan.

One complication of performing fMRI analyses with the intent to compare SZ with HC is that if the reliability of the measure is different in the two groups, it could lead to difficulty discerning true group effects. This results because reliability sets an upper bound on the observed relationship between variables. There is some reason to believe that schizophrenia patients may produce less reliable data than healthy controls. One piece of evidence is the finding of Manoach et al. (2001) that schizophrenia patients had less reliable activation data than healthy controls in “higher order cognitive” regions but equally reliable data in primary visual and other sensory regions. Maiza et al. (2011) found no differences between schizophrenia patients and healthy controls in regions underlying early language processing, potentially supporting the results of Manoach et al. It is possible that patient groups in general are less reliable, based on the findings of reduced test-retest reliability in an activation study comparing stroke patients and healthy controls (Chen & Small, 2007) as well as a resting state functional connectivity study of children with and without attention deficit hyperactivity disorder (Somandepalli et al., 2015).

In addition to multiple subject groups, analyzing data from multiple sites with different scanners may affect the reliability of a measurement. Multi-site studies are an effective way of increasing power in neuroimaging studies; however, differences in scanner technology and other methodological details have the potential to affect group comparisons in these studies. These site-specific factors add noise to the analysis model that can reduce the ability to discern real effects of interest. For instance, a large multi-

site study examining 10 sites found low between-site reliability before taking steps in the analysis to control these effects (Friedman et al., 2008). Likewise, Wurnig et al. (2013) also found important site differences in a large multi-site study. However, other studies (Gountouna et al., 2010; Yendiki et al., 2010) failed to find significant site differences in reliability. It is unknown how the interaction of site effects and multiple subject groups affect the reliability of fMRI data analysis techniques.

In addition to site and subject group effects, there exists the potential for a measurement confound if certain brain regions are measured more reliably than others. Akin to the psychometric confound (Chapman & Chapman, 1973), this confound would make brain regions that are measured more reliably more likely than other regions to show group differences given no true difference between those brain regions. This circumstance could result in apparently specific deficits in the activity or functional connectivity of a certain region, but it might actually only reflect the ability to measure that region better than others. Some evidence for differences in the reliability of some brain regions over others in HC comes from a study that created reliability maps based on voxel time courses during a change detection task (Simak, Liou, Zhigalov, Liou, & Cheng, 2012). They found that group results showed greatest reliability in primary visual, dlPFC, posterior parietal lobe, and areas of the DMN. These regions are commonly implicated in activation and functional connectivity studies comparing SZ with HC. If these regions are indeed measured more reliably than others, it raises the question of whether these apparent specific effects are actually indicative of a global difference between SZ and HC.

The first goal of the present study was to estimate the reliability of GLM, ICA, tFNC, and dFNC in the same multisite dataset including SZ and HC while those subjects performed a cognitive task. The second goal was to examine the variability of those reliability estimates across data collection sites to measure how sensitive each metric is to site effects. The third goal was to make a direct comparison of the four analysis methods on both the overall magnitude of their reliabilities and their variability across data collection sites and subject groups. The final goal was to examine whether the reliability of GLM and ICA differed across different brain regions.

Method

Subjects. Subjects were recruited as part of the Cognitive Neuroscience Test Reliability and Clinical applications for Schizophrenia consortium (CNTRACS), which is made up of five sites. Data were collected at two time points, approximately three weeks apart. Data collection and subject recruitment methodology for this sample and the CNTRACS project has been previously published (Henderson et al., 2012; Poppe et al., 2016; Ragland et al., 2015; Silverstein et al., 2015). In addition to exclusion criteria described in appendix 1, subjects without two sets of complete data (one from each time point) were excluded from further analysis.

The final sample consisted of 86 subjects (45 HC, 41 SZ) with no significant differences between included and excluded subjects on demographic, behavioral, or symptom indices (p 's > .06). The final groups did not differ on any measured demographic variable except education (see Table 4.1). After removing subjects with

excessive movement (see appendix 1 for details), groups did not differ on average absolute or relative head movement (both p 's > .06).

Dot pattern expectancy task. The Dot Pattern Expectancy task (DPX) is a test of the goal maintenance aspect of executive control and has been described (Henderson et al., 2012; Jones et al., 2010). Four DPX blocks were performed by each subject at each time point, with each trial consisting of a cue dot pattern followed by a probe dot pattern. One dot pattern was identified as a valid cue ('A' cue), and another as a valid probe ('X' probe), with all other cues and probes being invalid ('B' cues and 'Y' probes, respectively). This creates four trial types based on different combinations of valid and invalid cues and probes ('AX', 'AY', 'BX', and 'BY'). Each block of the DPX task consisted of 40 trials: twenty-four AX (60%), six AY (15%), six BX trials (15%), and four BY (10%).

fMRI data acquisition and preprocessing. Three CNTRaCS sites used Siemens Trio 3 Tesla scanners (Minnesota, Washington University, UC Davis), one site employed a Phillips 3 Tesla scanner system (MPRC), and the fifth site used a Siemens Allegra 3 Tesla system (Rutgers). The procedure of each scanning session was as follows: a localizer scan, a 3D T1-weighted anatomical scan (1 mm³ isotropic voxels), a 2D T2-weighted scan, a field map, and the functional EPI sequences. The functional scans were T2*-weighted gradient echo EPI sequences, with TR = 2000 ms, TE = 30 ms, flip angle = 77 degrees, 32 contiguous AC-PC aligned axial slices, voxel size = 3.5 x 3.5 x 4.0 mm. Each scan session included the collection of four, 180-volume scans during four blocks of

the DPX task. Preprocessing used FMRIB Software Library (FSL v. 4.1.8) and included motion correction, brain extraction, prewhitening, high-pass temporal filtering with sigma of 100 s; B_0 field unwarping, spatial smoothing with a 5 mm FWHM Gaussian kernel, and spatial normalization and linear registration to the MNI 152 standard brain. Subjects with excessive movement, scanner artifacts, or otherwise poor data quality were removed from the analysis (see appendix 1 for details).

General linear model. The GLM analysis was performed identically as described in a previous report of the current time 1 data (Poppe et al., 2016) and was performed separately at each time point. All GLM analyses were conducted using fMRI Expert Analysis Tool (FEAT) using fMRIB's linear optimized basis functions (Woolrich et al., 2004). Whole-brain analyses were performed at the group level, and the contrasts of interest were a comparison of SZ with HC on the lower level contrast of B Cue activation greater than A Cue activation and also on the lower level contrast of B Cue activation itself.

Independent component analysis.

Meta-ICA procedure. The ICA consisted of a "meta" procedure in which several lower-level group ICAs were performed using randomized subject orders and a final ICA is then conducted using the results of the lower levels (Wisner et al., 2013). The input data for these lower level ICAs were the four temporally concatenated scans of each subject. Two meta-ICA procedures were completed: one for each time point. The number of lower level ICAs (25) as well as the number of components to specify in each group

ICA (60) was suggested by previous research (Poppe et al., 2013). For display purposes and for some subsequent analyses, spatial maps resulting from the meta-ICA were thresholded using a normalized threshold of 0.4 (Poppe et al., 2013).

Dual regression. Following the meta-ICA, the 60 unthresholded spatial maps were used in dual regression to produce subject-specific spatial maps and time courses. For the purposes of reliability estimation, only time 1 maps were used in this dual regression for both time 1 and time 2 data. This was to allow a direct comparison of the same components when calculating reliability.

Task relationship. Following dual regression, the group maps were visually examined to exclude artifactual components. The subject-specific time-courses produced by dual regression were next used to calculate Pearson correlations between component time courses and the time course of B cues of the DPX task after convolving that time course with a double gamma hemodynamic response function. These correlation coefficients were then Fisher z-transformed and t tests were conducted across groups at each time point to determine which components' time courses were significantly related to the DPX task time course.

Tonic functional network connectivity. Tonic Functional Network Connectivity (tFNC) was performed in the same manner as in chapter 3. Briefly, the following steps were taken: 1) subject-specific time courses resulting from dual regression were interpolated four times, 2) time courses were allowed to slide in time five seconds in each

direction relative to one another, 3) a Pearson's product moment correlation coefficient was calculated at each amount of lag between components, 4) the absolute maximum correlation was chosen to represent the tFNC between each component pair. This resulted in a tensor of tFNC values of size $N \text{ Components} \times N \text{ Components} \times N \text{ Subjects}$.

Dynamic functional network connectivity. Dynamic Functional Network Connectivity (dFNC) was calculated as in chapter 3. First, the lag values calculated in tFNC were used to "pre-lag" component time courses relative to one another. Next, a sliding window was used to calculate a series of correlations between each component pair. The width of the window was 15 seconds, and it advanced 4 seconds with each step. To measure the task-modulation of the relationship between any two components, a "task demands" variable was created by calculating a moving average of the time course of B Cues of the DPX task after that time course had been convolved with a double gamma hemodynamic response function. The dFNC values for a component pair were then correlated with the task demands variable. This process resulted in a tensor of task modulation correlation coefficients of size $N \text{ Components} \times N \text{ Components} \times N \text{ Subjects}$.

Reliability analyses.

GLM reliability. To assess the reliability of GLM results, two procedures were employed. The first was to assess spatial overlap of results between time points by calculating Dice coefficients (Dice, 1945) between group contrasts. The second procedure involved calculating intraclass correlation coefficients (ICC) for each voxel within the mask for time 1 B Cue > 0 group contrast map and taking the average of those

values. This analysis was also performed using the map of voxels significant in the group B Cue > A Cue group contrast. The specific form of ICC used in this study is as follows:

$$\frac{BMS - EMS}{BMS + (k - 1)EMS + \frac{k(JMS - EMS)}{n}}$$

where *BMS* is the between-subjects mean square, *JMS* is the between-judges mean square, *EMS* is the residual mean square, *k* is the number of ratings, and *n* is the number of subjects (Shrout & Fleiss, 1979). This formula corresponds with the consistency ICC(2,1) formula in Shrout and Fleiss (1979), and to ICC2 in the R package *psych*'s ICC function (Revelle, 2010). Subject-level contrast maps were used to calculate these GLM ICC values, which were calculated separately for each subject group (HC and SZ) as well as for each CNTRACS site. Data from the Rutgers site was excluded from all reliability analyses involving site membership because only eight subjects came from that site (4 HC, 4 SZ).

ICA reliability. To assess spatial overlap of group components derived from each time point, a matrix of Dice coefficients was created representing the spatial overlap of components between time points. Voxelwise ICC calculations were also performed using subject-specific maps resulting from dual regression in voxels within time 1 thresholded group spatial maps. These ICCs were calculated separately for subject group and for each CNTRACS site. Only non-artifact components whose time courses were significantly correlated with the task time course of B Cues were included in reliability analyses.

An additional analysis was performed to assess the extent to which components associated with different functions differed in reliability. First, components were sorted

into four groups: executive, default mode, visual, and other. Next, an ANOVA was performed on the average of voxel ICC values with subject group and component group as predictor variables.

tFNC reliability. ICCs were calculated for tFNC values of each connection between non-artifact components, and were calculated separately for each subject group and each CNTRACS group. These ICCs were then averaged across component connections. Additionally, ICCs were averaged across components with the highest tFNC values in an attempt to determine what the reliability of an analysis might be that defined a graph based in highest tFNC values. Likewise, correlation coefficients were calculated between tFNC values and tFNC reliability separately for each subject group to determine if there is a relationship between tFNC strength and reliability.

dFNC reliability. To assess the reliability of the correlations between dFNC and the DPX task, ICCs were computed using these Fisher's *z*-transformed (Fisher, 1915) correlations for each connection between non-artifact components (as in the tFNC reliability computations) and then averaged. dFNC reliability was also estimated using 30-second and 60-second sliding windows to determine how this parameter affects reliability.

Comparison of reliability amongst analysis methods. To compare the magnitudes of reliabilities of the four analysis methods, an analysis of variance (ANOVA) was performed with ICC as the dependent variable and analysis method as the

independent variable. Two similar models were fit to determine if subject groups and CNTRACS sites, respectively, differed in average reliability. Tukey's range test (Tukey, 1949) was used to assess which variables were significantly different in the event of a significant omnibus test.

To assess if analysis methods differed with regard to the variability of reliability across sites, a Levene's test of variance homogeneity (Levene, 1960) was performed on ICC calculations of each analysis method. Given a significant omnibus test, further pairwise tests were performed between analysis methods to determine which methods were significantly more variable than the others. Additional Levene's tests were computed employing subject group and CNTRACS site as grouping variables in order to test whether observed differences in variance between analysis methods were driven by those variables.

Results

GLM results and reliability. The Time 1 GLM group contrast of HC > SZ and B Cues > A Cues resulted in significant activation in bilateral frontal pole as shown in figure 4.1. The corresponding analysis at Time 2 did not result in significant activation, which produced a Dice coefficient of zero for that comparison. In the B Cues > 0 average activation contrast (across subject groups), activation was found in bilateral occipital lobes, bilateral superior parietal lobes, and left postcentral gyrus at both time points. The dice coefficient for this contrast was 0.77. With regard to reliability, the average ICC of voxels within the HC > SZ mask of B Cues > A Cues was .15 for HC and .09 for SZ, corresponding to "poor" reliability (Cicchetti & Sparrow, 1981; Fleiss, 1986). These

results are presented in figure 4.2. However, there was variability across sites and across subject groups, with ICCs ranging from -0.30 to 0.41. Full results are presented in table 4.2 and displayed in figure 4.3. For the B Cues > 0 average activation map, the ICC for HC was .19 and for SZ was .33.

ICA and task relationship. The group ICA at time 1 resulted in 37 non-artifact components, and the same analysis at time 2 resulted in 36 non-artifact components. The Dice comparison of time 1 maps with time 2 maps produced a matrix of coefficients with a strong diagonal, indicating a nearly one-to-one correspondence of the components produced by each analysis, suggesting good reproducibility. The average of the maximum Dice value in each column of this matrix was .57. These results are presented in figure 4.4.

The analysis of task-relatedness of the components, taken as a whole, at time 1 resulted in 11 components positively correlated with the B Cues time course in HC and 5 negatively correlated, as is presented in figure 4.5. When the same analysis was completed with time 1 components dual regressed in time 2 functional data, the same 11 components were found to be positively correlated with the DPX task in HC, and 4 components were negatively correlated. For SZ, at time 1 there were 7 positively correlated components and 6 negatively task-related components. At time 2, there were 5 positive and 4 negative. The average ICC of task-relatedness across non-artifact components was 0.224 for HC and 0.250 for SZ, indicating poor reliability. However, when considering the dichotomous choice of whether a component was task-related or

not, the kappa statistic for that decision was 0.86 for HC and 0.42 for SZ (strong and weak, respectively; McHugh, 2012). These results are displayed in figure 4.6.

As presented in table 4.2, the average voxelwise reliability of the task-related group ICA spatial maps across sites was 0.686 for HC and 0.667 for SZ, corresponding to “fair” reliability (Cicchetti & Sparrow, 1981; Fleiss, 1986). When broken down by site, these values remained fairly consistent (ranging from .52-.70). These results are displayed in figure 4.3.

To assess the variability of reliability across components, we performed an ANOVA of average ICC predicted by subject group and by component group (i.e., executive, default mode, visual, or other). This ANOVA resulted in a significant main effect of component group, $F(3,66) = 5.82, p = .001$, whereas neither the main effect of subject group nor the interaction of component group with subject group were significant, both p 's $> .05$. Tukey's HSD post hoc tests revealed that executive components ($M = .74, SD = .01$) were significantly more reliable than the group of components that were not executive, default mode, or visual ($M = .55, SD = .12, p = .01$). Executive components did not differ significantly from either default mode or visual components, both p 's $> .05$. These results are presented in figure 4.7.

Tonic FNC reliability. To assess reliability of tFNC connection strength, ICCs were calculated for each connection between non-artifact components. As presented in table 4.2, the average of these ICCs across sites for HC was 0.560 and for SZ was 0.503, corresponding to “fair” reliability. Across sites, there was little variability (see table 4.2 for ICCs for individual sites). These values are displayed in figure 4.3.

To determine the relationship between tFNC magnitude and reliability, correlation coefficients were computed between these variables and average ICCs were computed for connections above various thresholds of tFNC magnitude. There were significant positive relationships between tFNC magnitude and ICC in both HC ($r = .48, p < .001$) and SZ ($r = .33, p < .001$). Table 4.3 and figure 4.8 present these results.

Dynamic FNC reliability. To measure the reliability of the dynamic FNC analysis, ICCs were calculated based on the strength of the correlation between dFNC value and the task demands variable for each connection between non-artifact components. The average ICC across connections for HC was .014 and for SZ was .003, corresponding to “poor” reliability. Table 4.2 and figure 4.3 present these results as well as ICCs broken down by CNTRACS site.

Because of the relatively narrow window size employed in the current dFNC analysis, it was repeated with a window of 30 seconds and 60 seconds. Reliability estimates failed to improve using this wider window of 30 seconds ($M_{HC} = .00, SD_{HC} = .15; M_{SZ} = .00, SD_{SZ} = .16$) or 60 seconds ($M_{HC} = .01, SD_{HC} = .15; M_{SZ} = .01, SD_{SZ} = .16$), p of both main effects and interaction $> .05$.

Comparison of reliability across methods. The average reliability was significantly different between methods and across subject groups and CNTRACS sites, $F(3,28) = 108.8, p < .001$. Tukey’s range test revealed that GLM and dFNC were each significantly less reliable than both ICA and tFNC (p ’s $< .001$) and that dFNC was less reliable than GLM ($p < .001$). ICA and tFNC did not differ in reliability ($p = .99$). These

results are presented in figure 4.2. The ANOVA models with subject group and CNTRACS were not significant ($F_s < 0.02$, $p_s > .98$).

The Levene's test of homogeneity of variance showed that there were significant differences between analysis methods in the amount of variance in their ICCs, $F(3,28) = 7.10$, $p < .005$. Post-hoc pairwise comparisons between methods showed that GLM had significantly higher variance in reliability across sites and subject groups compared with ICA and dFNC (p 's $< .02$) but did not differ from tFNC ($p = .07$). tFNC had more variability than dFNC ($p < .01$), but ICA did not differ from tFNC or dFNC (p 's $> .21$).

Discussion

In order to assess and compare the reliability of GLM, ICA, tFNC, and dFNC, we calculated ICCs of those techniques using a large, multisite fMRI dataset including HC and SZ performing the DPX task. This represents the first study to directly compare the reliability of these metrics and the first study to assess the reliability of either FNC method. We found that GLM and dFNC had generally poor reliability whereas ICA and tFNC had fair, and in some regions good, reliability. Additionally, GLM showed higher variability across data collection site and subject groups compared with ICA and dFNC, suggesting these other analysis methods may be less sensitive to site effects when analyzing data from multisite studies. Finally, we found that frontoparietal "executive" components had higher reliability than other components in the ICA analysis, which may introduce a measurement confound for fMRI analyses.

Reliability estimation.

GLM reliability. We found that GLM analysis results were in the “poor” range according to calculated voxelwise ICCs (Cicchetti & Sparrow, 1981; Fleiss, 1986). Although the obtained mean ICCs for HC and SZ were lower than the mean ICC reported in a recent review of activation fMRI studies (Bennett & Miller, 2010), they were within the range of obtained reliability estimates of studies included in that review. Some reasons for lower reliability in the present study compared with the average in that review derive from the task and study design in the current study. The task in the present study involved executive functioning and a fast event-related design, whereas several of the tasks reviewed previously were simple (e.g., finger tapping). Of note, in the current study higher reliability estimates were obtained in the simpler contrast of B Cues > 0 as opposed to the more specific (but more interpretable) contrast of B Cues > A Cues. Additionally, nearly all studies in that review used a block design as opposed to an event-related design when creating contrast images. The use of a block design is more likely to produce more robust effects in fMRI analyses when compared with fast event related designs. Last, all studies included in that review were single-site studies. As the results of the current study show, GLM reliability varied significantly across study sites. Overall, these results suggest that for complicated cognitive tasks, beta-maps from GLM analyses may not be the best choice for an imaging biomarker. This is in part because the maximum correlation between a measure and another variable is determined by the reliability of that measure. It is comparatively more difficult to observe an association between an imaging biomarker and another variable such as psychotic symptoms or diagnosis given a measure with poor reliability than one with high reliability.

The spatial overlap of GLM results assessed by the Dice statistic showed a high overlap for activation associated with B Cues. This overlap was greater than the average overlap reported by Bennett and Miller (2010). However, in the more specific contrast of B Cues minus A Cues, we failed to observe significant activation at the second time point, which resulted in a Dice coefficient of zero. Again, the effect of higher reproducibility for the more general activation contrast is in line with the idea that more specific, complex contrasts are likely to have a smaller signal to noise ratio than more general analyses. The failure to observe significant activation in the B Cues > A Cues contrast at the second time point reflects the relatively poor reliability of this specific GLM contrast.

ICA reliability. With regard to ICA reliability, we found that the average reliability of task-related components was in the “fair” range (Cicchetti & Sparrow, 1981). These results suggest that ICA coherence values are sufficiently reliable to use for examining group differences and individual differences between HC and SZ in multisite task-based studies. The ICC values obtained in this study compare favorably to those reported in previous test-retest studies of ICA (Blautzik et al., 2012; Guo et al., 2012; Jeong, Choi, & Kim, 2012; Wisner et al., 2013). However, all these studies utilized resting state data. Given the limitations of the current study listed in the GLM reliability section, the relatively high reliability obtained from ICA in both HC and SZ is encouraging for task-based ICA studies of group differences and for the use of this analytic method for individual differences research, such as identification of a biomarker. These results complement the findings of Zuo and Xing (2014) who found that ICA was

one of the most reliable methods in their analyses (although they did not assess GLM, tFNC, or dFNC).

We also found the task-relatedness of ICA components to be highly reliable. HCs had higher reliability than SZs with regard to the identification of task related components. This may indicate that SZs are inconsistent in how they approach or react to the DPX task from one time point to another, perhaps due to underlying cellular processes or cytoarchitecture (Harrison, 2000). These results are promising for the utility of task-based ICA in identifying biomarkers for schizophrenia. One criticism of resting state functional connectivity studies is that subject behavior is unconstrained, which limits the behavioral interpretations that can be made based on the results of such studies compared with GLM activation studies. These results show that largely the same set of networks was found to be task-related at both time points.

Finally, we found that frontoparietal “executive control” components had significantly higher reliability compared with other components that were not visual or default mode components. Executive components did not differ from visual or default mode components, and this effect did not depend on subject group. These results are in line with those of Zuo and Xing (2014), who found that executive, default mode, and visual components were the most reliably measured in a resting state functional connectivity study. It is not clear why these particular regions are measured more reliably than other regions. It is possible that these regions are more stably connected functionally, such that the time courses of voxels within these regions are more temporally correlated. Another explanation for the results of the current study is that these particular regions are associated with performing the DPX task, which results in

more robust measurements than more transiently active networks. It is true that executive and visual components were found to be positively correlated with the DPX task in this study and that default mode components were found to be negatively correlated. It is common in the fMRI literature to find positive correlations with task in executive and visual components/regions while finding negative task correlations in default mode regions. However, that does not explain why Zuo and Xing also found these networks to be the most reliable in their resting state ICA.

An alternative explanation is that the relatively higher reliability of the measurements in these regions is why they are found to be task-related and found to differ between groups more often than other regions are. This measurement confound would make it more likely to observe group differences in these regions compared with other regions given the same underlying true group difference amongst regions. Regions involved with the frontoparietal executive network and the default mode network consistently emerge in group contrasts between schizophrenia patients and healthy controls in both task activation studies (Minzenberg et al., 2009) and resting state functional connectivity studies (Sheffield & Barch, 2016). Additionally, they are involved in some of the leading putative functional connectivity intermediate phenotypes of schizophrenia (Cao, Dixon, Meyer-Lindenberg, & Tost, 2016). A less obvious but similar effect can be seen in the GLM reliability results shown in figure 4.5. It will be important to more thoroughly examine the nature of this effect in order to mitigate its effects on the interpretation of clinical fMRI studies.

tFNC reliability. The present study represents the first reliability study of both tFNC and dFNC. The average reliability of the magnitude of tFNC connections was found to be fair across all non-artifact components. These results suggest that tFNC is sufficiently reliable to use for examining group differences and individual differences between HC and SZ in multisite task-based studies. We also found a linear relationship between the strength of a dFNC connection and its reliability, meaning that weak connections were less reliably measured than strong connections. This effect may have to do with the ICC calculation, in that high ICCs require high interindividual variance and low intraindividual variance. If there is not much variance between individuals, the reliability will be low even if there is consistency within individuals. Figure 4.8C shows the relationship between the variance and the mean of tFNC magnitude, which suggests the low reliability of weak tFNC connections is likely due to the restricted variance in their measurement.

dFNC reliability. The reliability of the degree to which dFNC values were task-modulated by the DPX task was found to be poor. The poor reliability observed with dFNC indicates caution should be employed when utilizing this method. There was little variability across sites and across subject groups when using this method, so more research is necessary to determine the source of the poor reliability.

To determine the effect of window size on dFNC reliability, we increased the window to 30 seconds and 60 seconds and repeated the dFNC analyses and reliability computation. We did not observe a change in reliability estimate when the window size was increased. More research is needed to determine what factors contribute to the low

reliability of task modulation of dFNC results, and whether changes to the analytic methods or research design will be needed to improve reliability. This work is important because dFNC may reveal an important aspect of connectivity dynamics in the brain that provide a window into psychopathology (Damaraju et al., 2014).

Reliability comparison between methods. We found that ICA had the highest ICCs across sites and subject groups, and dFNC had the lowest. tFNC had higher reliability than GLM and dFNC, and GLM had higher reliability than dFNC. These results suggest that ICA and tFNC provide the most power to detect group differences between HC and SZ, and they are the best candidates for biomarkers of schizophrenia. This is because they allow better measurement of individual differences amongst subjects. One explanation for the lower reliability of GLM compared with ICA and tFNC is that physiological noise may contribute to measures of GLM more than in ICA unless special noise extraction is done. This is because an ICA is capable of isolating artifactual BOLD signals such as those associated with respiration or movement and separating those signals from those of non-artifact components. With a naïve GLM analysis, those artifacts remain in the data and contribute to voxels' timecourses, adding noise to the data. More research is needed to determine how effective noise-removal strategies are to increasing test-retest reliability.

In addition to absolute magnitude of reliability, we also examined the variability of reliability across data collection sites and subject groups. We found that GLM had the highest variability in reliability and tFNC had the second highest. ICA and dFNC were the most stable in their reliability estimates across sites and subject groups. These results

suggest that GLM and tFNC may be more susceptible and ICA less susceptible to site and group effects, which has implications for multisite imaging studies or that include clinical groups. Specifically, for multisite fMRI studies, especially those including clinical samples in addition to healthy controls, it may be more effective to use an ICA based analysis method as opposed to GLM.

Limitations and future directions. The current study included four analysis types in its comparison, however many more approaches have been used . It would be informative to re-analyze these data using other common analysis methods (e.g., seed-based correlation, graph theoretical methods, dynamic causal modeling) and compare test-retest reliability among them. The sample in this study included HC and SZ performing a complex cognitive task. Future research might explore the reliability of these methods in particular in the resting state. Such research may also choose to include other clinical samples to determine if the general trend of reduced reliability in SZ is present to the same degree in other populations.

Table 4.1. Subject Demographics

	Group		Test
	Patients	Controls	
N	46	54	
Mean Age (yrs)	36.4 (12.1)	35.7 (12.1)	$t(84) = 0.01$
% Male	70.7	77.8	$\chi(1) = 0.25$
% Caucasian	53.7	62.2	$\chi(1) = 0.34$
% Right-Handed	85.4	84.4	$\chi(1) = 0.00$
Mean Education (yrs)	13.9 (2.0)	15.5 (2.6)	$t(84) = -3.34^*$
Mean Parental Education (yrs)	13.6 (2.3)	14.1 (2.8)	$t(76) = 0.73$
BPRS Total	38.7 (8.0)	n/a	
Positive Symptoms ^a	9.0 (4.7)	n/a	
Negative Symptoms ^b	7.2 (2.4)	n/a	
Disorganization ^c	4.7 (1.2)	n/a	

Note: BPRS refers to the Brief Psychiatric Rating Scale. Parenthetical numbers following means represent standard deviations. Asterisks following test statistics represent $p < .05$.

^a BPRS items 8, 9, 10, and 11.

^b BPRS items 13, 16, 17, and 18.

^c BPRS items 12, 14, 15, and 24.

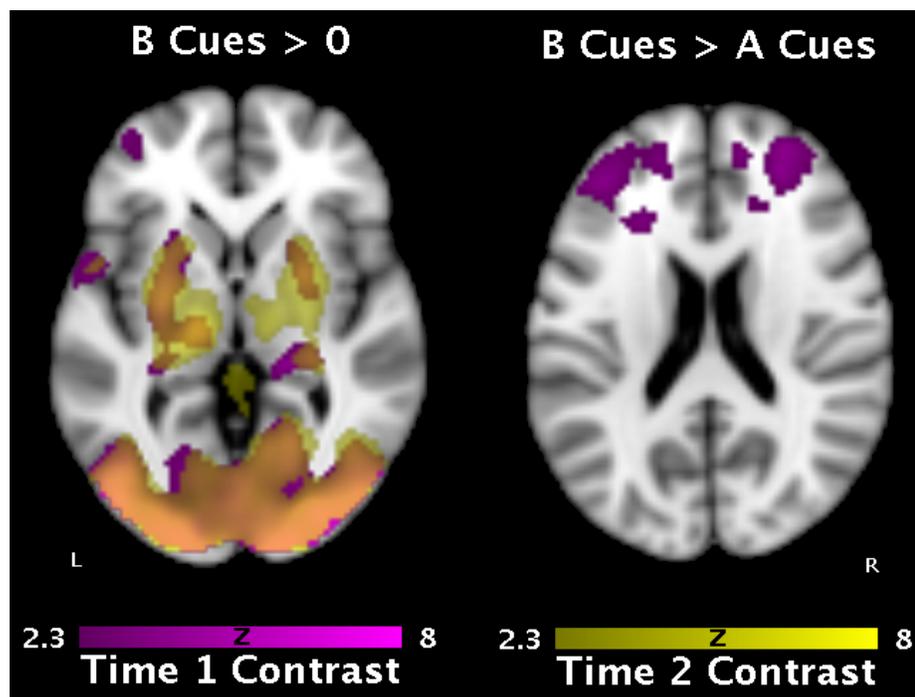
Table 4.2. Reliability of Analysis Methods

Method	Site									
	Baltimore		UC Davis		Minnesota		Washington University		Overall	
	HC	SZ	HC	SZ	HC	SZ	HC	SZ	HC	SZ
GLM	0.12	0	0.15	8	0.37	0.25	0.09	0.40	0.19	3
ICA	0.52	1	0.70	7	0.58	0.61	0.60	0.54	0.69	7
tFNC	0.67	6	0.63	8	0.56	0.51	0.50	0.54	0.56	0
dFN		0.0		0.0		-				0.0
C	-0.06	0	0.03	0	-0.04	0.02	-0.03	0.00	0.01	0

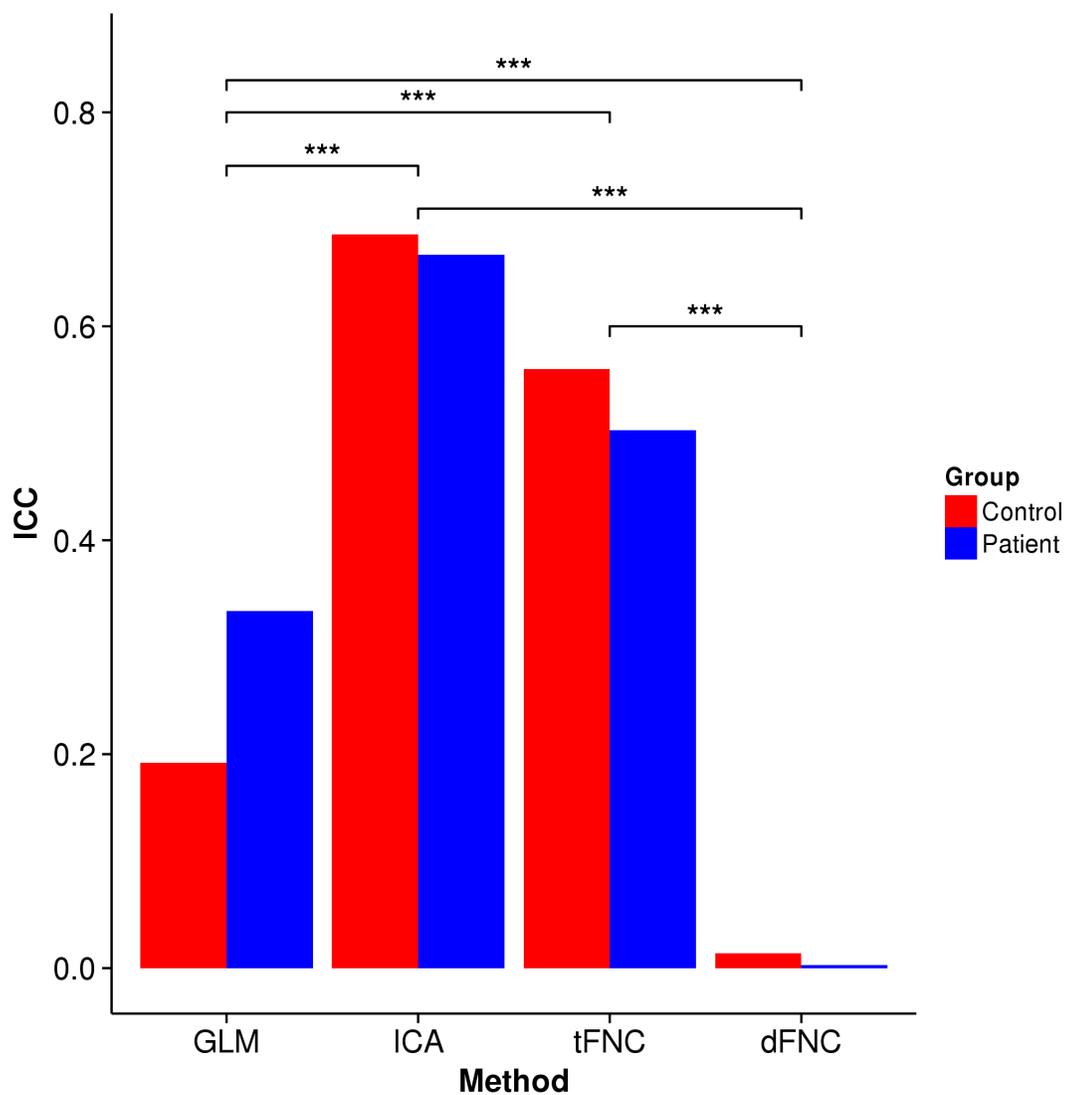
Table 4.3. tFNC Reliability by Magnitude of tFNC

Section of Data	Group	
	HC	SZ
Top 5%	0.70	0.62
Top 10%	0.68	0.60
Top 25%	0.66	0.57
Top 50%	0.63	0.54
Top 75%	0.60	0.52
Bottom 50%	0.49	0.47

Figure 4.1. GLM Activation Spatial Maps.

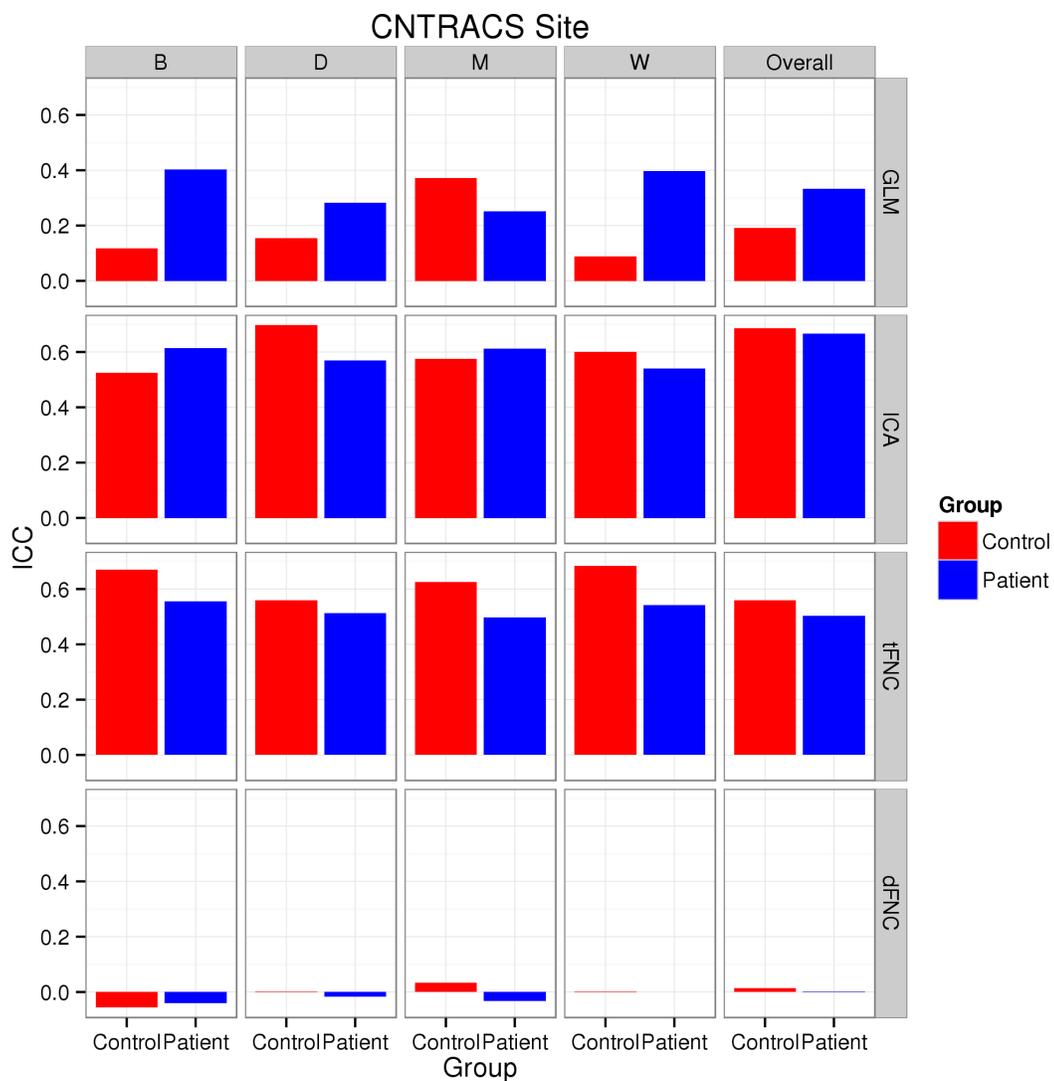


These images represent voxels significantly active in contrasts of B Cues (left) and B Cues minus A Cues (right). Purple represents activity from the first time point, and yellow represents activity from the second time point.

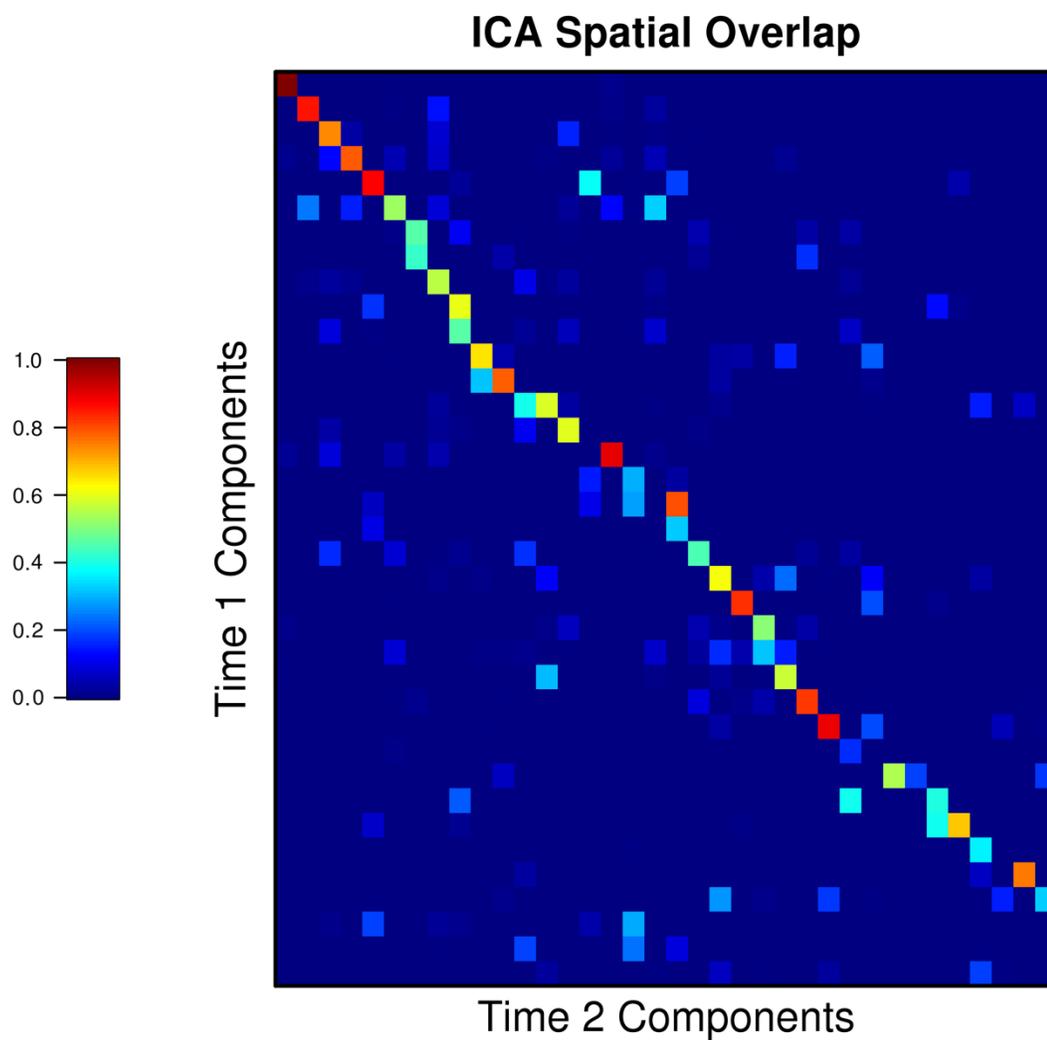
Figure 4.2. Average Reliability of Analysis Methods.

The bars represent the overall reliability of each group and each analysis method across data collection sites. Three asterisks (***) represents a significant difference between analysis methods (across groups) at the level of $p < .001$.

Figure 4.3. Average Reliability Separated by CNTRACS Site.

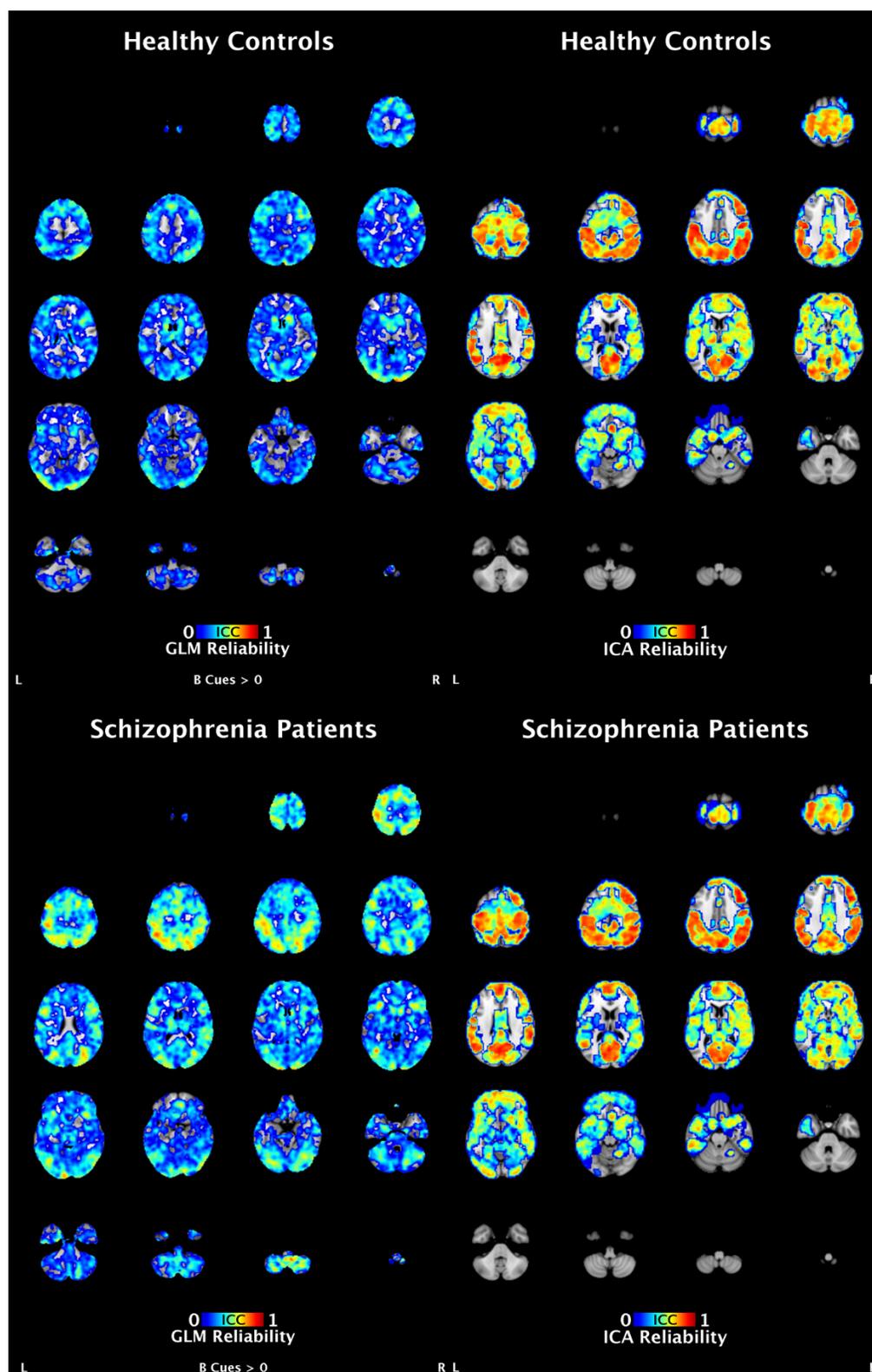


This figure shows reliability of each analysis method and each subject group broken down by data collection site. Overall reliability is included for comparison.

Figure 4.4. Spatial Overlap of ICA Results.

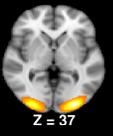
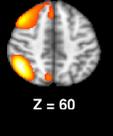
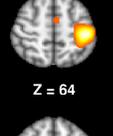
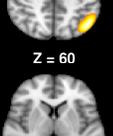
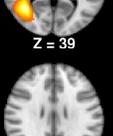
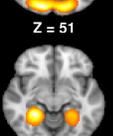
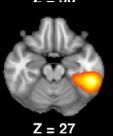
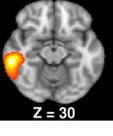
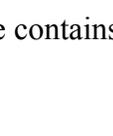
This figure shows the Dice overlap metric between thresholded group components at each time point. The higher the value, the more the components overlapped, where a Dice value of 1 indicates a 1 to 1 spatial relationship. The presence of a diagonal of relatively high Dice values indicates the ICA decomposition at each time point were spatially similar to each other at the group level.

Figure 4.5. Spatial Maps of ICA and GLM Reliability.



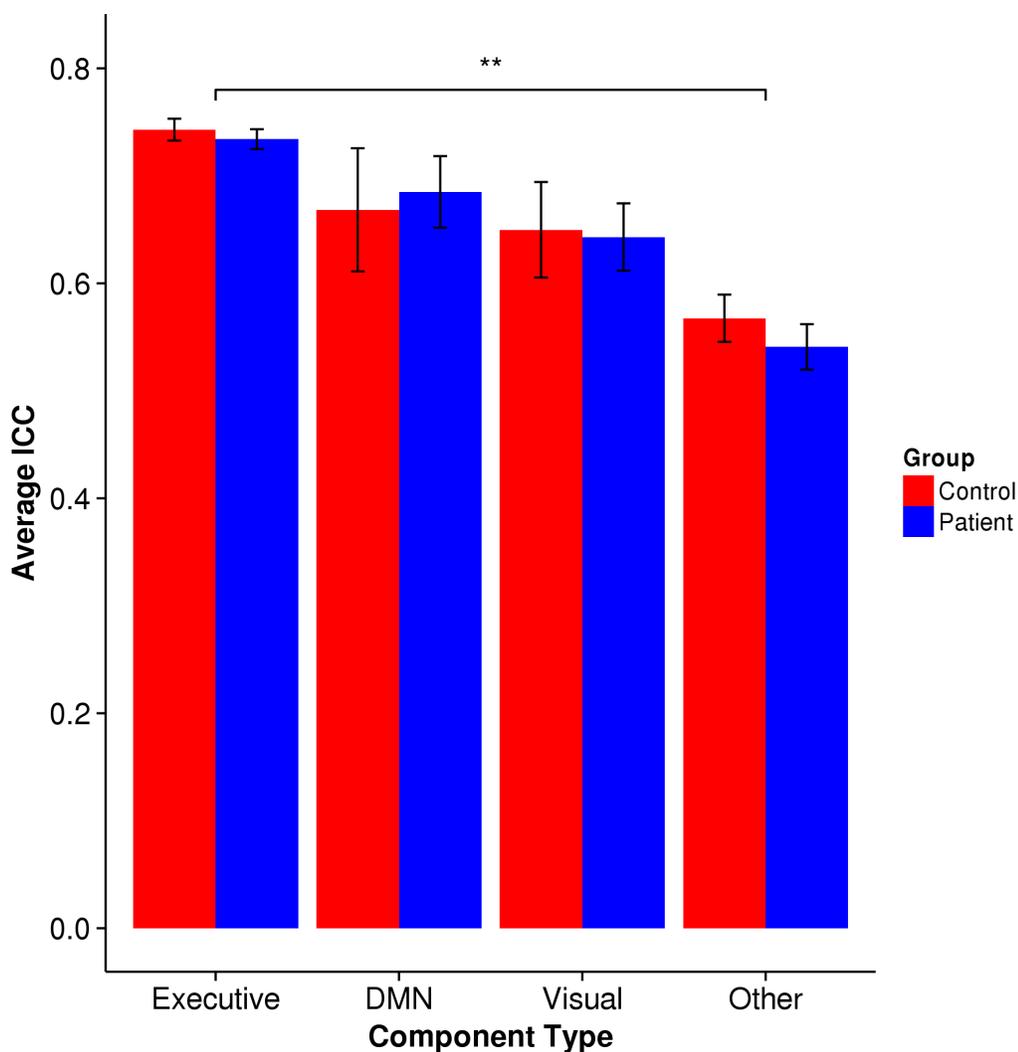
These figures show the obtained voxelwise ICC values for GLM and ICA separated by subject group. The GLM contrast in this figure is B Cues $>$ 0. The ICA maps include all non-artifact components.

Figure 4.6. Relationship between ICA Results and DPX B Cue Time Course.

Component	Image	HC1	HC2	HC_ICC	SZ1	SZ2	SZ_ICC
5	 Z = 37	0.12*	0.11*	0.33	0.14*	0.11*	0.27
10	 Z = 60	0.06*	0.04*	0.43	0.04*	0.04	0.46
13	 Z = 64	0.05*	0.05*	0.26	0.07*	0.04	0.52
15	 Z = 60	0.1*	0.09*	0.34	0.08*	0.08*	0.74
19	 Z = 60	0.04*	0.03*	0.17	0.01	0.03	0.49
25	 Z = 39	0.05*	0.05*	0.46	0.03*	0.04*	0.35
28	 Z = 51	0.04*	0.05*	0.3	0.03	0.03	0.42
36	 Z = 32	0.04*	0.04*	0.18	0.03	0.03	0.15
43	 Z = 36	0.05*	0.04*	0.56	0.04*	0.04*	0.3
58	 Z = 27	0.11*	0.08*	0.36	0.08*	0.08*	0.5
60	 Z = 30	0.04*	0.04*	0.49	0.01	0.01	0.29

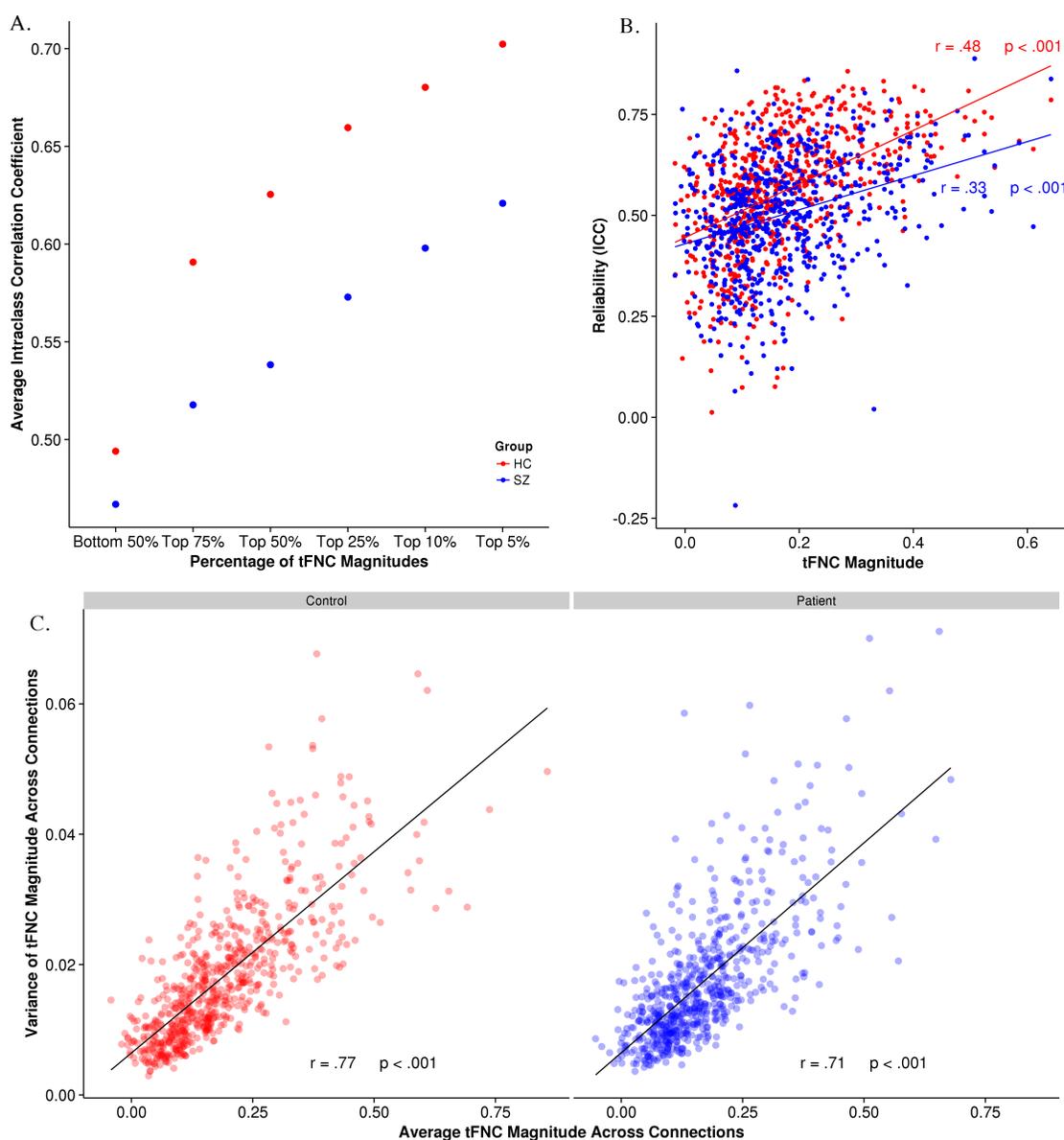
The figure contains every component that was found to be positively related to the B

Cues of the DPX task by at least one subject group at at least one time point. The data includes the average correlation with the B Cues of the DPX task for time points one and two as well as the ICC of those correlation values.

Figure 4.7. Average Reliability of ICA Separated by Component Type.

This figure shows the reliability of groups of components based on these categories: executive, default mode, visual, and other. Components within the “Other” category represent all non-artifact components not included in the other three categories and include such regions as motor cortex, ventral and medial PFC, temporal lobes, and subcortical regions.

Figure 4.8. Relationship between tFNC Magnitude and Reliability.



A) This shows the reliability of tFNC connections that surpass certain thresholds of tFNC magnitude (as measured at time 1). For example, the “Top 5%” section includes the highest 5% of tFNC connections. B) This is a basic scatterplot between tFNC magnitude at time 1 and reliability as measured by ICC. C) This shows the relationship between the

average magnitude of tFNC per connection and the variance of tFNC magnitude per connection divided by subject group.

Chapter 5: Overall Summary, Conclusions, and Future Directions

This series of studies had the collective goal of investigating the neural underpinnings of goal maintenance deficits in schizophrenia patients while estimating the test-retest reliability of the data analysis methods used in that investigation. Establishing a tool's reliability is an important step in quantifying its validity. If a measure is found to have high reliability, a researcher can be more confident in its results. However, as was the case in the present dissertation, differential reliability amongst brain regions might lead to erroneous conclusions from the results of fMRI analysis methods, even if reliability is relatively high across regions.

The first study demonstrated a replication of previous goal maintenance fMRI activation studies in schizophrenia patients while also providing a rationale for examining functional connectivity using the same data set. Study two found differential functional network connectivity between schizophrenia patients and healthy controls between networks previously implicated in top down cognitive control. It also showed that this connectivity between brain regions predicted task accuracy on the DPX task in both subject groups. Study three provided estimates for the reliability of GLM, ICA, tFNC, and dFNC in both subject groups both across data collection sites and broken down by site. It also examined differences in reliability between ICA components, which suggested the possibility of a measurement confound.

Study 1 Summary and Conclusions: GLM Analysis

The results of this study were important for replicating previous findings with other goal maintenance tasks as well as for providing a rationale for study 2. This was the first fMRI study to test group differences between SZ and HC while subjects performed the DPX task. The replication of previous findings of hypoactivation in MFG in SZ compared with HC suggests that the DPX is measuring the same construct as the expectancy AX-CPT while doing so in a more efficient way. These findings also bolster the idea that MFG is integral in cognitive control.

However, a major limitation of this study was the inability to detect significant correlations between GLM activation results and either DPX task performance or psychosis symptom severity. As was noted in chapter 1 of this dissertation, poor reliability in a measure limits the possible correlation between that measure and another variable. This possibility represented justification for examining the reliability of GLM as an analysis method of DPX fMRI data. Another explanation for the failure to observe correlations with symptom measures involves the specific symptom measure used in this study. In previous studies that observed such correlations between BOLD activity and disorganization symptoms (e.g., A. W. MacDonald, Carter, et al., 2005), a composite measure including the BPRS as well as SANS/SAPS was used as opposed to only the BPRS in the current study. Other studies that failed to find such correlations also used the BPRS only. In the current sample, the disorganization subscale of the BPRS had a Cronbach's alpha of .46 (whereas the positive symptom subscale had alpha = .71). It is possible that the disorganization subscale of the BPRS is insufficiently reliable on its own to allow detection of symptom/activation correlations.

This study additionally arrived at a minimum scan length to observe group activation differences. This result is helpful for future studies that use the DPX in imaging studies that include SZ, because it can potentially reduce the cost inherent in such studies as well as the discomfort of participants. However, given the reliability estimates obtained in study 3, longer scans are recommended to increase reliability.

Finally, the activation patterns observed to be different between SZ and HC included MFG and posterior parietal lobe. These two regions routinely appear in functional connectivity studies and have long been theorized to constitute an “executive” network that is involved in cognitive control and has been previously found to be differentially active in SZ compared with HC (e.g., Poppe et al., 2015). These activation results did not allow a direct testing of the functional connectivity of this network, which provided a clear rationale for study 2.

A future direction for this study might be to examine the specificity of these effects to SZ. The first step might be to determine if hypoactivation in MFG and posterior parietal is present in individuals with bipolar disorder with psychotic features or if this pattern of activation is specific to SZ.

Study 2 Summary and Conclusions: ICA Analyses

The examination of the same dataset as in study 1 using ICA allowed for an assessment of functional connectivity. This study marked the first to employ ICA, tFNC, and dFNC of DPX imaging data. The tFNC results included five connections that were stronger in HC compared with SZ. Of these, one involved an executive component including regions implicated in study 1. The strength of this component’s connection

with a salience network, which included anterior insula, frontal operculum, paracingulate cortex, and superior frontal regions, predicted DPX task performance in SZ (but not HC). Additionally, the lag feature of tFNC showed that the salience network preceded the executive network. These results provide support for the notion that the salience network in some way “tags” relevant information and forwards it to the executive network to guide behavior. The fact that HC’s task performance was not predicted by this connection’s strength may indicate that this communication from salience network to executive network represents a hurdle that must be cleared, but once cleared other neural functions are also required for cognitive control.

In the dFNC analysis, only the connection between left and right executive components was found to be significantly task modulated across groups, and that connection was differentially task modulated between groups. The degree to which the connection between left and right executive components was task modulated predicted task performance in HC but not SZ. This is an intriguing finding, as it is showing that the degree to which left and right executive components coordinate during key portions of the DPX task is important for successful task performance. This finding shows that constant synchrony between these two components is not necessary or helpful, but rather coordination during key parts of the task.

One major limitation of this study was the failure to observe ICA task-relationship differences between SZ and HC in right executive component as in previous research (Poppe et al., 2015). This failure may stem from the differential attributes of the DPX and expectancy AX tasks, but it could also result from any of the various sample and study design differences between the two studies. However, it may also reflect the findings of

the tFNC and dFNC analyses, indicating that task modulation of connections between bilateral executive components may be more important than the activity of the executive components correlating with the task themselves.

Future directions in this domain involve a closer examination of the salience network and the two executive networks. Dynamic causal modeling may be a good choice for this examination. Another direction may be to examine these connections during different tasks or even during the resting state. If the explanations I have proposed are correct, tasks or activities that do not require cognitive control should not produce group differences in tFNC between the salience network and executive, and the strength of that connection should not predict task performance.

Study 3 Summary and Conclusions: Reliability Analyses

The need for reliability studies of fMRI data analysis methods cannot be overstated. It is largely unknown how several factors affect the reliability of even widely used metrics (such as preprocessing options on GLM analyses) and healthy controls, let alone how such factors interact with clinical samples. As there has never been a reliability study of tFNC or dFNC, this study was important to do. Likewise, to my knowledge there has never been a direct comparison of the reliability of GLM and ICA in schizophrenia patients during a cognitive task, much less one that allowed for testing multisite effects on reliability. The results of the study were surprising but hopefully useful for future studies.

This study found that GLM had generally poor reliability and that it was also highly variable across data collection sites. Likewise, dFNC had very poor reliability

(essentially zero). tFNC had fair reliability across sites, but it also showed some variability across sites. ICA had the highest reliability and was stable across sites. These results suggest that ICA and tFNC are sufficiently reliable analysis methods across sites and subject groups, and it showed that ICA reliability was not sensitive to site effects. This point suggests that ICA may be a more appropriate analysis method when combining data from different sites, as is becoming a more common practice.

Perhaps the most important finding was that executive components were measured significantly more reliably than other components, besides visual and default mode components. Given that executive networks and the default mode network are some of the most commonly found networks to be found dysfunctional in SZ whether at rest or during task, it is entirely possible that differential reliability is causing a measurement confound that is at least augmenting the relative difference between subject groups compared with with other networks. A future direction would be to conduct a more direct analysis of differential reliability amongst brain regions to determine how much of a measurement confound may exist.

One limitation of the current study is that other common measures of fMRI data were not included and that only task-based data were available to be studied. Although these facts reduce the interpretations that can be made based on the results of this study, I feel that its results will still be helpful for researchers hoping to use any of the four methods included in the study.

In general, I hope that the material contained in this dissertation is helpful to researchers going forward.

Bibliography

Abou Elseoud, A., Littow, H., Remes, J., Starck, T., Nikkinen, J., Nissilä, J., ...

Kiviniemi, V. (2011). Group-ICA Model Order Highlights Patterns of Functional Brain Connectivity. *Frontiers in Systems Neuroscience*, 5(June), 37.

<http://doi.org/10.3389/fnsys.2011.00037>

Ambrosini, E., & Vallesi, A. (2015). Asymmetry in Prefrontal Resting-state EEG

Spectral Power Underlies Individual Differences in Phasic and Sustained Cognitive Control. *NeuroImage*, 124, 843–857.

<http://doi.org/10.1016/j.neuroimage.2015.09.035>

Andreasen, N. C., Nopoulos, P., O'Leary, D. S., Miller, D. D., Wassink, T., & Flaum, M.

(1999). Defining the phenotype of schizophrenia: Cognitive dysmetria and its neural mechanisms. *Biological Psychiatry*. [http://doi.org/10.1016/S0006-3223\(99\)00152-3](http://doi.org/10.1016/S0006-3223(99)00152-3)

Arbabshirani, M. R., & Calhoun, V. D. (2011). Functional network connectivity during

rest and task: Comparison of healthy controls and schizophrenic patients. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4418–4421. <http://doi.org/10.1109/IEMBS.2011.6091096>

<http://doi.org/10.1109/IEMBS.2011.6091096>

Assaf, M., Jagannathan, K., Calhoun, V., Kraut, M., Hart, J., & Pearlson, G. (2010).

Temporal Sequence of Hemispheric Network Activation during Semantic

Processing: A Functional Network Connectivity Analysis. *Brain Cogn.*, 70(2), 238–

246. <http://doi.org/10.1016/j.bandc.2009.02.007>.Temporal

Barch, D. M., Carter, C. S., Braver, T. S., Sabb, F. W., MacDonald, A., Noll, D. C., &

Cohen, J. D. (2001). Selective deficits in prefrontal cortex function in medication-

naive patients with schizophrenia. *Archives of General Psychiatry*, 58(3), 280–288.

<http://doi.org/10.1001/archpsyc.58.3.280>

Beckmann, Mackay, Filippini, & Smith. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *NeuroImage*, 47(Suppl 1), S148.

<http://doi.org/10.1073/pnas.0811879106>

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*, 23(2), 137–

152. <http://doi.org/10.1109/TMI.2003.822821>

Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage*, 25(1), 294–311.

<http://doi.org/10.1016/j.neuroimage.2004.10.043>

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, 1191,

133–155. <http://doi.org/10.1111/j.1749-6632.2010.05446.x>

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine : Official Journal of the Society of Magnetic Resonance in*

Medicine / Society of Magnetic Resonance in Medicine, 34(4), 537–541.

<http://doi.org/10.1002/mrm.1910340409>

Blackman, R. K., Crowe, D. A., DeNicola, A. L., Sakellaridi, S., MacDonald, Angus W., I., & Chafee, M. V. (n.d.). Monkey prefrontal neurons reflect logical operations for

cognitive control in a variant of the AX continuous performance task (AX-CPT).

Journal of Neuroscience.

Blautzik, J., Keeser, D., Berman, A., Paolini, M., Kirsch, V., Mueller, S., ... Meindl, T.

- (2012). Long-Term Test-Retest Reliability of Resting-State Networks in Healthy Elderly Subjects and Mild Cognitive Impairment Patients. *J Alzheimers Dis*.
<http://doi.org/10.3233/JAD-111970>
- Braver, T. S., & Bongiolatti, S. R. (2002). The role of frontopolar cortex in subgoal processing during working memory. *NeuroImage*, *15*(3), 523–536.
<http://doi.org/10.1006/nimg.2001.1019>
- Brown, G. D., Yamada, S., & Sejnowski, T. J. (2001). Independent component analysis at the neural cocktail party. *Trends in Neurosciences*. [http://doi.org/10.1016/S0166-2236\(00\)01683-0](http://doi.org/10.1016/S0166-2236(00)01683-0)
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, *10*(3), 186–198.
<http://doi.org/10.1038/nrn2575>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.
<http://doi.org/10.1038/nrn3475>
- Calhoun, V. D., & Adali, T. (2012). Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery. *IEEE Reviews in Biomedical Engineering*, *5*, 60–73.
<http://doi.org/10.1109/RBME.2012.2211076>
- Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, *14*(3), 140–151. <http://doi.org/10.1002/hbm.1048>

- Calhoun, V. D., Eichele, T., & Pearlson, G. (2009). Functional brain networks in schizophrenia: a review. *Frontiers in Human Neuroscience*, 3(August), 17.
<http://doi.org/10.3389/neuro.09.017>
- Cao, H., Dixon, L., Meyer-Lindenberg, A., & Tost, H. (2016). Functional connectivity measures as schizophrenia intermediate phenotypes: advances, limitations, and future directions. *Current Opinion in Neurobiology*, 36, 7–14.
<http://doi.org/10.1016/j.conb.2015.07.008>
- Carter, C. S., & Barch, D. M. (2007). Cognitive Neuroscience-Based Approaches to Measuring and Improving Treatment Effects on Cognition in Schizophrenia: The CNTRICS Initiative. *Schizophrenia Bulletin*, 33(5), 1131–1137.
- Casey, B. J., Cohen, J. D., O'Craven, K., Davidson, R. J., Irwin, W., Nelson, C. a., ... Turski, P. a. (1998). Reproducibility of fMRI results across four institutions using a spatial working memory task. *NeuroImage*, 8(3), 249–61.
<http://doi.org/10.1006/nimg.1998.0360>
- Celone, K. a, Calhoun, V. D., Dickerson, B. C., Atri, A., Chua, E. F., Miller, S. L., ... Sperling, R. a. (2006). Alterations in memory networks in mild cognitive impairment and Alzheimer's disease: an independent component analysis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 26(40), 10222–31. <http://doi.org/10.1523/JNEUROSCI.2250-06.2006>
- Chapman, L., & Chapman, J. (1973). Problems in the measurement of cognitive deficits. *Psychological Bulletin*.
- Chen, E. E., & Small, S. L. (2007). Test–retest reliability in fMRI of language: Group and task effects. *Brain and Language*, 102(2), 176–185.

<http://doi.org/10.1016/j.bandl.2006.04.015>

- Choi, K. H., Wykes, T., & Kurtz, M. M. (2013). Adjunctive pharmacotherapy for cognitive deficits in schizophrenia: Meta-analytical investigation of efficacy. *British Journal of Psychiatry*. <http://doi.org/10.1192/bjp.bp.111.107359>
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–137.
- Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *99*(1), 45–77. <http://doi.org/10.1037/0033-295X.99.1.45>
- Cohen, J. R., Gallen, C. L., Jacobs, E. G., Lee, T. G., & D'Esposito, M. (2014). Quantifying the reconfiguration of intrinsic networks during working memory. *PloS One*, *9*(9), e106636. <http://doi.org/10.1371/journal.pone.0106636>
- Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D. H., ... Calhoun, V. D. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, *5*(July), 298–308. <http://doi.org/10.1016/j.nicl.2014.07.003>
- De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). An introduction to independent component analysis. *Journal of Chemometrics*, *14*(3), 123–149. [http://doi.org/10.1002/1099-128X\(200005/06\)14:3<123::AID-CEM589>3.0.CO;2-1](http://doi.org/10.1002/1099-128X(200005/06)14:3<123::AID-CEM589>3.0.CO;2-1)
- Delawalla, Z., Csernansky, J. G., & Barch, D. M. (2008). Prefrontal Cortex Function in Nonpsychotic Siblings of Individuals with Schizophrenia. *Biological Psychiatry*, *63*(5), 490–497. <http://doi.org/10.1016/j.biopsych.2007.05.007>

- Deserno, L., Sterzer, P., Wüstenberg, T., Heinz, A., & Schlagenhaut, F. (2012). Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(1), 12–20. <http://doi.org/10.1523/JNEUROSCI.3405-11.2012>
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*, 297–302. <http://doi.org/10.2307/1932409>
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. <http://doi.org/10.1016/j.tics.2008.01.001>.
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., ... Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 11073–11078. <http://doi.org/10.1073/pnas.0704320104>
- Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., & Calhoun, V. D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Human Brain Mapping*, *32*(12), 2075–2095. <http://doi.org/10.1002/hbm.21170>
- Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., ... Di Salle, F. (2005). Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage*, *25*(1), 193–205. <http://doi.org/10.1016/j.neuroimage.2004.10.042>
- Filippini, N., MacIntosh, B. J., Hough, M. G., Goodwin, G. M., Frisoni, G. B., Smith, S. M., ... Mackay, C. E. (2009). Distinct patterns of brain activity in young carriers of

the APOE-epsilon4 allele. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), 7209–14.

<http://doi.org/10.1073/pnas.0811879106>

First, M. B. et, Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. for DSMIV.*

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Patient Edition (SCID-I/P, 11/2002 revision). for DSMIV.* <http://doi.org/M>

Fisher, R. a. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507–521.

<http://doi.org/10.2307/2331838>

Fleiss, J. L. (1986). *Design and analysis of clinical experiments* (Vol. 73). John Wiley & Sons.

Ford, J. M., Roach, B. J., Jorgensen, K. W., Turner, J. a, Brown, G. G., Notestine, R., ... Mathalon, D. H. (2009). Tuning in to the voices: a multisite fMRI study of auditory hallucinations. *Schizophrenia Bulletin*, 35(1), 58–66.

<http://doi.org/10.1093/schbul/sbn140>

Forsythe, G. E., Moler, C. B., & Malcolm, M. A. (1977). Computer methods for mathematical computations.

Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., ... Potkin, S. G. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, 29(8), 958–972. <http://doi.org/10.1002/hbm.20440>

- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13–36. <http://doi.org/10.1089/brain.2011.0008>
- Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome? *Clinical Neuroscience (New York, N.Y.)*, *3*, 89–97. <http://doi.org/10.1007/s00117-004-1160-3>
- Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*. <http://doi.org/10.1038/jcbfm.1993.4>
- Gold, J. M., Barch, D. M., Carter, C. S., Dakin, S., Harvey, P., Keefe, R., ... Strauss, M. E. (n.d.). Clinical, functional and inter-task correlations of measures developed by the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia consortium. *Schizophrenia Bulletin*.
- Gold, J. M., & Harvey, P. D. (1993). Cognitive Deficits in Schizophrenia. *Psychiatric Clinics of North America*, *16*(2), 295–312.
- Gountouna, V.-E., Job, D. E., McIntosh, A. M., Moorhead, T. W. J., Lymer, G. K. L., Whalley, H. C., ... Lawrie, S. M. (2010). Functional Magnetic Resonance Imaging (fMRI) reproducibility and variance components across visits and scanning sites with a finger tapping task. *NeuroImage*, *49*(1), 552–60. <http://doi.org/10.1016/j.neuroimage.2009.07.026>
- Guo, C. C., Kurth, F., Zhou, J., Mayer, E. a, Eickhoff, S. B., Kramer, J. H., & Seeley, W. W. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *NeuroImage*, *61*(4), 1471–1483. <http://doi.org/10.1016/j.neuroimage.2012.03.027>

- Harrison, P. J. (2000). Postmortem studies in schizophrenia. *Dialogues in Clinical Neuroscience*, 2(4), 349–57.
- He, H., Bustillo, J., Du, Y., Yu, Q., Jones, T., Jiang, T., ... Sui, J. (2015). Resting fMRI measures are associated with cognitive deficits in schizophrenia assessed by the MATRICS consensus cognitive battery, 9417(January 2016), 94171V–94171V–8. <http://doi.org/10.1117/12.2083171>
- Heinrichs, R. W. (2005). The primacy of cognition in schizophrenia. *The American Psychologist*, 60(3), 229–42. <http://doi.org/10.1037/0003-066X.60.3.229>
- Henderson, D., Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., ... MacDonald, A. W. (2012). Optimization of a goal maintenance task for use in clinical applications. *Schizophrenia Bulletin*, 38(1), 104–13. <http://doi.org/10.1093/schbul/sbr172>
- Hollingshead, A. de B., & Redlich, F. C. (1958). Social class and mental illness: A community study. *BOOK*, 442. <http://doi.org/10.1111/j.1939-0025.1959.tb00180.x>
- Holmes, A. J., MacDonald, A., Carter, C. S., Barch, D. M., Stenger, V. A., & Cohen, J. D. (2005). Prefrontal functioning during context processing in schizophrenia and major depression: An event-related fMRI study. *Schizophrenia Research*, 76(2-3), 199–206. <http://doi.org/10.1016/j.schres.2005.01.021>
- Hyvärinen, a, & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks : The Official Journal of the International Neural Network Society*, 13(4-5), 411–30. [http://doi.org/10.1016/S0893-6080\(00\)00026-5](http://doi.org/10.1016/S0893-6080(00)00026-5)
- Hyvärinen, A., & Oja, E. (1997). A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7), 1483–1492.

<http://doi.org/10.1162/neco.1997.9.7.1483>

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin Del La Société Vaudoise Des Sciences Naturelles*, 37(JANUARY 1901), 547–579. <http://doi.org/http://dx.doi.org/10.5169/seals-266450>

Jafri, M. J., Pearlson, G. D., Stevens, M., & Calhoun, V. D. (2008). A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *NeuroImage*, 39, 1666–1681. <http://doi.org/10.1016/j.neuroimage.2007.11.001>

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. M. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17, 825–841. <http://doi.org/10.1006/nimg.2002.1132>

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5, 143–156. [http://doi.org/10.1016/S1361-8415\(01\)00036-6](http://doi.org/10.1016/S1361-8415(01)00036-6)

Jeong, B., Choi, J., & Kim, J. W. (2012). MRI study on the functional and spatial consistency of resting state-related independent components of the brain network. *Korean Journal of Radiology*, 13(3), 265–274. <http://doi.org/10.3348/kjr.2012.13.3.265>

Jones, J. A. H., Sponheim, S. R., & MacDonald, A. W. (2010). The dot pattern expectancy task: reliability and replication of deficits in schizophrenia. *Psychological Assessment*, 22, 131–141. <http://doi.org/10.1037/a0017828>

Karbasforoushan, H., & Woodward, N. D. (2012). Resting-state networks in

schizophrenia. *Current Topics in Medicinal Chemistry*, 12(21), 2404–14.

<http://doi.org/10.2174/1568026611212210011>

Kiviniemi, V., Starck, T., Remes, J., Long, X., Nikkinen, J., Haapea, M., ... Tervonen, O.

(2009). Functional segmentation of the brain cortex using high model order group PICA. *Human Brain Mapping*, 30(12), 3865–3886.

<http://doi.org/10.1002/hbm.20813>

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, 399(6732), 148–151.

<http://doi.org/10.1038/20178>

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science (New York, N.Y.)*, 302(5648), 1181–1185.

<http://doi.org/10.1126/science.1088545>

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, 11(6), 229–235.

<http://doi.org/10.1016/j.tics.2007.04.005>

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., ... Fox, P. T. (2011). Behavioral Interpretations of Intrinsic Connectivity Networks. *Journal of Cognitive Neuroscience*. http://doi.org/10.1162/jocn_a_00077

Langlois, D., Chartier, S., & Gosselin, D. (2010). An Introduction to Independent

Component Analysis: InfoMax and FastICA Algorithms. *Tutorials in Quantitative Methods for Psychology*, 6(1), 31–38.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian

sources. *Neural Computation*, 11(2), 417–441.

<http://doi.org/10.1162/089976699300016719>

Lesh, T. A., Tanase, C., Geib, B. R., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., ...

Carter, C. S. (2015). A multimodal analysis of antipsychotic effects on brain structure and function in first-episode schizophrenia. *JAMA Psychiatry*, 72(3), 226–234. <http://doi.org/10.1001/jamapsychiatry.2014.2178>

Lesh, T. A., Westphal, A. J., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., Ragland, J.

D., ... Carter, C. S. (2013). Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. *NeuroImage: Clinical*, 2(1), 590–599. <http://doi.org/10.1016/j.nicl.2013.04.010>

Levandowsky, M., & Winter, D. (1971). Distance between sets. *Nature*, 234(5323), 34–35.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. *Stanford University Press*, 278–292. <http://doi.org/10.1002/bimj.19630050119>

Lewis, D. A., Pierri, J. N., Volk, D. W., Melchitzky, D. S., & Woo, T. U. W. (1999).

Altered GABA neurotransmission and prefrontal cortical dysfunction in schizophrenia. In *Biological Psychiatry* (Vol. 46, pp. 616–626).

[http://doi.org/10.1016/S0006-3223\(99\)00061-X](http://doi.org/10.1016/S0006-3223(99)00061-X)

Li, K., Guo, L., Nie, J., Li, G., & Liu, T. (2009). Review of methods for functional brain

connectivity detection using fMRI. *Computerized Medical Imaging and Graphics*, 33(2), 131–139. <http://doi.org/10.1016/j.compmedimag.2008.10.011>

Li, Y.-O., Adali, T., & Calhoun, V. D. (2007). Estimating the number of independent

- components for functional magnetic resonance imaging data. *Human Brain Mapping*, 28(11), 1251–1266. <http://doi.org/10.1002/hbm.20359>
- Lopez-Garcia, P., Lesh, T. A., Salo, T., Barch, D. M., MacDonald III, A. W., Gold, J. M., ... Carter, C. S. (2015). The neural circuitry supporting goal maintenance during cognitive control: a comparison of expectancy AX-CPT and dot probe expectancy paradigms. *Cognitive, Affective, & Behavioral Neuroscience*, 1–12.
- Ma, L., Wang, B., Chen, X., & Xiong, J. (2007). Detecting functional connectivity in the resting brain: a comparison between ICA and CCA. *Magnetic Resonance Imaging*, 25(1), 47–56. <http://doi.org/10.1016/j.mri.2006.09.032>
- MacDonald, a W., Cohen, J. D., Stenger, V. a, & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science (New York, N.Y.)*, 288(5472), 1835–1838. <http://doi.org/10.1126/science.288.5472.1835>
- MacDonald, A. W., & Carter, C. S. (2003). Event-related fMRI study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. *Journal of Abnormal Psychology*, 112(4), 689–697. <http://doi.org/10.1037/0021-843X.112.4.689>
- MacDonald, A. W., Carter, C. S., Kerns, J. G., Ursu, S., Barch, D. M., Holmes, A. J., ... Cohen, J. D. (2005). Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *American Journal of Psychiatry*, 162, 475–484. <http://doi.org/10.1176/appi.ajp.162.3.475>
- MacDonald, A. W., Goghari, V. M., Hicks, B. M., Flory, J. D., Carter, C. S., & Manuck,

- S. B. (2005). A convergent-divergent approach to context processing, general intellectual functioning, and the genetic liability to schizophrenia. *Neuropsychology*, *19*(6), 814–821. <http://doi.org/10.1037/0894-4105.19.6.814>
- MacDonald, A. W., Pogue-Geile, M. F., Johnson, M. K., & Carter, C. S. (2003). A specific deficit in context processing in the unaffected siblings of patients with schizophrenia. *Archives of General Psychiatry*, *60*, 57–65.
<http://doi.org/10.1001/archpsyc.60.1.57>
- Magnuson, M. E., Thompson, G. J., Schwarb, H., Pan, W.-J., McKinley, A., Schumacher, E. H., & Keilholz, S. D. (2015). Errors on interrupter tasks presented during spatial and verbal working memory performance are linearly linked to large-scale functional network connectivity in high temporal resolution resting state fMRI. *Brain Imaging and Behavior*, 854–867. <http://doi.org/10.1007/s11682-014-9347-3>
- Maïza, O., Mazoyer, B., Hervé, P.-Y., Razafimandimby, A., Dollfus, S., & Tzourio-Mazoyer, N. (2011). *Reproducibility of fMRI activations during a story listening task in patients with schizophrenia. Schizophrenia Research* (Vol. 128).
- Manoach, D. S., Halpern, E. F., Kramer, T. S., Chang, Y., Goff, D. C., Rauch, S. L., ... Gollub, R. L. (2001). Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects. *American Journal of Psychiatry*, *158*(6), 955–958. <http://doi.org/10.1176/appi.ajp.158.6.955>
- Manoliu, A., Riedl, V., Zherdin, A., Mühlau, M., Schwerthöffer, D., Scherr, M., ... Sorg, C. (2014). Aberrant dependence of default mode/central executive network interactions on anterior insular salience network activity in schizophrenia. *Schizophrenia Bulletin*, *40*(2), 428–437. <http://doi.org/10.1093/schbul/sbt037>

- Margulies, D. S., Böttger, J., Long, X., Lv, Y., Kelly, C., Schäfer, A., ... Villringer, A. (2010). Resting developments: A review of fMRI post-processing methodologies for spontaneous brain activity. *Magnetic Resonance Materials in Physics, Biology and Medicine*. <http://doi.org/10.1007/s10334-010-0228-5>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–82.
- McKeown, M. (2003). Independent component analysis of functional MRI: what is signal and what is noise? *Current Opinion in Neurobiology*, 13(5), 620–629. <http://doi.org/10.1016/j.conb.2003.09.012>
- McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., & Sejnowski, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6(3), 160–188. [http://doi.org/10.1002/\(SICI\)1097-0193\(1998\)6:3<160::AID-HBM5>3.0.CO;2-1](http://doi.org/10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBM5>3.0.CO;2-1)
- Meda, S. a., Gill, A., Stevens, M. C., Lorenzoni, R. P., Glahn, D. C., Calhoun, V. D., ... Pearlson, G. D. (2012). Differences in resting-state functional magnetic resonance imaging functional network connectivity between schizophrenia and psychotic bipolar probands and their unaffected first-degree relatives. *Biological Psychiatry*, 71(10), 881–889. <http://doi.org/10.1016/j.biopsych.2012.01.025>
- Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*. <http://doi.org/10.1016/j.tics.2011.08.003>
- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure & Function*, 214(5-6), 655–67.

<http://doi.org/10.1007/s00429-010-0262-0>

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function.

Annual Review of Neuroscience, 24(FEBRUARY 2001), 167–202.

<http://doi.org/10.1146/annurev.neuro.24.1.167>

Milner, D., & Goodale, M. (2012). *The Visual Brain in Action. The Visual Brain in*

Action. <http://doi.org/10.1093/acprof:oso/9780198524724.001.0001>

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2009). Meta-

analysis of 41 functional neuroimaging studies of executive function in

schizophrenia. *Archives of General Psychiatry*, 66(8), 811–22.

<http://doi.org/10.1001/archgenpsychiatry.2009.91>

Moran, L. V., Tagamets, M. A., Sampath, H., O'Donnell, A., Stein, E. A., Kochunov, P.,

& Hong, L. E. (2013). Disruption of anterior insula modulation of large-scale brain

networks in schizophrenia. *Biological Psychiatry*, 74(6), 467–474.

<http://doi.org/10.1016/j.biopsych.2013.02.029>

Niendam, T. A., Lesh, T. A., Yoon, J., Westphal, A. J., Hutchison, N., Daniel Ragland, J.,

... Carter, C. S. (2014). Impaired context processing as a potential marker of

psychosis risk state. *Psychiatry Research - Neuroimaging*, 221(1), 13–20.

<http://doi.org/10.1016/j.psychresns.2013.09.001>

Nikiforuk, A., Hołuj, M., Kos, T., & Popik, P. (2016). The effects of a 5-HT_{5A} receptor

antagonist in a ketamine-based rat model of cognitive dysfunction and the negative

symptoms of schizophrenia. *Neuropharmacology*, 105, 351–360.

<http://doi.org/10.1016/j.neuropharm.2016.01.035>

Nowrangi, M. a, Lyketsos, C., Rao, V., & Munro, C. a. (2014). Systematic review of

- neuroimaging correlates of executive functioning: converging evidence from different clinical populations. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 26, 114–25. <http://doi.org/10.1176/appi.neuropsych.12070176>
- Ojemann, J. G., Buckner, R. L., Akbudak, E., Snyder, a Z., Ollinger, J. M., McKinstry, R. C., ... Conturo, T. E. (1998). Functional MRI studies of word-stem completion: reliability across laboratories and comparison to blood flow imaging with PET. *Human Brain Mapping*, 6(4), 203–15. [http://doi.org/10.1002/\(SICI\)1097-0193\(1998\)6:4<203::AID-HBM2>3.0.CO;2-7](http://doi.org/10.1002/(SICI)1097-0193(1998)6:4<203::AID-HBM2>3.0.CO;2-7) [pii]
- Palaniyappan, L., Mallikarjun, P., Joseph, V., White, T. P., & Liddle, P. F. (2011). Reality distortion is related to the structure of the salience network in schizophrenia. *Psychological Medicine*, 41(08), 1701–1708.
- Palaniyappan, L., Simmonite, M., White, T., Liddle, E., & Liddle, P. (2013). Neural primacy of the salience processing system in schizophrenia. *Neuron*, 79(4), 814–828. <http://doi.org/10.1016/j.neuron.2013.06.027>
- Perlstein, W. M., Dixit, N. K., Carter, C. S., Noll, D. C., & Cohen, J. D. (2003). Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biological Psychiatry*, 53(1), 25–38. [http://doi.org/10.1016/S0006-3223\(02\)01675-X](http://doi.org/10.1016/S0006-3223(02)01675-X)
- Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., & MacDonald, A. W. (2016). Reduced Frontoparietal Activity in Schizophrenia Is Linked to a Specific Deficit in Goal Maintenance: A Multisite Functional Imaging Study. *Schizophrenia Bulletin*. <http://doi.org/10.1093/schbul/sbw036>
- Poppe, A. B., Carter, C. S., Minzenberg, M. J., & MacDonald, A. W. (2015). Task-based

functional connectivity as an indicator of genetic liability to schizophrenia.

Schizophrenia Research, 162(1-3), 118–23.

<http://doi.org/10.1016/j.schres.2014.11.022>

Poppe, A. B., Wisner, K., Atluri, G., Lim, K. O., Kumar, V., & Macdonald, A. W.

(2013). Toward a neurometric foundation for probabilistic independent component analysis of fMRI data. *Cognitive, Affective & Behavioral Neuroscience*, 13(3), 641–59. <http://doi.org/10.3758/s13415-013-0180-8>

Ragland, J. D., Ranganath, C., Harms, M., Barch, D., Gold, J., Layher, E., ... Carter, C.

(2015). Functional and Neuroanatomic Specificity of Episodic Memory Dysfunction in Schizophrenia: A Functional Magnetic Resonance Imaging Study of the Relational and Item-Specific Encoding Task. *JAMA Psychiatry*.

<http://doi.org/10.1001/jamapsychiatry.2015.0276>

Raichle, M. E., MacLeod, a M., Snyder, a Z., Powers, W. J., Gusnard, D. a, & Shulman,

G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 676–682.

<http://doi.org/10.1073/pnas.98.2.676>

Ramsay, I. S., & Macdonald, A. W. (2015). Brain Correlates of Cognitive Remediation in

Schizophrenia: Activation Likelihood Analysis Shows Preliminary Evidence of Neural Target Engagement. *Schizophrenia Bulletin*, 41(6), 1276–1284.

<http://doi.org/10.1093/schbul/sbv025>

Ray, K. L., McKay, D. R., Fox, P. M., Riedel, M. C., Uecker, A. M., Beckmann, C. F., ...

Laird, A. R. (2013). ICA model order selection of task co-activation networks.

Frontiers in Neuroscience, (7 DEC). <http://doi.org/10.3389/fnins.2013.00237>

- Revelle, W. (2010). psych: Procedures for psychological, psychometric, and personality research. *Northwestern University: Evanston, Illinois*, 0–90.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*(5), 465–471.
[http://doi.org/10.1016/0005-1098\(78\)90005-5](http://doi.org/10.1016/0005-1098(78)90005-5)
- Ritsner, M. S., & Gottesman II. (2011). *The Schizophrenia Construct After 100 Years of Challenges. Handbook of Schizophrenia Spectrum Disorders, Vol 1: Conceptual Issues and Neurobiological Advances*. http://doi.org/10.1007/978-94-007-0837-2_1
- Sakoğlu, Ü., Pearlson, G. D., Kiehl, K. a., Wang, Y. M., Michael, A. M., & Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *23*(5-6), 351–366. <http://doi.org/10.1007/s10334-010-0197-8>
- Schneider, F., Habel, U., Reske, M., Kellermann, T., Stöcker, T., Shah, N. J., ... Gaebel, W. (2007). Neural correlates of working memory dysfunction in first-episode schizophrenia patients: An fMRI multi-center study. *Schizophrenia Research*, *89*(1-3), 198–210. <http://doi.org/10.1016/j.schres.2006.07.021>
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(9), 2349–56.
<http://doi.org/10.1523/JNEUROSCI.5587-06.2007>
- Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J. G., Bilecen, D., ... Di Salle, F. (2002). Spatiotemporal pattern of neural processing in the human auditory

cortex. *Science (New York, N.Y.)*, 297(5587), 1706–1708.

<http://doi.org/10.1126/science.1074355>

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. *Archives of General Psychiatry*, 53, 1105–1112. [http://doi.org/10.1016/0920-9964\(95\)95347-C](http://doi.org/10.1016/0920-9964(95)95347-C)

Sheffield, J. M., & Barch, D. M. (2016). Cognition and resting-state functional connectivity in schizophrenia. *Neuroscience & Biobehavioral Reviews*, 61, 108–120. <http://doi.org/10.1016/j.neubiorev.2015.12.007>

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <http://doi.org/10.1037/0033-2909.86.2.420>

Silverstein, S. M., Harms, M. P., Carter, C. S., Gold, J. M., Keane, B. P., MacDonald, A., ... Barch, D. M. (2015). Cortical contributions to impaired contour integration in schizophrenia. *Neuropsychologia*, 75, 469–80. <http://doi.org/10.1016/j.neuropsychologia.2015.07.003>

Simak, A. A., Liou, M., Zhigalov, A. Y., Liou, J.-W., & Cheng, P. E. (2012). Reliability maps in event related functional MRI experiments. *FUNCTIONAL MAGNETIC RESONANCE IMAGING--ADVANCED NEUROIMAGING APPLICATIONS*, 149.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17, 143–155. <http://doi.org/10.1002/hbm.10062>

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., ... Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. In *NeuroImage* (Vol. 23).

<http://doi.org/10.1016/j.neuroimage.2004.07.051>

- Snitz, B. E., MacDonald, A. W., & Carter, C. S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: A meta-analytic review of putative endophenotypes. *Schizophrenia Bulletin*. <http://doi.org/10.1093/schbul/sbi048>
- Somandepalli, K., Kelly, C., Reiss, P. T., Zuo, X.-N., Craddock, R. C., Yan, C.-G., ... Di Martino, A. (2015). Short-term test–retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. *Developmental Cognitive Neuroscience*, *15*, 83–93. <http://doi.org/10.1016/j.dcn.2015.08.003>
- Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., & Jiang, T. (2008). Brain spontaneous functional connectivity and intelligence. *NeuroImage*, *41*(3), 1168–76. <http://doi.org/10.1016/j.neuroimage.2008.02.036>
- Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Networks*, *105*(34), 12569–12574. <http://doi.org/10.1073/pnas.0800005105>
- Stone, J. V. (2002). Independent component analysis: An introduction. *Trends in Cognitive Sciences*. [http://doi.org/10.1016/S1364-6613\(00\)01813-1](http://doi.org/10.1016/S1364-6613(00)01813-1)
- Stone, J. V. (2002). Independent Component Analysis: A Tutorial Introduction. *Trends in Cognitive Sciences*, *6*(2), 59–64.
- Stone, J. V., Porrill, J., Porter, N. R., & Wilkinson, I. D. (2002). Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *NeuroImage*, *15*(2), 407–421. <http://doi.org/10.1006/nimg.2001.0986>
- Strauss, M. E., McLouth, C. J., Barch, D. M., Carter, C. S., Gold, J. M., Luck, S. J., ...

- Silverstein, S. M. (2013). Temporal Stability and Moderating Effects of Age and Sex on CNTRaCS Task Performance. *Schizophrenia Bulletin*.
<http://doi.org/10.1093/schbul/sbt089>
- Stuss, D. T. (2011). Functions of the frontal lobes: relation to executive functions. *Journal of the International Neuropsychological Society : JINS*, 17(5), 759–765.
<http://doi.org/10.1017/S1355617711000695>
- Suchy, Y. (2009). Executive functioning: Overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine*, 37(2), 106–116.
<http://doi.org/10.1007/s12160-009-9097-4>
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6), 2769–2794.
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99. <http://doi.org/10.2307/3001913>
- van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519–534.
<http://doi.org/10.1016/j.euroneuro.2010.03.008>
- Vlieger, E.-J., Lavini, C., Majoie, C. B., & den Heeten, G. J. (2003). Reproducibility of functional MR imaging results using two different MR systems. *AJNR. American Journal of Neuroradiology*, 24(4), 652–7.
- Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *NeuroImage*, 16(2), 401–414.
<http://doi.org/10.1006/nimg.2002.1109>

- Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A. C. (2015). Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. *Journal of Cognitive Neuroscience*, *27*(10), 2019–2034.
- Wang, Q., Cheung, C., Deng, W., Li, M., Huang, C., Ma, X., ... Li, T. (2013). Frontoparietal white matter microstructural deficits are linked to performance IQ in a first-episode schizophrenia Han Chinese sample. *Psychological Medicine*, *43*(10), 2047–56. <http://doi.org/10.1017/S0033291712002905>
- Wechsler, D. (2001). Wechsler Test of Adult Reading. San Antonio, TX: The Psychological Corporation.
- Weir, J. P. J. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, *19*(1), 231–40. <http://doi.org/10.1519/15184.1>
- White, T. P., Joseph, V., Francis, S. T., & Liddle, P. F. (2010). Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. *Schizophrenia Research*, *123*(2-3), 105–115. <http://doi.org/10.1016/j.schres.2010.07.020>
- Wisner, K. M., Atluri, G., Lim, K. O., & MacDonald, A. W. (2013). Neurometrics of intrinsic connectivity networks at rest using fMRI: Retest reliability and cross-validation using a meta-level method. *NeuroImage*, *76*, 236–251. <http://doi.org/10.1016/j.neuroimage.2013.02.066>
- Woolrich, M. W., Behrens, T. E. J., & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage*, *21*, 1748–1761. <http://doi.org/10.1016/j.neuroimage.2003.12.024>

- Wurnig, M. C., Rath, J., Klinger, N., Höllinger, I., Geissler, A., Fischmeister, F. P., ...
Beisteiner, R. (2013). Variability of Clinical Functional MR Imaging Results: A
Multicenter Study. *Radiology*, *268*(2), 521–531.
<http://doi.org/10.1148/radiol.13121357>
- Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis
of cognitive remediation for schizophrenia: Methodology and effect sizes. *American
Journal of Psychiatry*. <http://doi.org/10.1176/appi.ajp.2010.10060855>
- Yang, H. H., & Amari, S. (1997). Adaptive Online Learning Algorithms for Blind
Separation: Maximum Entropy and Minimum Mutual Information. *Neural
Computation*, *9*(7), 1457–1482. <http://doi.org/10.1162/neco.1997.9.7.1457>
- Yendiki, A., Greve, D. N., Wallace, S., Vangel, M., Bockholt, J., Mueller, B. A., ...
Gollub, R. L. (2010). Multi-site characterization of an fMRI working memory
paradigm: Reliability of activation indices. *NeuroImage*, *53*(1), 119–131.
<http://doi.org/10.1016/j.neuroimage.2010.02.084>
- Yoon, J. H., Minzenberg, M. J., Ursu, S., Walters, R., Wendelken, C., Ragland, J. D., &
Carter, C. S. (2008). Association of dorsolateral prefrontal cortex dysfunction with
disrupted coordinated brain activity in schizophrenia: Relationship with impaired
cognition, behavioral disorganization, and global function. *American Journal of
Psychiatry*, *165*(8), 1006–1014. <http://doi.org/10.1176/appi.ajp.2008.07060945>
- Zar, J. H. (2010). *Biostatistical Analysis*. Prentice Hall New Jersey USA.
<http://doi.org/10.1037/0012764>
- Zhang, H., Zuo, X. N., Ma, S. Y., Zang, Y. F., Milham, M. P., & Zhu, C. Z. (2010).
Subject order-independent group ICA (SOI-GICA) for functional MRI data analysis.

NeuroImage, 51(4), 1414–1424. [http://doi.org/S1053-8119\(10\)00326-5](http://doi.org/S1053-8119(10)00326-5)

[pii]10.1016/j.neuroimage.2010.03.039

Zhao, X., Glahn, D., Tan, L. H., Li, N., Xiong, J., & Gao, J. H. (2004). Comparison of TCA and ICA Techniques in fMRI Data Processing. *Journal of Magnetic Resonance Imaging*, 19(4), 397–402. <http://doi.org/10.1002/jmri.20023>

Zhou, Y., Fan, L., Qiu, C., & Jiang, T. (2015). Prefrontal cortex and the dysconnectivity hypothesis of schizophrenia. *Neuroscience Bulletin*, 31(2), 207–219. <http://doi.org/10.1007/s12264-014-1502-8>

Zuo, X.-N. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., & Milham, M. P. (2010). Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. *NeuroImage*, 49(3), 2163–2177. [http://doi.org/S1053-8119\(09\)01152-5](http://doi.org/S1053-8119(09)01152-5) [pii]10.1016/j.neuroimage.2009.10.080

Zuo, X.-N., & Xing, X.-X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*, 45, 100–118. <http://doi.org/10.1016/j.neubiorev.2014.05.009>

Appendix 1. Supplement to Study 1

Method

Subjects. Data were collected as part of the Cognitive Neuroscience Test Reliability and Clinical applications for Schizophrenia consortium (CNTRaCS), which is made up of five sites: University of California at Davis, Maryland Psychiatric Research Center at the University of Maryland (MPRC), Rutgers University, University of Minnesota, and Washington University in St. Louis. Informed consent was obtained from all participants at each site, and the study was approved by the institutional review board of each respective site. Diagnoses (or lack thereof) were confirmed using the Structured Clinical Interview for DSM-IV, patient and non-patient editions (First, Spitzer, Gibbon, & Williams, 2002; First, Spitzer, Gibbon, & Williams, 2002). Groups were matched on sex, age, handedness, SES based on the Hollingshead Index (Hollingshead & Redlich, 1958), and estimated premorbid intelligence (using Wechsler Test of Adult Reading; WTAR (Wechsler, 2001)).

Data collection was attempted on 120 subjects (60 HCs, 60 SZs). Of these 120 subjects, 2 were excluded due to excessive movement (average relative movement > 0.37 mm; 1 HC, 1 SZ), 6 were excluded due to minimal task competence (2 HCs, 4 SZs) using the following criteria: error rates greater than 90% on A-X, A-Y, or B-X trials or error rate greater than 50% on B-Y trials. Four subjects were excluded due to inability to understand the task (4 SZs), 3 were excluded due to poor image quality (3 SZs), and 2 were excluded due to poor effort or refusal to continue (1 HC, 1 SZ).

Dot pattern expectancy task and analysis. Each trial in the DPX consisted of a cue followed by a probe, with each cue and each probe lasting 500 milliseconds. Trials were jittered such that inter-stimulus intervals ranged from 2.5 seconds to 3.5 seconds and inter-trial intervals ranged from 2.5 seconds to 12.5 seconds.

The fact that the majority of trials were AX trials was designed to induce a prepotent response to an A cue, namely that the subject will expect an X probe to follow. In the rare event that an A cue is followed by a Y probe, the subject must overcome the prepotent response and respond to that trial as a non-target trial. It is expected that people with intact goal maintenance have relatively more difficulty with AY trials for this reason. On the other hand, BX trials should prove more difficult for people with impaired goal maintenance, because they fail to maintain the non-target context throughout the trial, responding target to the X probe. BY trials acted as a validity check, as subjects who make many errors in this condition likely did not understand the task or were not paying attention.

fMRI data acquisition and preprocessing. The scanning session consisted of a localizer scan to define the anterior commissure-posterior commissure (AC-PC) line for subsequent scans, a 3D T1-weighted anatomical scan (1 mm³ isotropic voxels), a 2D T2-weighted scan, a field map, and the functional EPI sequences. The functional scans were T2*-weighted gradient echo EPI sequences, with TR = 2000 ms, TE = 30 ms, flip angle = 77 degrees, 32 contiguous AC-PC aligned axial slices, voxel size = 3.5 x 3.5 x 4.0 mm.

Confirmatory ROI GLM analysis. A previous fMRI study of the expectancy AX task in schizophrenia patients (MacDonald, Carter, et al., 2005) found significant group differences in activation for B Cue versus A Cue trials in three brain regions: right and left MFG and right IFG. The coordinates and volumes reported in that study were used to construct spherical regions of interest (ROIs) for small volume corrections. Activation data were extracted from these regions separately for each subject and included in an independent samples *t*-test, with the expectation that HCs would exhibit significantly greater activation relative to SZs in B cues compared to A cues.

Special considerations for multi-site analyses. To assess and mitigate site differences in image quality across sites, several procedures were employed. First, phantom data were collected during the scanning session for every subject at every site. Differences in scanner performance across sites and over time were thus able to be assessed easily. Next, metrics of scan quality were measured and compared across sites, including signal to fluctuating noise ratio, estimated spatial smoothness, and absolute and relative movement. A common threshold was set and extreme data were removed from further analysis. Sites were directly compared on these QA metrics, and these (as well as site itself) were also included in the group-level GLM contrasts as covariates of non-interest. In this way, effects of interest that were driven by site effects may be assessed and partially controlled.

Results

DPX task speed/accuracy trade-off. To determine if SZs' poorer accuracy was due to a tradeoff between speed and accuracy, response times were compared overall and on specific trial types. In all cases, SZs had slower reaction times than HCs (overall: $t(80.69) = 3.30, p = .001$; AX: $t(86.20) = 3.24, p = .002$; AY: $t(78.70) = 2.41, p = .02$; BX: $t(80.65) = 3.30, p = .001$; BY: $t(75.75) = 3.51, p < .001$), but neither group demonstrated significantly slower times on correct versus incorrect trials (all $ps > .28$). In fact, SZs were (non-significantly) slower on incorrect trials than correct ones. An ANOVA model with reaction time as the DV and group, trial type, and accuracy as predictor variables found significant main effects for all predictor variables (accuracy: $F(1, 697) = 3.97, p = .047$; trial type: $F(3, 697) = 7.83, p < .001$; group: $F(1, 697) = 45.77, p < .001$) but failed to show significant interactions between any predictor variables with the exception of accuracy and trial type ($p < .001$). These results suggest SZs were not prioritizing speed over accuracy more than HCs were.

Confirmatory ROI fMRI results. Confirmatory ROI analyses consisted of extracting data from voxels within *a priori* ROI masks and comparing groups using Welch's independent samples *t*-tests in the contrast of B cues versus A cues. In the right MFG mask, HCs displayed significantly more activation (B > A cues) compared with SZs, $t(90.02) = 2.47, p = .016, Cohen's d = 0.52$. Likewise, in the left MFG mask, HCs showed significantly more B cue activation compared with SZs, $t(96.85) = 2.33, p = 0.022, Cohen's d = 0.47$. Groups did not differ significantly in the right IFG mask, $t(97.82) = -1.71, p = 0.09, Cohen's d = 0.35$. These data are presented in supplementary

figure 1.4. There were no significant correlations between any ROI activation and either BPRS symptoms or task performance (all p 's > .05).

To assess the effect of site in these analyses, all of the ANOVAs predicting BOLD activation within each mask used both subject group and CNTRaCS site. In no case was there a significant main effect of site (p 's > .06) or a site by group interaction (p 's > .18). The main effect of group status remained significant in left MFG ($F(1, 93) = 4.98, p = .03$) whereas the group difference was attenuated in right MFG ($F(1, 93) = 2.52, p = .12$) and remained non-significant in right IFG ($F(1, 93) = 0.18, p = .67$).

Supplementary Table 1.1. Clusters of Significant Activation in fMRI Analysis.

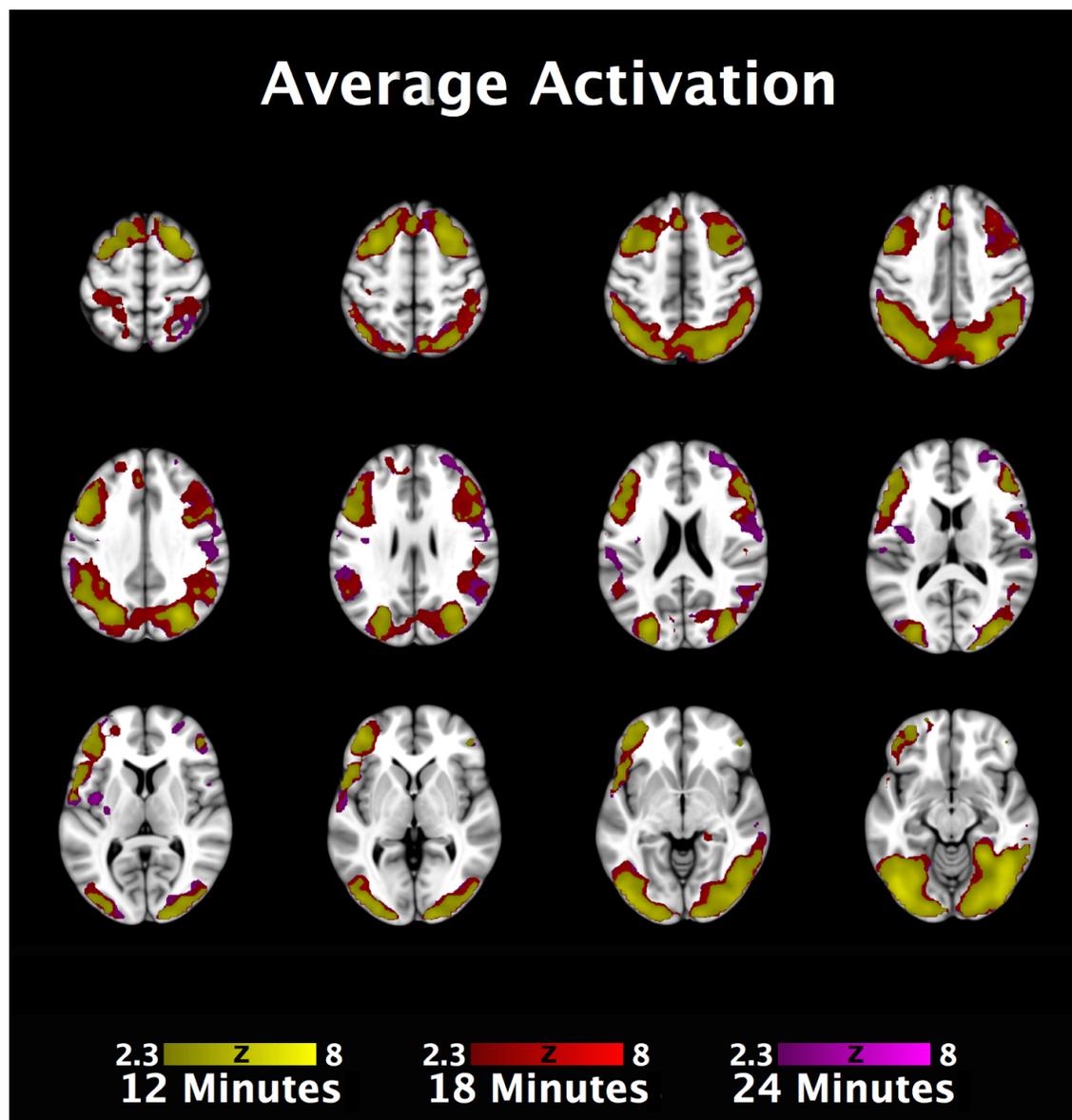
Contrast	Cluster	Peak Voxel Z Score	Cluster Volume (mm ³)	Anatomical Regions	MNI Coordinates		
					X	Y	Z
SZ > 0	1	4.39	10,152	L Middle Frontal Gyrus	-50	12	42
				L Inf Frontal Gyrus pars opercularis	-48	12	20
				L Sup Frontal Gyrus	-28	2	66
				L Inf Frontal Gyrus pars triangularis	-50	32	16
				2	6.09	39,408	L Lateral Inf Occipital Cortex
	L Occipital Pole	-24	-98	-10			
	L Lateral Sup Occipital Cortex	-30	-84	22			
	L Temporal Occipital Fusiform Cortex	-44	-54	-16			
	L Pos Temporal Fusiform Cortex	-34	-38	-24			
	L Pos Supramarginal Gyrus	-48	-48	44			
	L Angular Gyrus	-48	-54	38			
	L Inf Temporal Gyrus temporooccipital part	-50	-60	-20			
	L Occipital Fusiform Gyrus	-32	-70	-12			
	L Precuneous Cortex	-12	-72	40			
	3	5.91	48,856	R Lateral Inf Occipital Cortex	38	-76	-16
	R Lateral Sup Occipital Cortex	24	-74	40			

				R Occipital Fusiform Gyrus	30	-84	-14
				R Temporal Occipital Fusiform Cortex	44	-52	-16
				R Occipital Pole	26	-98	8
				R Angular Gyrus	40	-50	36
				R Pos Supramarginal Gyrus	52	-42	46
HC > 0	1	6.78	495,912	R Lateral Inf Occipital Cortex	46	-66	-14
				R Occipital Fusiform Gyrus	42	-68	-16
				L Lateral Inf Occipital Cortex	-38	-80	-12
				L Occipital Fusiform Gyrus	-34	-82	-12
				R Lateral Sup Occipital Cortex	28	-74	36
				L Pos Supramarginal Gyrus	-46	-48	50
				R Temporal Occipital Fusiform Cortex	30	-42	-24
				L Temporal Occipital Fusiform Cortex	-40	-60	-14
				R Middle Frontal Gyrus	32	16	54
				R Superior Frontal Gyrus	24	20	54
				L Lateral Sup Occipital Cortex	-28	-84	18
				L Frontal Pole	-42	38	22
				R Frontal Pole	28	54	18
				L Central Opercular Cortex	-40	-4	12
				L Middle Frontal Gyrus	-32	14	50
				R Pos Supramarginal Gyrus	52	-44	48
HC > SZ	1	4.28	11,056	R Frontal Pole	26	54	14
				R Paracingulate Gyrus	12	50	24
				R Middle Frontal Gyrus	24	28	36
				R Ant Cingulate Gyrus	10	32	16

2	4.31	23,408	L Parietal Operculum Cortex	-50	-40	22
			L Pos Supramarginal Gyrus	-62	-46	24
			L Pos Superior Temporal Gyrus	-62	-22	2
			L Postcentral Gyrus	-62	-14	24
			L Lateral Inf Occipital Cortex	-56	-64	-2
			L Supracalcarine Cortex	-22	-58	14
			L Lateral Sup Occipital Cortex	-38	-70	18
			L Ant Supramarginal Gyrus	-60	-32	32
			L Inf Temporal Gyrus temporooccipital part	-46	-52	-4
			L Planum Temporale	-40	-34	6
3	4.43	30,664	L Central Opercular Cortex	-40	-4	12
			L Frontal Pole	-38	40	24
			L Sup Frontal Gyrus	-14	0	62
			L Supplementary Motor Cortex	-16	-14	42
			Ant Cingulate Gyrus	0	8	24
			L Pos Cingulate Gyrus	-4	-20	34
			L Sup Pos Temporal Gyrus	-52	-12	-8
			R Ant Cingulate Gyrus	2	-8	32
			L Precentral Gyrus	-6	-16	54
L Planum Polare	-52	2	-6			

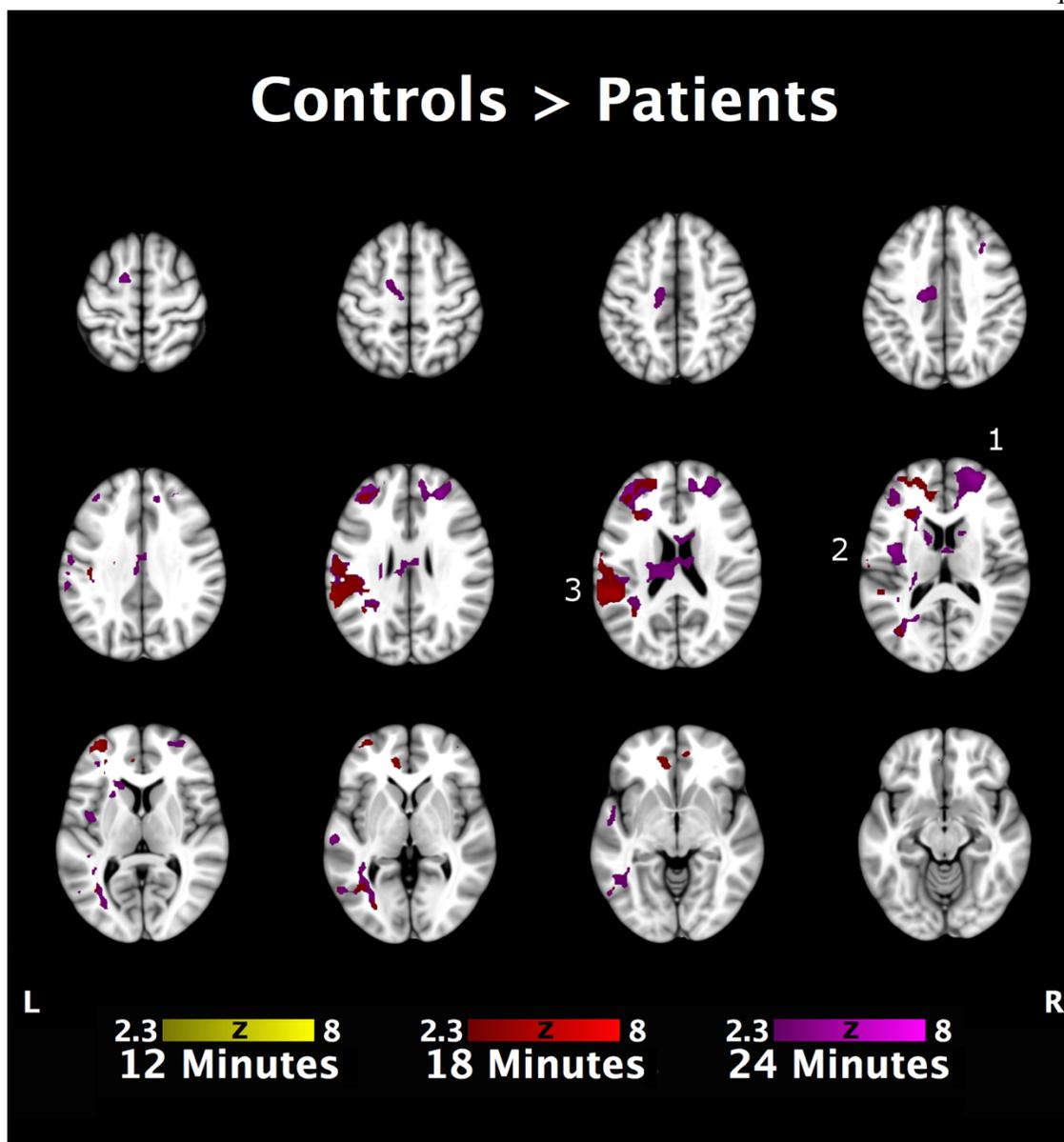
Note: "SZ" refers to schizophrenia patients; "HC" refers to healthy controls. "L" and "R" refer to "left" and "right" hemisphere, respectively. "Ant" refers to "anterior," "Pos" refers to "posterior," "Inf" refers to "inferior," and "Sup" refers to "superior". "MNI" refers to the Montreal Neurological Institute.

Supplementary Figure 1.1. fMRI Efficiency Analysis.



This figure shows average significant activation irrespective of group status on the B Cues – A Cues contrast for scan lengths of 12, 18, and 24 minutes, highlighting the ability to observe similar activation at 12 minutes of data that is observed using 24 minutes.

Supplementary Figure 1.2. fMRI Efficiency Analysis, B Cues – A Cues.

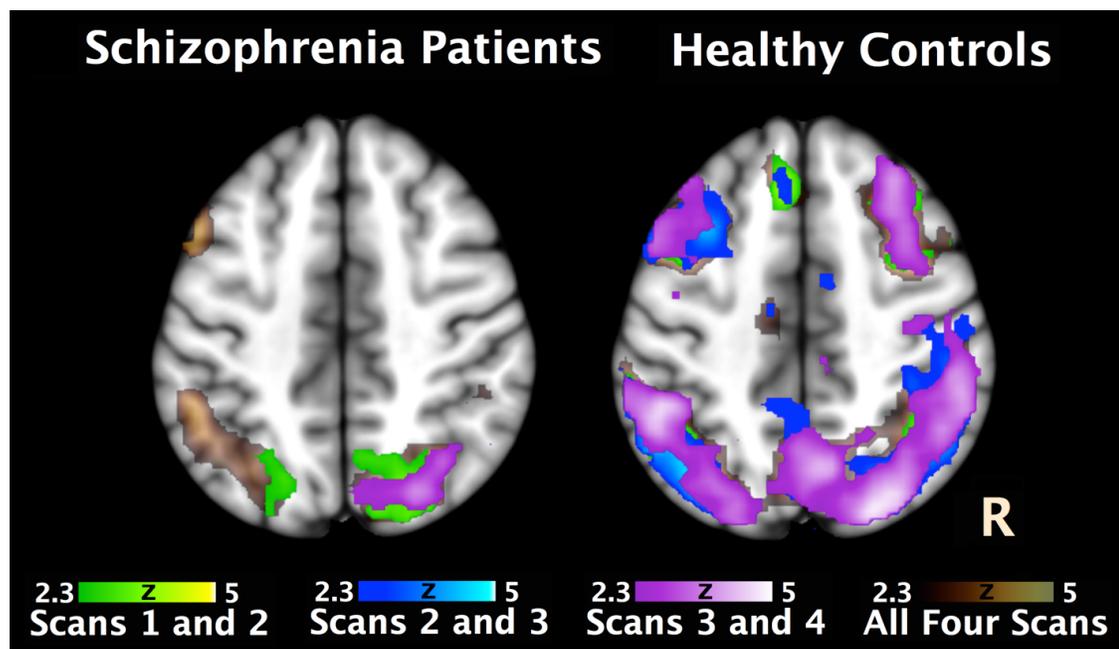


Regions with significantly greater activation in HCs than SZs, made up of 3 clusters.

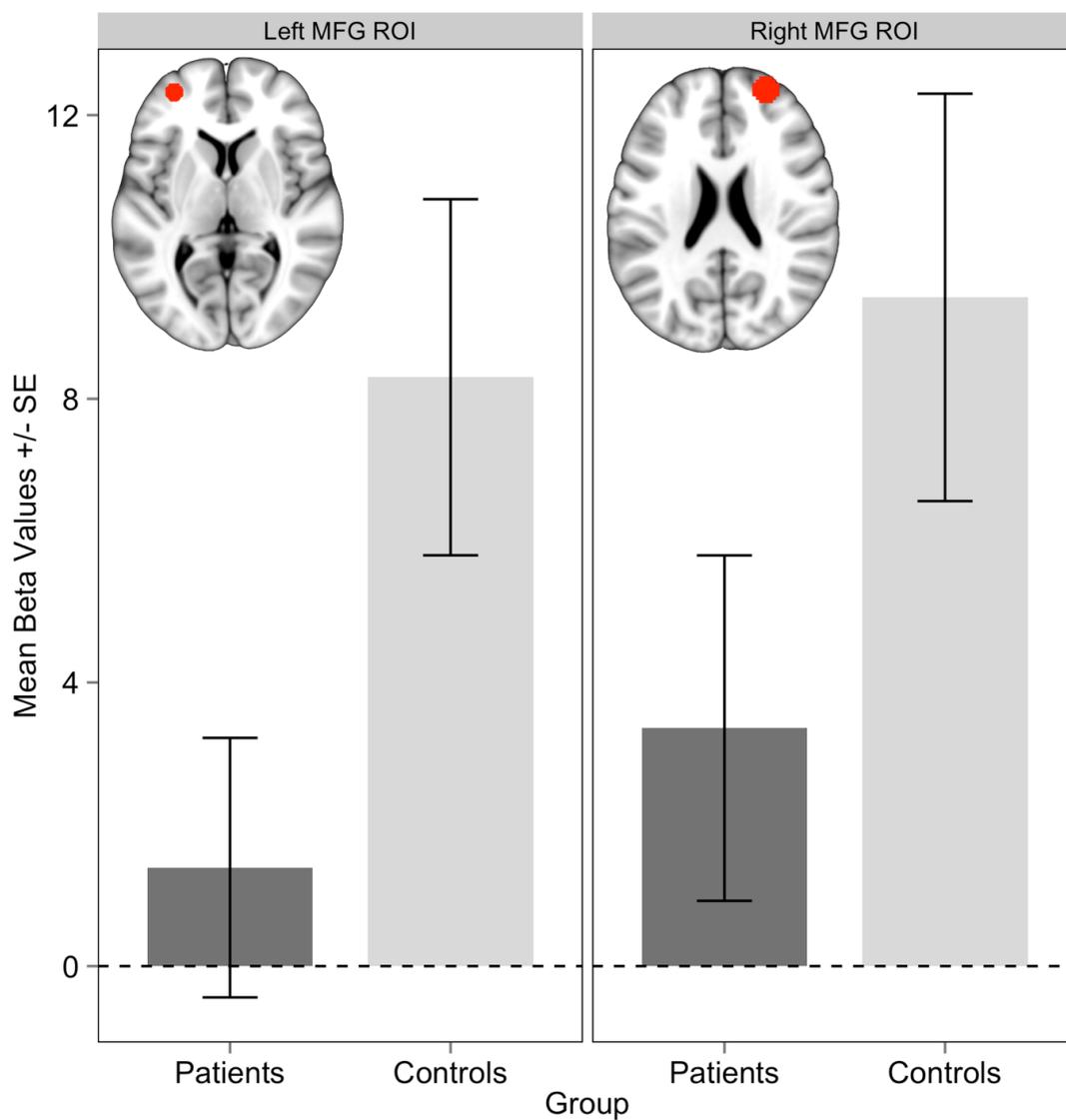
Cluster 1 had a peak voxel Z score of 4.28, a volume of 11,056 mm³, and MNI coordinates (x,y,z) of 26, 54, 14. Cluster 2 had a peak voxel Z score of 4.43, a volume of 30,664 mm³, and coordinates of -40, 4, 12. Cluster 3 had a peak voxel Z score of 4.31, a volume of 23,408 mm³, and coordinates of -50, -40, 22. The scan lengths that define the

color of activation are cumulative, such that the 18 minute data includes the 12 minute data, and the 24 minute data includes the previous two.

Supplementary Figure 1.3. fMRI Efficiency Analysis.



This figure shows significant activation in the B Cues – A Cues contrast using a sliding window approach to show that the efficiency results described previously are not due to changes in the scans over time but rather due to the amount of data used. The green activation represents an analysis using the first two scans (12 minutes) of data, the blue represents the second two scans (12 minutes), and the purple represents the last two scans (12 minutes). The bronze shadow represents the same contrast using all four scans (24 minutes).

Supplementary Figure 1.4. Confirmatory fMRI GLM results.

Beta values are taken from the B Cues – A Cues contrast. Brain images show the size and location of ROIs derived from previous research (A. W. MacDonald, Carter, et al., 2005).

Appendix 2. Supplement to Study 2

Method

Subjects. Data were collected on a total of 120 subjects (60 HC, 60 SZ), of which we excluded two subjects for excessive movement (relative movement $> 0.37\text{mm}$; 1 HC, 1SZ), six subjects for sub-threshold task performance using previously reported criteria (Henderson et al., 2012; 2 HC, 4 SZ), four subjects for inability to understand task instructions (4 SZ), three subjects for poor imaging data quality (3 SZ), and two subjects for poor effort/refusal to continue (1 HC, 1SZ). Additionally, three subjects were excluded because they had fewer than four usable imaging scans (2 HC, 1 SZ). The final sample consisted of 100 subjects (54 HC, 46 SZ). Included HC did not differ from excluded HC on demographic variables (all $p > .16$), nor did included SZ differ from excluded SZ on these or symptom measures (all $p > .11$).

fMRI data acquisition and preprocessing. Scanning protocol included a localized to align scans with AC-PC line, a 3D T1-weighted anatomical scan (1 mm^3 isotropic), a T2-weighted scan, a field map, and T2*-weighted whole brain gradient echo EPI functional scans. These functional scans had the following parameters: TR = 2000 ms, TE = 30 ms, flip angle = 77° , 32 AC-PC aligned axial slices, voxel size = $3.43 \times 3.43 \times 4.0\text{ mm}$. Four scans of 180 functional volumes were collected during four blocks of the DPX task. Quality control “phantom” scans were also collected on each scanner at the time of each subject’s data collection.

Preprocessing using FMRIB Software Library (FSL v. 4.1.8; Smith et al., 2004) included motion correction using MCFLIRT (Jenkinson et al., 2002), brain extraction

using BET (Smith, 2002), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma=50.0s$); B_0 field unwarping, spatial smoothing with a 7 mm FWHM Gaussian kernel, and spatial normalization and linear registration to the MNI 152 standard brain using FLIRT (Jenkinson & Smith, 2001). Motion regression was then performed by conducting a multiple regression whereby the six motion parameters calculated for each subject's data during motion correction were used to predict the functional data. The residuals from this regression were retained and used for subsequent analyses of the functional data. Data for each of the four, 6-minute blocks of the DPX task were then concatenated temporally, producing one functional data set for each subject.

Tonic functional network connectivity. We also sought to ascertain whether subject groups differed with regard to the amount components lagged behind one another. To do this, the average amount of lag was calculated for each component pair within each group. Next, for each component pair we subtracted the SZ mean value from the HC mean value. To determine if one group had more lag than the other, a sign test (Zar, 2010) was performed on this array of difference values.

Results

Tonic functional network connectivity.

To determine if groups differed with regard to the amount of lag between component pairs, a binomial test was conducted. In 359 out of 595 cases, SZ had larger

lag values compared with HC. A binomial test comparing this rate to 0.5 resulted in a 95% confidence interval of 0.56 to 0.64 and a p value less than .001, rejecting the null hypothesis that the probability of SZ having a larger lag than HC was 0.5.

Dynamic functional network connectivity. Because of the significant correlations between tFNC and d' context in SZ and between dFNC/task timeline with d' context in HC, it suggested that the association between right ECN and SN may act as a prerequisite for the two ECN to coordinate to meet task demands. To test this, SZ whose tFNC values between right ECN and SN were greater than one standard deviation below the mean for SZ were removed from the analysis. As displayed in supplementary figure 2.4, the correlation between the task modulation of the two ECN with d' context was repeated, and SZ then showed an equivalent relationship with task performance as did HC.

Supplementary Table 2.1. Sample Demographics.

	Group		Test
	SZ	HC	
N	46	54	
Mean Age (yrs)	35.8 (12.1)	34.9 (11.9)	$t(98) = 0.38$
% Male	73.9	75.9	$\chi^2(1) = 0.00$
% Caucasian	56.5	61.1	$\chi^2(1) = 0.07$
% Right-Handed	84.8	85.2	$\chi^2(1) = 0.00$
Mean Education (yrs)	14.0 (2.0)	15.3 (2.6)	$t(98) = -2.93^*$
Mean Parental Education (yrs)	13.9 (2.5)	13.8 (2.7)	$t(89) = 0.09$
BPRS Total	40.3 (10.3)	n/a	
Positive Symptoms ^a	9.4 (5.2)	n/a	
Negative Symptoms ^b	7.2 (2.3)	n/a	
Disorganization ^c	5.0 (1.7)	n/a	

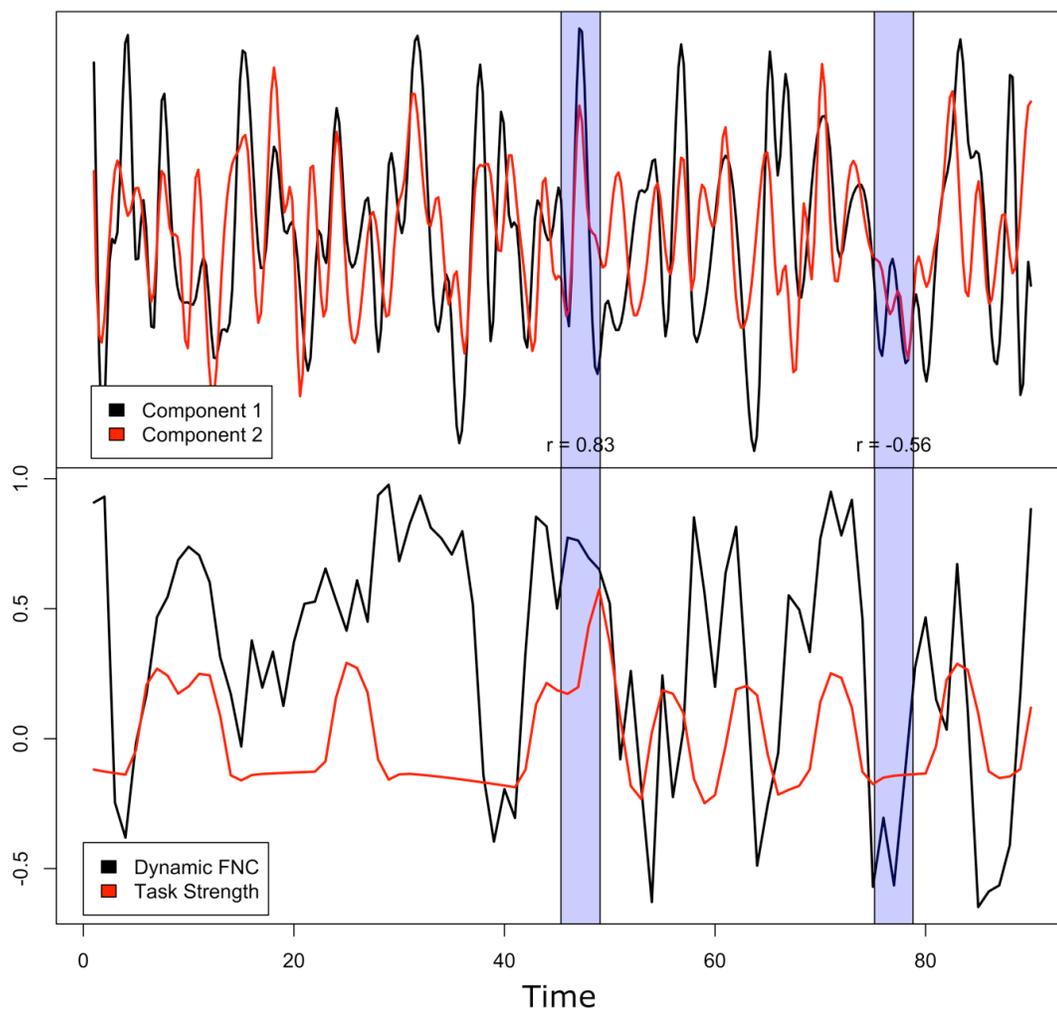
Note: BPRS, Brief Psychiatric Rating Scale; SZ, schizophrenia patients; HC, healthy controls. Parenthetical numbers following means represent standard deviations. Asterisks following test statistics represent $p < .05$.

^a BPRS items 8, 9, 10, and 11.

^b BPRS items 13, 16, 17, and 18.

^c BPRS items 12, 14, 15, and 24.

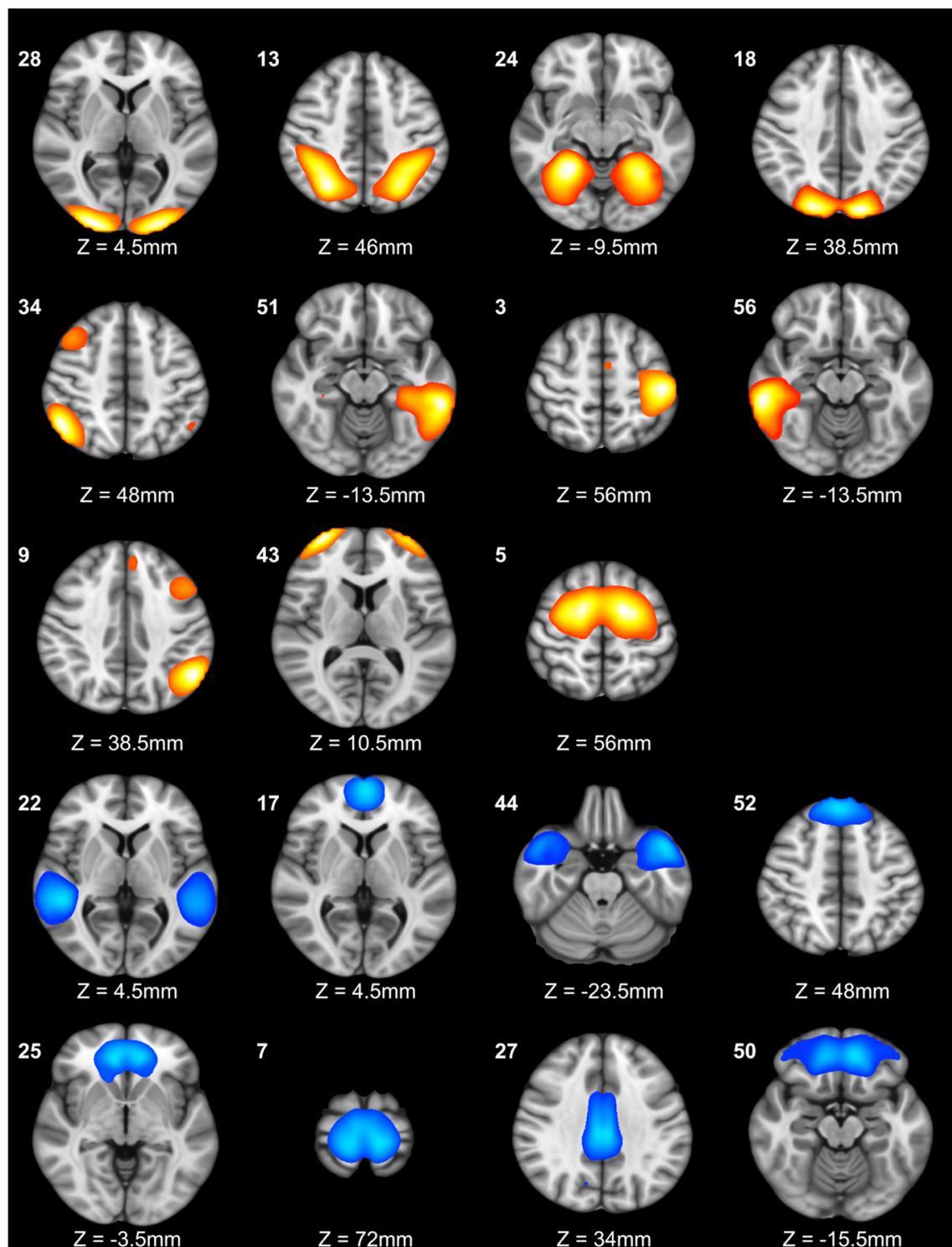
Supplementary Figure 2.1. Demonstration of dynamic functional network connectivity.



The top portion shows the time courses associated with two components for a single subject. The two blue rectangles represent two example time windows within which correlation coefficients are measured. The bottom portion shows the dFNC between those two components and also the task demands. The bottom black line represents dFNC and shows that when the two components' timecourses are more similar, dFNC is higher. The bottom red line represents the task time course of B cues after having been convolved

with a hemodynamic response function and after a moving average has been applied. A correlation coefficient is calculated between the task demands and the dFNC to assess task modulation of the dFNC between these two components.

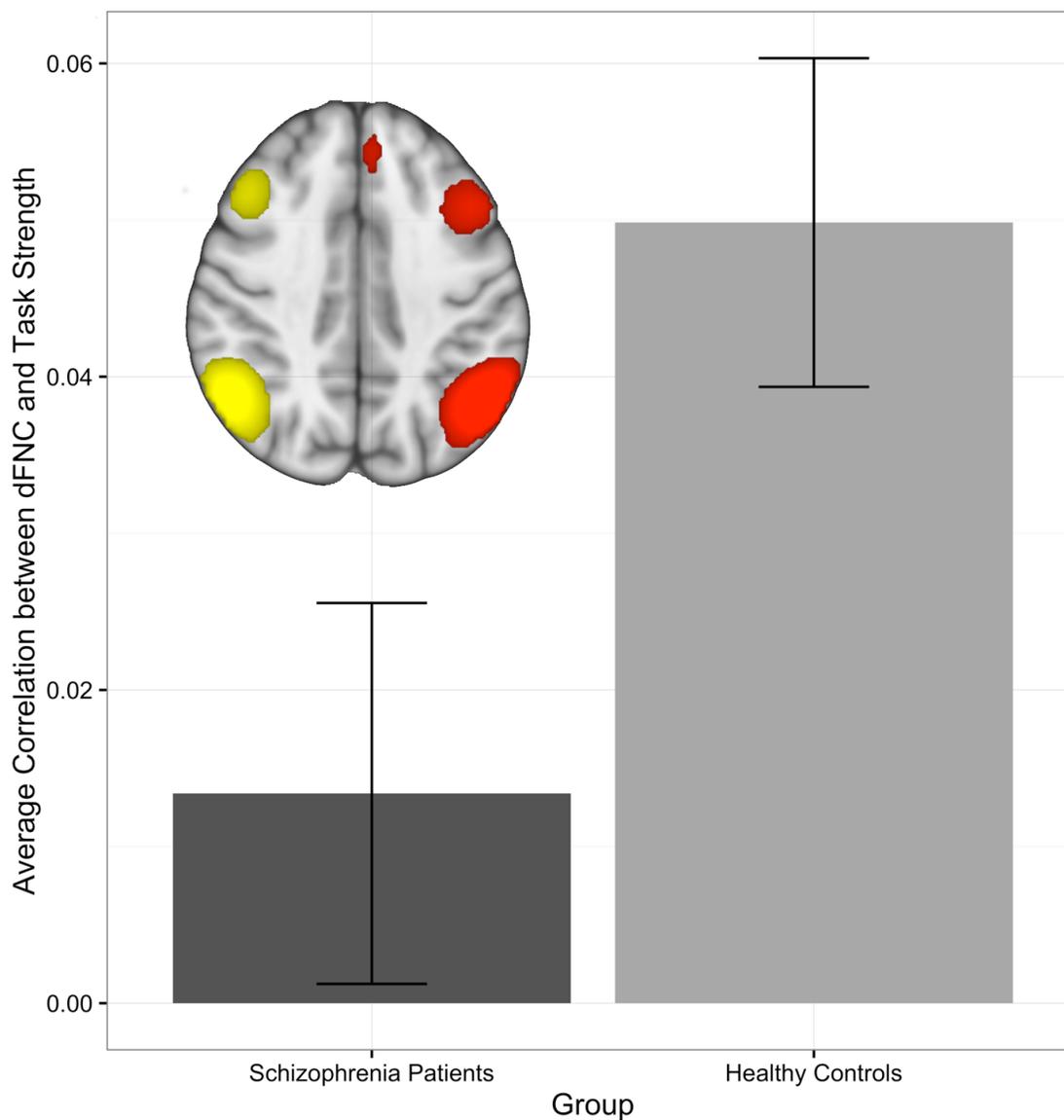
Supplementary Figure 2.2. Task-related component maps.



These nineteen components' time courses were significantly related to the DPX task's B

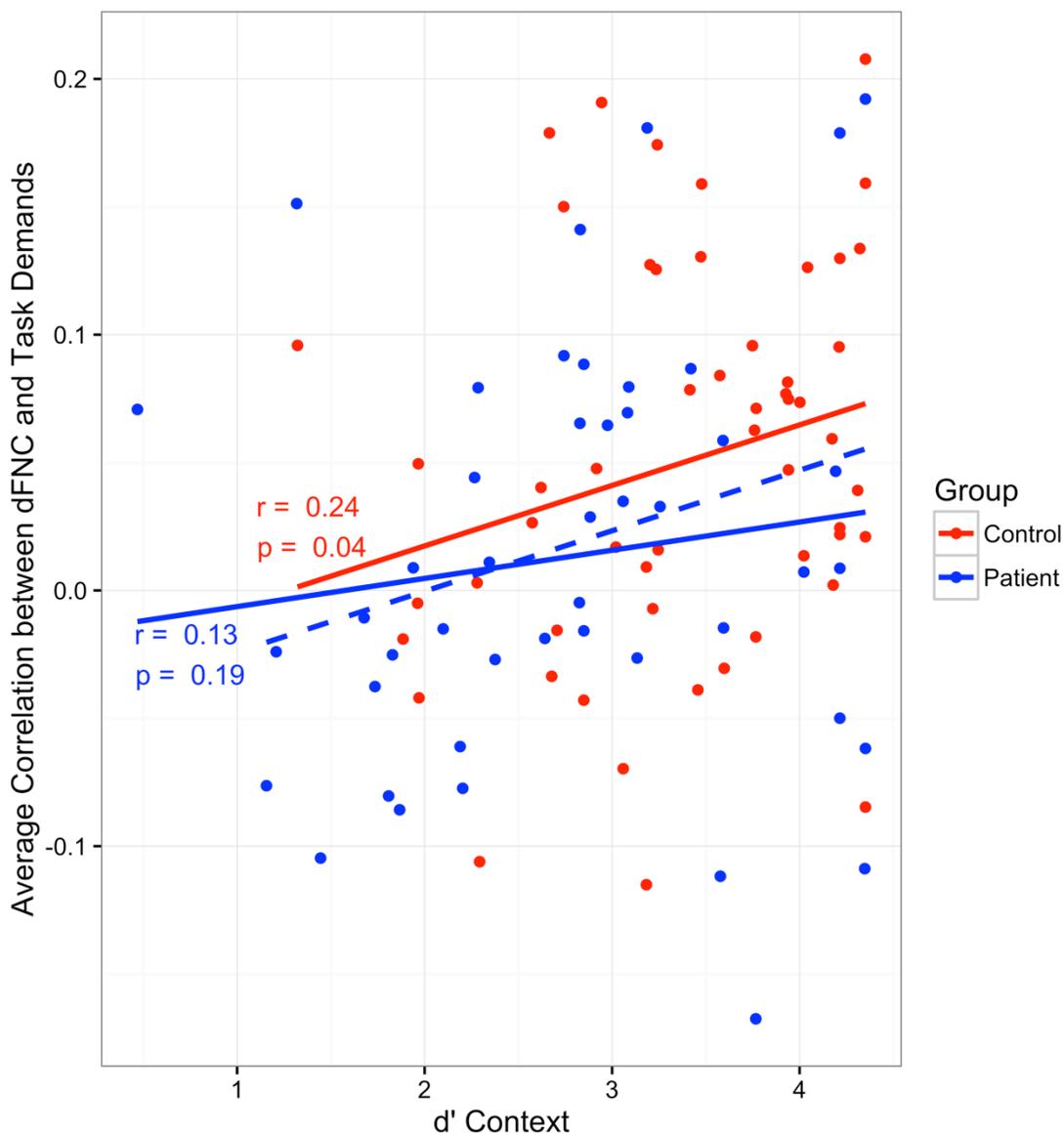
Cue time course across groups. Numbers to the upper left of each map represents the arbitrary index of that component. Warm colors representing component spatial maps indicate components that were positively correlated with the task time course, and blue colors indicate negative correlations. Within each type (positive and negative), components are ordered by the extremity of the correlation value, with the most extreme values at the top left. Maps are presented in radiological view, so images are reversed left to right.

Supplementary Figure 2.3. Task modulation of dynamic functional network connectivity between ECN components.



This figure shows the extent to which dFNC, measured between the two executive components, correlated with the DPX task's B Cue time course for each subject group. The brain image shows the two frontoparietal component spatial maps. This map is in radiological view, so left and right are reversed.

Supplementary Figure 2.4. DPX Performance Predicted by dFNC in Schizophrenia Patients.



This figure shows a dashed blue line that represents the correlation between the task modulation of the dynamic relationship between left and right ECNs with d' Context in

schizophrenia patients after removing those patients with tFNC strength between ECN and SN greater than one standard deviation below the mean.