DIAGNOSIS- AND FAMILY MEMBERSHIP-DEPENDENT COGNITIVE DOMAIN IMPAIRMENT IN SCHIZOPHRENIA PATIENTS AND THEIR UNAFFECTED SIBLINGS

by

KATIE M. LAVIGNE

B.A., Concordia University, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Neuroscience)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2013

Abstract

The goals of the present research were two-fold: (1) to examine whether diagnosis-dependent group differences in cognitive performance among schizophrenia patients, their unaffected siblings and healthy controls are fundamentally the result of a general cognitive impairment and/or of domain-specific deficits in schizophrenia; and (2) to examine the cognitive domains that characterize family membership-dependent and family membership-independent group differences in cognitive performance between schizophrenia patients and their siblings. In Study 1, results from a traditional statistical analysis method suggested impairment in all five cognitive domains tested, whereas constrained principal component analysis (CPCA) revealed a single cognitive domain accounting for group differences that extended across all five traditional domains. This component reflected impairment in a generalized cognitive domain in schizophrenia patients and, to a lesser degree, siblings, and was dominated by WAIS-R Digit Symbol and WMS-R Logical Memory subscales, a finding in line with literature reporting most severe impairment in information processing speed and verbal memory in schizophrenia. In Study 2, CPCA with hierarchical regression was used to examine the cognitive domains that accounted for the interaction between group and family membership, revealing three cognitive domains (Working Memory/Attention, Visual Memory, and Verbal Memory) where differences between patients and their siblings depended on family membership. A subsequent cluster analysis revealed several family clusters differing on patients' and siblings' performance across these three cognitive domains. The results of the current research suggest that (1) diagnosis-based group differences in cognitive performance are due to impairment in a generalized cognitive domain (and not primarily within more specific cognitive domains) that is common to all families, (2) this general impairment is best captured by measures of information processing speed and verbal memory, and that (3) family membership-dependent group differences are present in more specific cognitive domains that are distinguishable from the general domain describing overall group differences. This research helps synthesize the two sides of the debate surrounding the nature of cognitive impairment in schizophrenia, by suggesting that there is impairment in both a generalized cognitive domain and in more specific domains, but that the latter may depend on moderating factors, such as family membership.

Preface

This research was approved by the University of British Columbia Behavioural Research Ethics Board, Certificate Number: H12-03133.

Table of Contents

List of Tables

List of Figures

Introduction

Cognitive impairment is a predominant feature of schizophrenia, and can be observed in a wide range of domains, including memory, attention, and executive functioning (Aleman, Hijman, de Haan, & Kahn, 1999; Heinrichs & Zakzanis, 1998). Deficits in basic cognitive processes such as these translate into difficulties completing common daily tasks and can impact social functioning, psychosocial skills and problem-solving abilities (Green, Kern, Braff, & Mintz, 2000). The importance of further understanding cognitive impairment in schizophrenia has regained some momentum in recent years with the emergence of cognitive interventions as complementary treatments to conventional medications (e.g., cognitive remediation therapy; McGurk, Twamley, Sitzer, McHugo, & Mueser, 2007; Wykes, Huddy, Cellard, McGurk, & Czobor, 2011).

General Cognitive Impairment vs. Fundamental Deficits in Specific Cognitive Domains

Given the pervasiveness of cognitive impairment in schizophrenia, it can be difficult to identify the nature of such deficits. Indeed, there is some debate surrounding this issue, namely, whether impaired performance in certain cognitive domains is a fundamental aspect of schizophrenia, a secondary effect of dysfunction in a more basic cognitive process(es), or a combination of the two (Dickinson, Iannone, Wilk, & Gold, 2004; Gold & Dickinson, 2013; Green, Horan, & Sugar, 2013). A seminal meta-analysis conducted by Heinrichs & Zakzanis (1998) demonstrated that schizophrenia patients are impaired on measures of global functioning as well as in more specific domains assessed by standard clinical tests (e.g., selective verbal and non-verbal memory, motor skills, attention, spatial ability, executive functioning, and language). This finding is well-replicated in the literature, with many studies reporting poorer performance

in schizophrenia patients relative to healthy controls on tests of general cognitive ability (e.g., intelligence quotient; IQ) as well as on tests designed to measure more specific cognitive sub-processes, including memory, attention, information processing speed, and executive functioning (Dickinson, Ramsey, & Gold, 2007; Fioravanti, Carlone, Vitale, Cinti, & Clare, 2005; Gladsjo et al., 2004). Findings such as these led to a debate regarding whether schizophrenia is a disorder characterized by deficits in all of these domains, or whether a more fundamental impairment in general cognitive ability negatively affects performance in these more specific cognitive sub-processes. Given these widespread deficits in combination with impairment on tests of general cognitive ability, many researchers hypothesized that a general cognitive deficit was largely responsible for poor cognitive performance in schizophrenia patients. In line with this idea, another early meta-analysis (Laws, 1999), focusing on executive functioning in schizophrenia, demonstrated that patients' poor performance on the Wisconsin Card Sorting Test (Heaton, 1981) could be accounted for by a general cognitive deficit (i.e., lower IQ scores), challenging the notion that schizophrenia is fundamentally a "disorder of executive functioning".

However, measures of general cognitive ability, such as those assessing IQ, recruit various cognitive sub-processes, and a fundamental impairment in any one of those processes could translate into poor performance on a more generalized test. Research investigating whether schizophrenia patients are impaired in certain cognitive domains over others has supported the notion of domain-specific deficits in schizophrenia, reporting more severe impairment in semantic relative to lexical fluency (Bokat & Goldberg, 2003), and working memory deficits that could not be accounted for simply by differences in IQ (Forbes, Carrick, McIntosh, & Lawrie, 2009). In addition, although schizophrenia patients are generally found to be impaired on most, if not all, cognitive domains commonly tested, they appear to show more severe impairment on

certain measures, namely, those assessing verbal memory (Dickinson, Ragland, Gold, & Gur, 2008; Heinrichs & Zakzanis, 1998) and information processing speed (e.g., digit symbol coding; Dickinson et al., 2007; Knowles, David, & Reichenberg, 2010). Studies of first-episode patients with psychosis have supported these findings, suggesting that cognitive impairment in the domains of verbal memory and information processing speed is not due to the effects of neuroleptic medication, but is a more fundamental characteristic of the disorder (González-Blanch et al., 2011; Mesholam-Gately, Giuliano, Goff, Faraone, & Seidman, 2009; Mohamed, Paulsen, O'Leary, Arndt, & Andreasen, 1999).

Domain-Specific Deficits or Measures of General Cognitive Ability?

Evidence that some cognitive abilities, such as verbal memory and information processing speed, are more impaired than others in schizophrenia suggests that patients may have fundamental deficits in these specific domains. However, one alternative interpretation is that most tests used to assess performance in these domains may be better indices of general cognitive ability than of the specific cognitive sub-processes they are designed to measure (Dickinson, 2008; Dickinson et al., 2007; Gold & Dickinson, 2013; González-Blanch et al., 2011; Green et al., 2013). For example, digit symbol coding tasks require a range of cognitive abilities, including motor and perceptual processing speed, set-shifting (between the test items and code key), relational memory (when memorizing the paired digits and symbols), and visual scanning (when consulting the code key for each test item; Bachman et al., 2010; Davis & Pierson, 2012; Joy, Fein, & Kaplan, 2003). Similarly, common verbal memory tasks generally recruit language comprehension/production and attention-related cognitive processes in addition to memory (Heinrichs & Zakzanis, 1998). In contrast, some of the tasks that show the least impairment in schizophrenia (e.g., Finger Tapping (Dickinson et al., 2007); and Block Design

(Heinrichs & Zakzanis, 1998)) likely recruit a smaller number of more specific cognitive sub-processes and, therefore, may be less sensitive to a general cognitive deficit. As such, it is important to recognize not only the cognitive sub-processes these tests recruit, but also the degree to which they are driven by general cognitive ability.

Data Analysis Techniques

More sophisticated data analysis techniques have recently been employed to address this issue of the nature of cognitive impairment in schizophrenia, including different forms of factor analysis and structural equation modeling (SEM). These techniques aim to uncover the underlying structure of a set of variables through their patterns of intercorrelation, and can generally be classified into two categories: confirmatory and exploratory (Tabachnick & Fidell, 2007). Confirmatory techniques require the specification of a model of hypothesized interactions between the variables of interest (that usually emerge from the literature), after which that model is tested to determine its validity. Exploratory techniques do not require definition of a priori models; rather, the underlying structure of the variables, and the way in which they interact with one another, determines the nature of the results. In general, exploratory models have poorer statistical power than confirmatory models, but are advantageous and are often preferred when the relationship between variables is complex or unknown, or when there are no specific hypotheses about their underlying structure.

Confirmatory Techniques

Confirmatory techniques have been used to examine whether a general cognitive impairment in schizophrenia can account for poor performance in more specific cognitive domains. Using confirmatory factor analysis, Keefe and colleagues (2006) found that cognitive

performance in schizophrenia was best characterized by a single factor that was highly related to the five cognitive domains that reflected the underlying structure of the neurocognitive measures included. Dickinson and colleagues (2008) used SEM to evaluate a 7-factor model comprising executive/working memory, verbal ability, spatial ability, verbal learning & memory, visual learning & memory, and processing speed, which were all hypothesized to relate to a higher-order intelligence factor, *g*. In order to examine the structure of cognitive performance between groups (i.e., schizophrenia patients, unaffected siblings, and controls), the model also included a factor reflecting diagnosis that was related back to the higher-order factor, *g*, as well as to the lower-order cognitive domains. They found that a general cognitive deficit accounted for the majority of between-group differences, with smaller direct effects in information processing speed and verbal memory. These results supported previous work by the same group (Dickinson et al., 2004) and provide further evidence of a generalized cognitive deficit in schizophrenia, and of potential smaller deficits in the domains of verbal memory and information processing speed.

Exploratory Techniques

While SEM and other confirmatory methods are powerful in cases where established models exist, exploratory analysis methods allow for new patterns to emerge. These methods are often preferred in cases where hypotheses about the relationships between variables are complex or previously stated hypotheses are questionable.

Dimension Reduction Followed by Tests of Group Differences

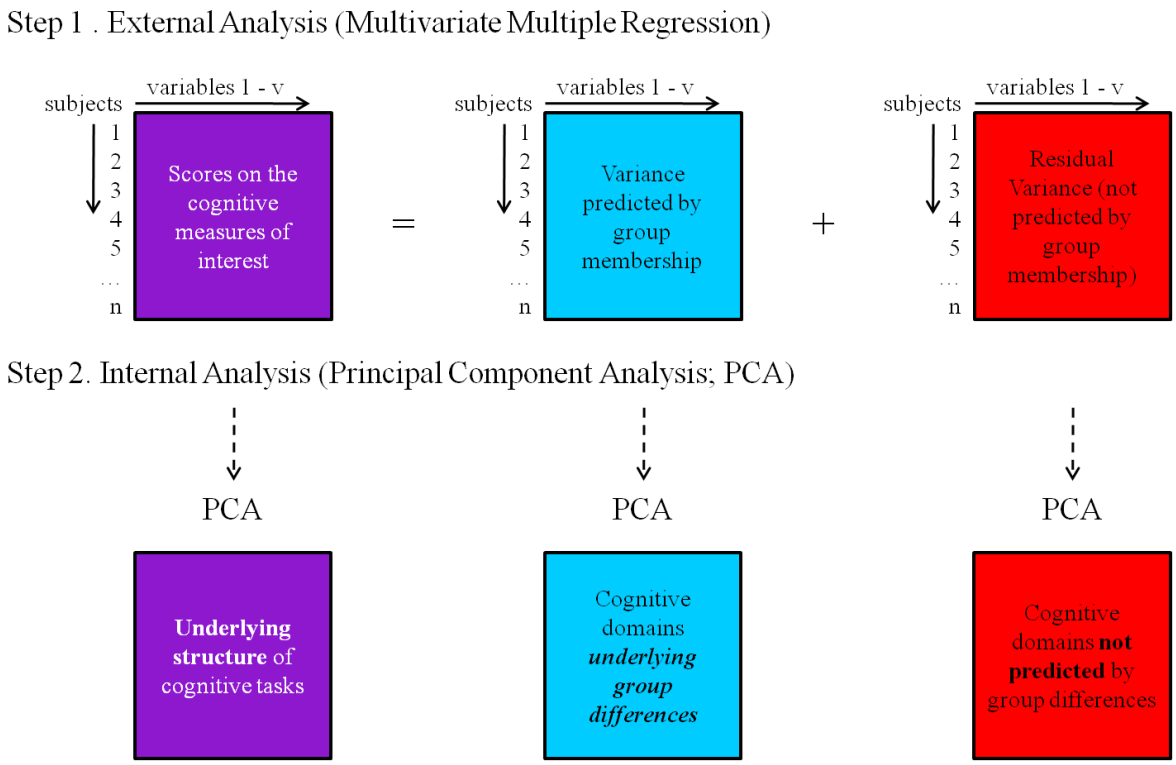
Researchers opting for exploratory methods when examining the underlying structure of cognitive deficits in schizophrenia have generally employed dimension reduction (e.g., principal component analysis; PCA) to identify cognitive domains, followed by statistical tests of group

differences (e.g., t-tests, analysis of variance; ANOVA) to compare performance across groups. Results from these investigations generally find impairment in schizophrenia patients relative to controls on all of the resulting cognitive domains, with a greater degree of impairment in domains reflecting verbal memory and processing speed, similar to findings gathered from other methods (Gladsjo et al., 2004; Nuechterlein et al., 2004). While these studies tend to support the multiple deficit side of the debate regarding the nature of cognitive deficits in schizophrenia, this method of examining whether groups differ on a *pre-defined* set of cognitive domains cannot specifically address the issue of whether cognitive impairment in schizophrenia is fundamentally the result of a general cognitive impairment and/or of domain-specific deficits. This is because the cognitive domains that are compared between groups in these types of studies are identified before group differences are taken into account. As such, these cognitive domains necessarily include variance that is due to factors other than group differences, and that may or may not be relevant to the research question at hand. This method is also the foundation of most early meta-analyses, in which cognitive domains were produced by combining measures based on the cognitive sub-processes they were designed to assess. As described above, research investigating whether schizophrenia patients are impaired on specific cognitive domains almost always finds that is the case. However, in order to optimize examination of cognitive deficits in schizophrenia, it is necessary to focus on only that portion of variance in cognitive test performance that is relevant to group differences, rather than simply determining whether group differences exist in pre-defined cognitive domains that are derived from variance primarily unrelated to group differences.

Constrained Principal Component Analysis

Identifying cognitive domains that characterize cognitive impairment in schizophrenia can be optimized through the use of constrained principal component analysis (CPCA), an exploratory multivariate statistical technique that combines multivariate multiple regression and principal component analysis. In the current study, CPCA allowed for the examination of the cognitive domains underlying group differences in cognitive performance, because it is a method of separating the overall variance in a dataset into that which is predictable by a set of independent variables of interest (e.g., group differences) and that which is not (Hunter & Takane, 1998, 2002; Lavigne, Hofman, Ring, Ryder, & Woodward, 2013; Takane & Hunter, 2001; Takane & Shibayama, 1991). This is done by simply reversing the order of the steps described above (i.e., separating out group differences prior to dimension reduction rather than vice-versa). Thus, CPCA is implemented in two steps, referred to as the external and internal analysis (see Figure 1). The external analysis splits the data into predicted and residual scores via multivariate multiple regression, and the internal analysis employs PCA to determine the component structures underlying the unconstrained, predicted, and residual scores (see the Constrained Principal Component Analysis section in the Appendix for a more in-depth description of this method).

Figure 1. CPCA Procedure with One Set of Dependent Variables and One Set of Independent Variables



Note. Basic CPCA procedure with one set of independent variables (e.g., group membership) and one set of dependent variables (e.g., cognitive performance); v = number of variables; n = number of subjects. The external analysis consists of a multivariate multiple regression, which divides the overall variance into the portion that is predictable by the independent variables (e.g., group membership) and that which is not. The internal analysis consists of three separate principal component analyses (PCA), which identify the underlying component structures of each source of variance (i.e., overall, predicted and residual).

It is important to note that the two exploratory methods described above (i.e., PCA followed by statistical analysis of group differences versus CPCA) are essentially addressing different research questions. Specifically, dimension reduction followed by t-tests or ANOVAs examines whether group differences exist within the cognitive domains that best describe the overall structure of the data, whereas CPCA allows identification of the cognitive domains that best characterize the portion of variance in the overall data that is predictable from group differences (e.g., the portion of variance in cognitive performance that can be predicted by differences between schizophrenia patients, their unaffected siblings, and healthy controls). This is a subtle,

but important, distinction given the debate surrounding the nature of cognitive deficits in schizophrenia. Employing dimension reduction prior to examining group differences within the resulting cognitive domains cannot adequately address the underlying structure of cognitive deficits in schizophrenia precisely because group differences are being investigated within a set of pre-defined domains. As noted in Figure 1 (purple), performing a dimension reduction technique, such as PCA, on the overall data leads to components that reflect the underlying structure of the cognitive tasks included. These resulting cognitive domains may not accurately reflect the cognitive domains that characterize group differences in cognitive performance, as their composition is influenced by factors other than group differences. By constraining the data based on group differences prior to performing dimension reduction (see Figure 1; blue), it is possible to examine *the cognitive domains underlying group differences in cognitive performance*, rather than simply determining whether impairment exists within the cognitive domains that reflect the underlying structure of the cognitive tasks. Those domains that emerge from this constrained (i.e., predictable) variance can provide insight into the nature of cognitive impairment in schizophrenia. Specifically, if cognitive impairment in schizophrenia is best characterized by dysfunction in multiple cognitive domains, these would be expected to emerge as separate components from the portion of variance predicted by group differences. Those cognitive domains not affected in schizophrenia would then emerge in the residual solution as they could not be predicted by group differences. In contrast, if cognitive deficits in schizophrenia are the result of a general impairment in cognitive ability, one would expect a single component to emerge from the predicted solution, with more specific cognitive domains from the overall solution being retained in the residual solution. Finally, if both a generalized cognitive impairment and domain-specific deficits play a role, the predicted solution might show

a dominant first component reflecting a generalized cognitive domain, with other smaller components reflecting those specific cognitive domains which are impaired in schizophrenia (e.g., verbal memory and information processing speed).

Cognitive Impairment in Unaffected Relatives and the Role of Family Membership

Schizophrenia is a highly heritable disorder whose expression is largely driven by genetics (Sullivan, 2005). There is evidence that many neurocognitive deficits characteristic of schizophrenia are also heritable (Cannon et al., 2000; Greenwood et al., 2007; Tuulio-Henriksson et al., 2002), and are present in first-degree relatives of individuals with the disorder (Sitskoorn, Aleman, Ebisch, Appels, & Kahn, 2004).While cognitive deficits in first-degree relatives of schizophrenia patients are generally less severe and/or less widespread than in patients, those cognitive domains that are most reliably impaired tend to be those that are most severely impaired in patients themselves (e.g., verbal memory, processing speed, executive functioning; Dickinson et al., 2007; Glahn et al., 2007; Sitskoorn et al., 2004). That the severity of cognitive impairment in schizophrenia patients and their first-degree relatives corresponds to the genetic proximity of the illness suggests that performance in specific cognitive domains might serve as potential endophenotypic markers of schizophrenia (Glahn et al., 2007; Sitskoorn et al., 2004). Endophenotypes are manifestations of biological processes believed to underlie a genetic predisposition to a disease or disorder, and can be neurophysiological, biochemical, endocrinological, neuroanatomical, neuropsychological, or cognitive in origin (e.g., cognitive impairment in one or more domains in schizophrenia patients and their first-degree relatives; Gottesman & Gould, 2003). Endophenotypes are associated with the disorder in question, heritable, state-independent (i.e., do not vary with stages of the disorder), co-segregate with the disorder within families, and are more prevalent in first-degree relatives of patients with the

disorder than within the normal population (Gottesman & Gould, 2003). Interest in identifying endophenotypes in schizophrenia has gained momentum in recent years given mounting evidence of its likely polygenetic nature and considerable clinical heterogeneity, which has led to difficulty identifying susceptibility genes (Egan, Goldberg, Gscheidle, et al., 2001; Sullivan, 2005). Endophenotypes can be observed in both affected (i.e., probands) and unaffected individuals (e.g., first-degree relatives) within a family at risk for developing the disorder in question, and may, therefore, be particularly useful in identifying susceptibility genes in the case of polygenetic and clinically heterogeneous disorders like schizophrenia. Moreover, the discovery of multiple endophenotypes that characterize a disorder (e.g., impairment in different cognitive domains for different families) has the potential to lead to the identification of subgroups of families demonstrating different patterns of cognitive impairment, which would allow for more in-depth genetic investigations to follow (Glahn et al., 2007).

There is some evidence that patterns of cognitive impairment in schizophrenia may be familial; that is, the cognitive domains on which patients and first-degree relatives are impaired differ between families (Egan, Goldberg, Gscheidle, et al., 2001). Although schizophrenia as a diagnostic category is strongly related to poorer performance relative to controls in many cognitive domains, not all schizophrenia patients demonstrate these deficits (Egan, Goldberg, Gscheidle, et al., 2001; Weickert et al., 2000). Similarly, it is possible that relatives of patients with schizophrenia would also show considerable variability in the domains on which they are cognitively impaired, which would suggest familial impairment in cognitive performance possibly due to distinct genetic profiles across families. One can investigate this hypothesis at a behavioural level by examining subgroup endophenotypes, or, in the case of cognitive ability, by investigating whether diagnosis-based performance on one or more cognitive domains differs

across families. Egan and colleagues (2000; 2001) found that siblings tended to perform poorly on cognitive tasks for which their psychotic sibling also showed impairment, supporting the notion that patterns of cognitive impairment can be familial. Notably, although siblings as a group did not perform significantly worse than control participants on a measure of sustained attention, siblings of patients who were impaired on this measure did demonstrate significantly lower scores than controls (Egan et al., 2000). This finding serves to clarify the inconsistencies in the literature regarding attention-related deficits in siblings of schizophrenia patients (Chkonia, Roinishvili, Herzog, & Brand, 2010; Dickinson et al., 2007; Sitskoorn et al., 2004) and highlights the importance of examining potential subgroups of patients and first-degree relatives who may demonstrate different patterns of cognitive impairment.

Aims

The aims of the present research were two-fold: (1) to examine whether diagnosis-dependent group differences in cognitive performance among schizophrenia patients, their unaffected siblings and healthy controls are fundamentally the result of a general cognitive impairment and/or of domain-specific deficits in schizophrenia; and (2) to examine the cognitive domains that characterize family membership-dependent and family membership-independent group differences in cognitive performance between schizophrenia patients and their unaffected siblings. In Study 1, group differences on cognitive performance among schizophrenia patients, their unaffected siblings, and healthy controls were investigated using two different methods: PCA followed by ANOVA (PCA-ANOVA); and CPCA. In Study 2, we investigated the role of family membership on cognitive impairment in schizophrenia by examining the cognitive domains that accounted for the main effects of, and the interaction between, group (schizophrenia patients versus unaffected siblings) and family membership using CPCA with

hierarchical regression. For a detailed explanation of the mathematical operations underlying the CPCA models employed in studies 1 and 2, see the Mathematical Operations section of the Appendix.

Hypotheses

With regard to the primary aim of the current research, we hypothesized that PCA-ANOVA and CPCA would lead to different conclusions regarding the nature of cognitive deficits in schizophrenia. Specifically, we expected that PCA-ANOVA would show that schizophrenia patients were impaired on all of the cognitive domains that emerged. In contrast, it was hypothesized that CPCA would produce a single cognitive domain reflecting a general cognitive impairment in schizophrenia, and that this domain would show the strongest contributions from indices of information processing speed and verbal memory, as these have been suggested as particularly strong measures of general cognitive ability; however, all variables were expected to show relatively similar contributions on this domain given that it should represent general cognitive ability.

With regard to the second aim of this research, we hypothesized that family membership-dependent group differences (between schizophrenia patients and their siblings) in cognitive performance would reflect more specific cognitive domains. Cognitive domains representing potential endophenotypes in schizophrenia would be expected to show impairment in both patients and, to a lesser degree, siblings of the same families. If multiple cognitive domains represented family-membership dependent group differences in cognitive performance and reflect multiple cognitive endophenotypes, these domains would be also be expected to show impairment in some families, but not others. In addition, we hypothesized that the main effect of

group, which represents diagnosis-dependent group differences common to all families, would be characterized by impairment in a generalized cognitive domain in schizophrenia, as in Study 1.

Methods

Participants

The original sample consisted of 863 schizophrenia patients, 673 of their unaffected siblings, and 1102 healthy controls who completed a battery of neuropsychological tests as part of the National Institute of Health's Clinical Brain Disorders Branch (CBDB) Siblings Study (Dickinson, Goldberg, Gold, Elvevag, & Weinberger, 2010; Egan et al., 2000; Egan, Goldberg, Kolachana, et al., 2001; Genderson et al., 2007). The CBDB Sibling Study is a comprehensive, ongoing investigation into the origin, physiology, genetics and course of schizophrenia, and involves neurological examination, electroencephalography and magnetic resonance imaging of the brain, and blood testing in addition to neurocognitive assessment. All participants were medically screened and interviewed by a board-certified psychiatrist using the Structured Clinical Interview for Axis I (patient and non-patient versions as appropriate) and Axis II disorders (First, Gibbon, Spitzer, & Williams, 1996a, 1996b). Patient diagnoses were confirmed by a second psychiatrist. Exclusion criteria for all groups included the following: head trauma with extended loss of consciousness, IQ < 70, evidence of learning disability, and substance abuse within the last 6 months, substance dependence within the last year, or a history of 5 or more years of substance abuse/dependence. Healthy controls were also excluded if they had a first-degree relative with schizophrenia, were diagnosed with any Axis I or II disorders, or were taking any psychotropic medications at the time of testing. Schizophrenia patients were stable and receiving neuroleptic medications during participation in the CBDB study.

Participants from this overall sample were excluded from the current research if they (1) had missing data on any of the variables of interest (see below), or (2) were part of the sibling group

and were re-classified as being affected with psychosis. This led to a sample of 334 schizophrenia patients, 386 siblings, and 481 healthy controls (N=1201), who were included in Study 1. For Study 2, families with at least one schizophrenia patient and sibling were selected from the above sample of 1201 participants, leading to a subsample of 165 families consisting of a total of 165 schizophrenia patients and 253 unaffected siblings (N=418). Controls were removed from study 2 as they were not part of the same families as patients and siblings. Group means and standard deviations for demographic and cognitive variables are listed in Tables 1 and 2 for studies 1 and 2, respectively.

Measures

Symptom Assessment

Patients' symptom severity was assessed using the Positive and Negative Syndrome Scale (PANSS; Kay, Fiszbein, & Opler, 1987; Kay, Opler, & Lindenmayer, 1989). Each of the PANSS' 30 items is rated on a scale of 1 (absent) to 7 (extremely severe), and is based on information pertaining to the previous week, derived from both clinical interview and reports of primary care staff. The PANSS items are generally combined into three subscales reflecting positive symptoms (7 items), negative symptoms (7 items) and general psychopathology (16 items); however, there is substantial evidence that the items are more accurately depicted by a 5-factor structure (i.e., positive, negative, disorganized/cognitive, excited, anxiety/depression; Emsley, Rabinowitz, & Torreman, 2003; Lançon, Auquier, Nayt, & Reine, 2000). Since the PANSS items were examined individually in the current study, patient mean scores for both study 1 and study 2 (listed in Table 3) are categorized according the original description for simplicity.

Table 1. Study 1: Group Means (and Standard Deviations Unless Otherwise Specified) for Demographic and Cognitive Variables

Variables	Schizophrenia Patients (N = 334)	Siblings (N = 386)	Controls (N = 481)
Demographics			
Sex (male; female)	(245 male; 89 female)	(157 male; 229 female)	(203 male; 278 female)
Age	33.55 (10.04)	36.58 (10.44)	33.36 (10.33)
Education	14.16 (2.26)	15.86 (2.44)	16.78 (2.58)
WAIS-R FSIQ	92.69 (11.40)	105.64 (10.53)	107.87 (9.71)
WRAT-R Reading	102.28 (11.48)	106.12 (10.51)	107.63 (10.00)
Cognitive Measures			
WAIS-R Arithmetic	8.61 (2.50)	10.71 (2.53)	11.15 (2.48)
WAIS-R Similarities	9.71 (2.34)	10.91 (2.32)	10.79 (2.04)
WAIS-R Picture Completion	8.81 (2.47)	10.28 (2.44)	10.55 (2.46)
WAIS-R Digit Symbol	8.32 (2.56)	11.58 (2.50)	12.45 (2.53)
Trails B-A	52.09 (39.51)	30.36 (16.43)	27.30 (15.07)
WMS-R Digit Span Forward	8.69 (2.03)	9.58 (1.86)	9.74 (1.85)
WMS-R Digit Span Backward	6.44 (2.30)	7.76 (2.29)	8.02 (2.24)
WMS-R Logical Memory I	17.68 (7.72)	26.58 (7.03)	27.99 (6.77)
WMS-R Logical Memory II	12.91 (7.90)	22.39 (7.85)	24.08 (7.44)
WMS-R Visual Reproduction I	30.98 (5.84)	34.43 (3.86)	35.20 (3.44)
WMS-R Visual Reproduction II	25.88 (8.40)	32.20 (5.01)	33.21 (4.66)
WMS-R Verbal Paired Associates I	15.97 (4.62)	19.46 (3.32)	19.86 (3.03)
WMS-R Verbal Paired Associates II	6.87 (1.36)	7.51 (0.76)	7.64 (0.74)
CPT Vigilance	27.65 (3.83)	29.44 (1.01)	29.41 (1.43)
CPT Distractibility	23.40 (6.87)	27.99 (3.77)	28.56 (2.64)
WCST Perseverative Errors	22.62 (15.71)	12.95 (8.06)	11.64 (8.27)
WCST Categories over Trials	0.04 (0.02)	0.06 (0.19)	0.06 (0.19)
Letter Fluency	35.00 (11.59)	41.54 (11.34)	44.45 (10.46)
Category Fluency	38.02 (10.83)	49.98 (10.55)	51.99 (10.07)
Nback One	0.67 (0.25)	0.83 (0.18)	0.88 (0.14)
Nback Two	0.50 (0.22)	0.69 (0.21)	0.75 (0.19)
Nback Three	0.44 (0.17)	0.55 (0.18)	0.63 (0.18)

Note. CPT = Continuous Performance Test; FSIQ = Full Scale IQ; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; WRAT-R = Wide Range Achievement Test – Revised.

Table 2. Study 2: Group Means (and Standard Deviations Unless Otherwise Specified) for Demographic and Cognitive Variables

Variable	Schizophrenia Patients (N = 165)	Siblings (N = 253)
Demographics		
Sex (male; female)	(125 male; 40 female)	(96 male; 157 female)
Age	34.30 (9.77)	36.02 (10.08)
Education	14.29 (2.08)	15.96 (2.37)
WAIS-R FSIQ	93.26 (10.88)	105.36 (10.48)
WRAT-R Reading	102.62 (11.36)	105.65 (10.94)
Cognitive Measures		
WAIS-R Arithmetic	8.74 (2.43)	10.59 (2.57)
WAIS-R Similarities	9.79 (2.29)	10.90 (2.24)
WAIS-R Picture Completion	9.01 (2.47)	10.20 (2.38)
WAIS-R Digit Symbol	8.23 (2.44)	11.66 (2.51)
Trails B-A	51.56 (35.59)	29.40 (15.99)
WMS-R Digit Span Forward	8.67 (2.09)	9.56 (1.84)
WMS-R Digit Span Backward	6.51 (2.29)	7.82 (2.28)
WMS-R Logical Memory I	18.32 (7.98)	26.85 (7.16)
WMS-R Logical Memory II	13.41 (8.10)	22.58 (8.05)
WMS-R Visual Reproduction I	31.23 (5.83)	34.53 (3.81)
WMS-R Visual Reproduction II	26.14 (8.43)	32.20 (5.08)
WMS-R Verbal Paired Associates I	16.34 (4.41)	19.34 (3.31)
WMS-R Verbal Paired Associates II	6.87 (1.25)	7.49 (0.74)
CPT Vigilance	27.52 (4.20)	29.42 (1.06)
CPT Distractibility	23.54 (6.72)	27.89 (3.94)
WCST Perseverative Errors	22.22 (14.60)	12.99 (7.64)
WCST Categories over Trials	0.04 (0.02)	0.06 (0.02)
Letter Fluency	35.26 (11.60)	41.83 (11.40)
Category Fluency	37.68 (11.27)	50.26 (10.79)
Nback One	0.68 (0.25)	0.83 (0.18)
Nback Two	0.51 (0.22)	0.68 (0.20)
Nback Three	0.44 (0.17)	0.54 (0.18)

Note. CPT = Continuous Performance Test; FSIQ = Full Scale IQ; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; WRAT-R = Wide Range Achievement Test - Revised.

Table 3. Studies 1 & 2: Mean Positive and Negative Syndrome Scale (PANSS) Symptom Scores for Schizophrenia Patients

	Study 1		*Study 2*	
PANSS Subscale	*N Available (/334)*	*Mean (SD)*	*N Available (/165)*	*Mean (SD)*
Positive Scale				
Delusions	300	2.61 (2.02)	149	2.66 (2.04)
Conceptual Disorganisation	304	2.38 (1.91)	151	2.46 (1.85)
Hallucinatory Behaviour	306	2.21 (1.90)	154	2.31 (1.87)
Excitement	309	1.21 (0.87)	153	1.20 (0.86)
Grandiosity	306	1.42 (1.20)	151	1.43 (1.22)
Suspiciousness/Persecution	309	2.37 (1.84)	154	2.29 (1.80)
Hostility	310	1.29 (0.96)	155	1.32 (0.99)
Negative Scale				
Blunted Affect	309	3.04 (1.85)	154	3.18 (1.79)
Emotional Withdrawal	306	3.05 (1.98)	155	3.10 (1.91)
Poor Rapport	310	2.18 (1.71)	155	2.24 (1.75)
Passive/Apathetic Social Withdrawal	306	2.71 (1.94)	155	2.80 (1.87)
Difficulty in Abstract Thinking	307	3.20 (2.01)	154	3.22 (1.98)
Lack of Spontaneity	309	3.03 (2.04)	154	3.03 (2.03)
Stereotyped Thinking	288	1.32 (0.97)	146	1.32 (0.96)
General Psychopathology				
Somatic Concern	307	1.66 (1.38)	153	1.64 (1.40)
Anxiety	311	2.14 (1.52)	157	2.14 (1.44)
Guilt Feelings	305	1.36 (0.93)	151	1.40 (0.96)
Tension	310	1.80 (1.27)	155	1.83 (1.29)
Mannerisms & Posturing	310	1.15 (0.58)	156	1.15 (0.55)
Depression	309	1.69 (1.23)	156	1.61 (1.14)
Motor Retardation	311	1.78 (1.24)	156	1.73 (1.12)
Uncooperativeness	310	1.29 (1.02)	155	1.31 (1.10)
Unusual Thought Content	303	2.72 (2.16)	151	2.79 (2.15)
Disorientation	309	1.06 (0.40)	154	1.05 (0.39)
Poor Attention	309	1.58 (1.28)	153	1.56 (1.23)
Lack of Judgment & Insight	304	2.35 (2.00)	151	2.34 (2.02)
Disturbance of Volition	307	1.58 (1.42)	153	1.55 (1.39)
Poor Impulse Control	307	1.24 (0.89)	153	1.27 (0.95)
Preoccupation	304	2.14 (1.82)	151	2.07 (1.81)
Active Social Avoidance	310	2.50 (1.96)	154	2.32 (1.83)

Cognitive Measures

The cognitive variables of interest included scores from common neuropsychological measures assessing a wide range of cognitive domains (e.g., executive functioning, memory, attention). The following measures were selected because they were administered throughout the entire CBDB study and allowed for the largest possible final sample for the current research.

Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981). The WAIS is the most widely-administered test of general intellectual functioning in both clinical and research settings (Spreen & Strauss, 1991). The WAIS-R Full Scale IQ (FSIQ) standard score, as well as raw scores from the Arithmetic, Similarities, Picture Completion, and Digit Symbol subscales, were included in the current study. The Arithmetic subscale consists of up to 14 mathematical word problems (e.g., "If you walk at three miles an hour, how long would it take you to walk 24 miles?") in increasing order of difficulty. The score used was the number of correct responses compounded with time bonuses acquired when responding quickly to the items of higher difficulty (Lezak, 1995). In the Similarities test, participants are presented with a series of word pairs and are asked to consider what each pair might have in common. Abstract interpretations are scored more favourably than concrete interpretations (2 points vs. 1 point) and the score used for the current study was the sum of all points across the task. The Picture Completion subscale consists of incomplete images of familiar objects/scenes or human features, and participants are required to state which part of the image is missing. The score used was the number of correct responses. Finally, for the Digit Symbol task, participants were presented with a code key pairing the digits 1 through 9 with a unique symbol and were required to reproduce the proper symbol for a list of numbers as quickly as possible. The score used was the number of symbols correctly reproduced.

Trail-Making Test (TMT; Reitan & Wolfson, 1985). The TMT assesses visual scanning, attention, and cognitive set-shifting, and involves tracing a line on a piece of paper to follow a set of 25 numbers and/or letters in chronological/alphabetical order (Spreen & Strauss, 1991). TMT-A is considered a pure information processing speed task, in which participants trace a line from the numbers 1 to 25; TMT-B includes both numbers and letters and participants are required to alternate between number and letter when tracing a line (i.e., 1-A-2-B-3-C, etc.), thereby introducing a set-shifting component. The final score used for the present study was the difference between the time in seconds taken to complete TMT-A and TMT-B, calculated by subtracting the former from the latter, leading to a difference score (Trails B-A). This score is assumed to reflect a more pure measure of cognitive set-shifting, with the influence of processing speed subtracted out (Lezak, 1995).

Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987). The Wechsler Memory Scale is a standard measure of memory consisting of various tests designed to assess verbal and non-verbal memory (Spreen & Strauss, 1991); the Digit Span, Logical Memory, Visual Reproduction, and Verbal Paired Associates (VPA) subscales were included in the current study. For Digit Span, participants repeat an increasingly long string of numbers either forward or backward (Digit Span Forward and Backward subscales, respectively). The remaining subscales (i.e., Logical Memory, Visual Reproduction, and Verbal Paired Associates) include both immediate and delayed (30 minute) recall conditions (henceforth referred to as I and II for immediate and delayed recall, respectively). The Logical Memory subscale consists of two stories that are read aloud to participants, followed by a recall period in which participants are asked to list the main points of each story. For Visual Reproduction, participants are very briefly presented with images of geometric designs and asked to reproduce them on a sheet of paper.

Finally, for the VPA subscale, participants are read a group of eight pairs of words and are required to correctly identify the paired word when given the first word of one of the pairs. Raw scores on Digit Span (number of strings correctly recalled), Logical Memory (sum of number of ideas correctly recalled across both stories), Visual Reproduction (revised scoring; maximum score=41; Lezak, 1995) and VPA (sum of correct pairs recalled on easy and difficult items) were used for the present study.

Continuous Performance Test (CPT; Gordon, 1983). The Gordon CPT is a measure of sustained and focused attention and includes two conditions: vigilance and distractibility. During the vigilance condition, participants are required to pay attention to a string of numbers and push a button when a target string appears on screen. The distractibility condition is identical to the vigilance condition, except that distractor numbers appear elsewhere on screen. The scores used for the current study were the number of correctly identified targets in each condition.

Wisconsin Card Sorting Test (WCST; Heaton, 1981). The WCST is a measure of cognitive set-shifting and abstract reasoning, and is often used as an index of executive functioning (Laws, 1999; Spreen & Strauss, 1991). This test involves several cards with pictures of varying shapes and colours (e.g., two red circles, four blue triangles), and can be administered electronically on a computer or with physical cards. Participants are asked to match these cards to one of four stimulus cards that differ on shape, colour and number based on an unspecified rule (i.e., match by colour, shape, or number). Participants are expected to determine the rule as they match the cards through trial and error. The rule by which cards are to be matched is covertly modified after ten trials and participants are expected to find the new rule by the same process. The number of perseverative errors (i.e., continuing to categorize the cards based on an old rule) and the number of categories achieved were used as the scores for the current study. It should be

noted that the scores for the number of categories achieved were divided by the number of trials completed in order to equate different versions of the WCST administered across the CBDB study, leading to a "WCST categories over trials" variable.

Controlled Oral Word Association (Letter Fluency; Benton & Hamsher, 1989). In this test of verbal fluency, participants are required to list as many words they can think of that begin with a certain letter, excluding proper nouns, numbers, and variations on the same word (e.g., plurals). This test includes three trials of one minute each, and the letters used were F, A, and S, which are those most commonly used for this task. The score used was the sum of the number of words correctly listed for each of the three trials.

Category Naming (Category Fluency; Goodglass & Kaplan, 1983). This test is similar to the letter fluency task except that participants are required to name objects from certain categories instead of words starting with certain letters. For example, participants might be asked to name as many animals, fruits, or vegetables they can think of in one minute. The score used in the present study was the sum of all correctly identified words within a category across three one-minute trials.

Nback (Kirchner, 1958). The Nback is a type of continuous performance test in which participants are presented with a continuous sequence of letters and asked to indicate when the current stimulus matches the one that appeared n steps earlier in the sequence. In the current study, one-, two-, and three-back versions of the Nback task were used. Zero-back was removed at the data analysis stage due to a lack of variability across participants (i.e., ceiling effects). The proportions of correctly-identified stimuli were the scores used for each of the Nback conditions.

Wide Range Achievement Test – Revised (Jastak & Wilkinson, 1984). The reading subtest of the WRAT-R was included as a measure of word reading and premorbid cognitive ability in the present study. This test consists of reading and pronouncing words aloud, with the final score being the number of words correctly produced.

Data Analysis Procedure

Principal Component Analysis Followed by Analysis of Variance

The data subjected to the PCA for the PCA-ANOVA in Study 1 were participants' ranked scores on the 22 cognitive variables described above (also see the Cognitive Measures portion of Table 1), standardized such that the mean and standard deviation of each column was 0 and 1, respectively. The scores were ranked and standardized as a means of correcting for non-normality and so that the variability in the variables' ranges (e.g., Nback scores ranged from 0-1, whereas the maximum category fluency score was 98) would not lead certain measures to influence the components that emerged from the PCA more than others. All PCA solutions (including those within CPCA) were separately rotated using varimax with Kaiser normalization. The number of components retained was determined by inspection of scree plots (Cattell, 1966; Cattell & Vogelmann, 1977). Computations were carried out using MATLAB version 7.6 (The MathWorks, 2008, Natick, MA) for the PCA and CPCAs, and IBM SPSS Statistics version 19 (IBM Corporation, 2010, Armonk, NY) for the ANOVA.

A one-way ANOVA with two contrasts (schizophrenia patients versus siblings and siblings versus controls) was used to examine group differences for each of the components extracted from the PCA described above. The dependent variables consisted of participants' component scores on each of the components that emerged from the PCA, and the independent variable was

group (i.e., schizophrenia patient, sibling, and healthy control). The contrast coefficients were -1, 1, and 0 for the contrast between patients and siblings and 0, -1, 1 for the contrast between siblings and controls for schizophrenia patients, siblings, and healthy controls, respectively.

Constrained Principal Component Analysis

For the CPCA, the criterion data, *Z*, consisted of participants' ranked scores on the 22 cognitive variables listed above (this is the same criterion data used in the PCA above). For Study 1 (Z = 1201 participants by 22 variables), participants' scores were standardized such that the mean and standard deviation of each column was 0 and 1, respectively. For Study 2 (Z = 418 participants by 22 variables), participants' scores were transformed into difference-from-control scores (comparing each group to controls) prior to ranking and standardization. This was achieved by subtracting each participant's score from the mean score of the control group on each variable.

For Study 1, the set of predictor variables, *G*, consisted of two group contrasts (1201 participants by 2 contrasts) comparing the schizophrenia group to the sibling group and comparing the sibling group to the control group. For the patient versus sibling contrast, schizophrenia patients, siblings and controls were coded as -1, 1, and 0, respectively. For the sibling versus control contrast, schizophrenia patients, siblings and controls were coded as 0, -1, and 1, respectively. Study 2 included two sets of predictor variables, the first of which consisted of a group contrast (*GRP*: 418 participants by 1 contrast), comparing the schizophrenia and sibling groups (coded as -1 and 1, respectively). The second set of predictor variables consisted of dummy-coded variables for each of the 165 families included in the analysis (*FAM*: 418 participants by 165 families). The interaction between the group contrast and families,

*GRP*x*FAM*, was calculated by multiplying *GRP* and *FAM*, leading to a matrix of 418 participants by 165 families, in which patients and siblings within each family were coded as -1 and 1, respectively.

Study 1: Component Structure of Cognitive Deficits in Schizophrenia

Results

Group differences were present for: age, $F(2, 1198) = 12.24$, $p < .001$, $\eta^2 = 0.02$; sex, χ^2 $(1, N = 1201) = 97.93$, $p < .001$, $\varphi = 0.29$; education, $F(2, 1198) = 113.20$, $p < .001$, $\eta^2 = 0.15$; and IQ (WRAT-R Reading, $F(2, 1198) = 25.68$, $p < .001$, $\eta^2 = 0.04$; WAIS-R FSIQ, $F(2, 1198) = 226.27$, $p < .001$, $\eta^2 = 0.27$). Bonferonni-adjusted post-hoc z-tests for the Chi-square indicated that the schizophrenia patient group consisted of more males than both the sibling and healthy control groups, who did not differ from each other. Least significant difference (LSD) post-hoc comparisons for the ANOVAs confirmed that: (1) siblings were older than both schizophrenia patients and controls, who did not differ from each other; (2) controls had achieved a higher level of education than siblings, who in turn were more educated than schizophrenia patients; (3) schizophrenia patients had significantly lower premorbid IQ (WRAT-R Reading) scores than siblings, who in turn had lower scores than controls; and (4) schizophrenia patients had lower WAIS-R FSIQ scores than siblings, who in turn had lower scores than controls.

Principal Component Analysis Followed by Analysis of Variance

The PCA revealed a five component solution, and component loadings for each are presented in Table 4. These components were labeled Working Memory/Attention, Verbal Memory, Perceptual Organization, Digit Span, and Visual Memory, and each accounted for over 8% of the overall variance (see Table 5 "% Overall"). The Working Memory/Attention domain showed the strongest contributions from the Nback and CPT subscales, as well as WAIS-R Digit Symbol and Trails B-A (reversed). Verbal Memory was dominated by WMS-R Logical Memory and VPA I and II, and also showed a strong contribution from Category Fluency.

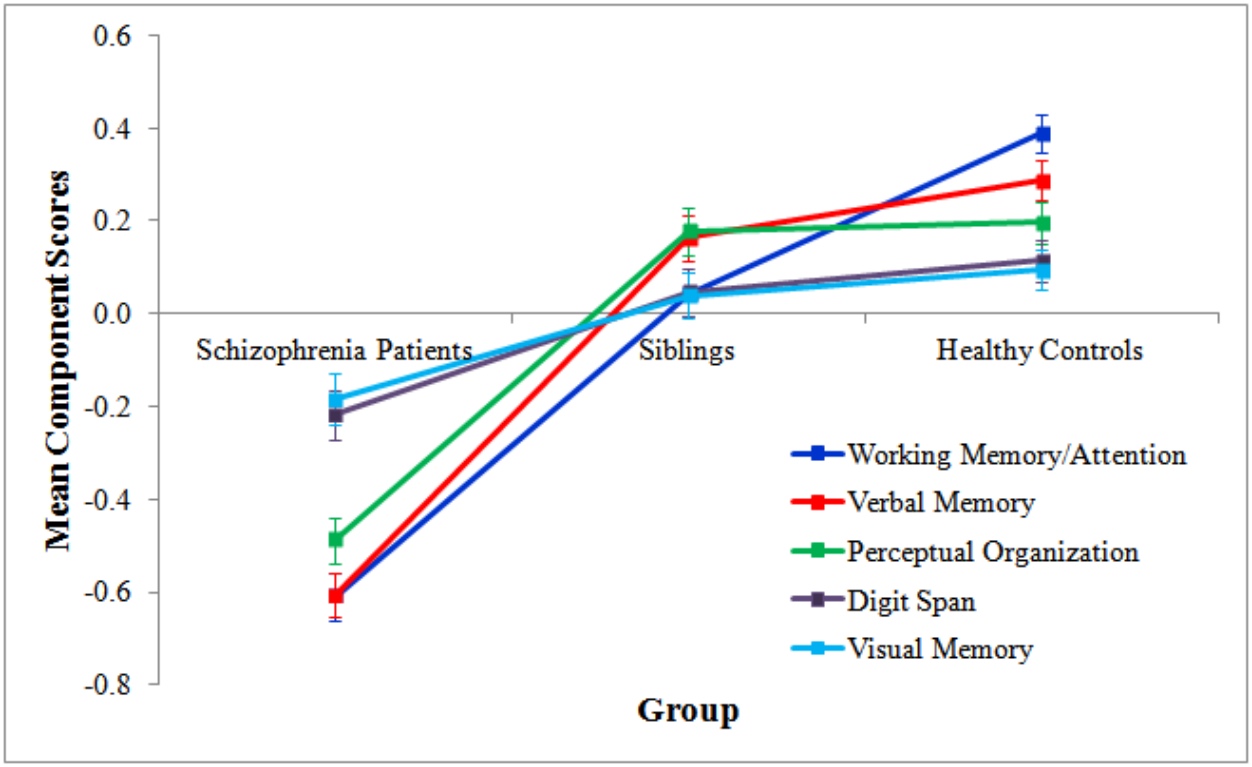
Table 4. Study 1: Component Loadings for the Five Components Extracted from the Principal Component Analysis

Variables	*Working Memory/ Attention*	*Verbal Memory*	*Perceptual Organization*	*Digit Span*	*Visual Memory*
Nback Two	**0.84**	0.17	0.10	0.16	0.11
Nback One	**0.80**	0.19	0.08	0.13	0.09
Nback Three	**0.79**	0.14	0.09	0.13	0.20
CPT Distractibility	**0.56**	0.11	0.26	0.12	0.18
WAIS-R Digit Symbol	**0.47**	0.28	**0.40**	0.17	0.05
Trails B-A	**-0.45**	-0.20	-0.19	-0.38	-0.08
CPT Vigilance	**0.42**	0.13	0.30	0.03	0.13
WMS-R Logical Memory II	0.26	**0.85**	0.16	0.11	0.06
WMS-R Logical Memory I	0.26	**0.83**	0.17	0.12	0.02
WMS-R Verbal Paired Associates I	0.12	**0.68**	0.17	0.13	0.30
WMS-R Verbal Paired Associates II	0.13	**0.56**	0.14	0.05	0.35
Category Fluency	0.21	**0.46**	**0.46**	0.29	-0.07
WCST Perseverative Errors	-0.36	-0.04	**-0.72**	0.03	-0.28
WCST Categories over Trials	0.36	0.12	**0.69**	0.00	0.28
WAIS-R Picture Completion	0.11	0.12	**0.59**	0.13	0.15
WAIS-R Similarities	-0.06	0.30	**0.57**	0.19	0.00
Letter Fluency	0.21	0.25	**0.42**	**0.42**	-0.23
WMS-R Digit Span Forward	0.09	0.07	-0.01	**0.83**	0.16
WMS-R Digit Span Backward	0.20	0.12	0.18	**0.74**	0.16
WAIS-R Arithmetic	0.35	0.33	0.34	**0.43**	0.05
WMS-R Visual Reproduction I	0.25	0.16	0.18	0.17	**0.78**
WMS-R Visual Reproduction II	0.27	0.24	0.21	0.16	**0.75**

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.40 are set in bold.

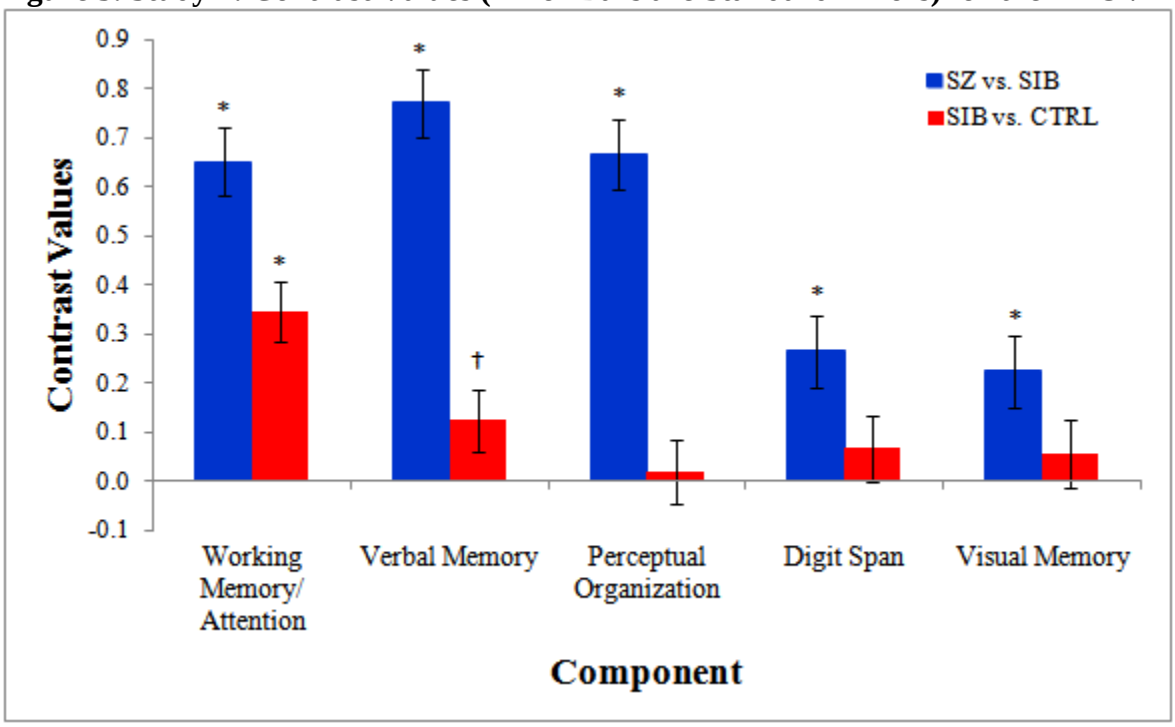
Perceptual Organization showed contributions from WCST Perseverative Errors (reversed) and Categories over Trials, WAIS-R Picture Completion and Similarities, and, to a lesser degree, Letter and Category Fluency, and WAIS-R Digit Symbol. The Digit Span domain showed strong contributions from WMS-R Digit Span Forward and Backward, and also showed lesser contributions from Letter Fluency and WAIS-R Arithmetic. Finally, Visual Memory was dominated by WMS-R Visual Reproduction I and II, with no other substantial contributions. Figure 2 displays group mean component scores and standard errors for each of these five components.

Figure 2. Study 1: Group Mean Component Scores (Error Bars are Standard Errors) for the Five Components Extracted from the Principal Component Analysis



Contrast values and standard errors for the ANOVA are displayed in Figure 3, and can be interpreted alongside Figure 2. Schizophrenia patients scored significantly lower than their siblings on all 5 components: Working Memory/Attention, $t(1198) = 9.53$, $p < .001$, $\eta^2 = 0.07$; Verbal Memory, $t(715.13) = 11.12$, $p > .001$, $\eta^2 = 0.15$; Perceptual Organization, $t(1198) = 9.34$, $p < .001$, $\eta^2 = 0.07$; Digit Span, $t(1198) = 3.58$, $p < .001$, $\eta^2 = 0.01$; Visual Memory, $t(1198) = 3.01$, $p < .005$, $\eta^2 = 0.01$. Siblings scored significantly lower than controls on Working Memory/Attention, $t(1198) = 5.54$, $p < .001$, $\eta^2 = 0.03$, with a trend towards significance for Verbal Memory, $t(805.02) = 1.91$, $p < .06$, $\eta^2 = 0.01$. There were no significant differences between siblings and controls on Perceptual Organization, Digit Span, or Visual Memory (all *p*s > .30).

Figure 3. Study 1: Contrast Values (Error Bars are Standard Errors) for the ANOVA



Note. * = p < .005; † = p < .06; CTRL = Controls; SIB = Siblings; SZ = Schizophrenia Patients

Constrained Principal Component Analysis

Table 5 presents the distribution of variance for each of the elements of the CPCA for Study 1 (i.e., overall, predicted, and residual solutions). The external analysis shows that the group contrasts accounted for 15.23% of the overall variance, while the internal analysis shows the percentage of variance accounted for by the components extracted from each solution. Five components were extracted from the overall variance (see above for a description of these results), a single component was extracted from the variance predictable by group membership, and five components were extracted from the residual variance, which is that portion of variance that is not predictable by group membership.

Table 5. Study 1: Variance (Cell Values in Regular Font) and Percentage of Variance (Cell Values in Italics) Accounted for by the Constrained Principal Component Analysis

	External	Internal					All Comps
		1	2	3	4	5	
Overall	22.00	3.70	3.03	2.77	2.09	1.79	13.37
% Overall	*100.00*	*16.83*	*13.77*	*12.57*	*9.48*	*8.14*	*60.80*
ME Group	3.35	3.32	-	-	-	-	3.32
% ME Group	*100.00*	*99.15*	-	-	-	-	*99.15*
% Overall	*15.23*	*15.10*	-	-	-	-	*15.10*
Residual	18.65	2.53	2.12	2.08	1.85	1.54	10.13
% Residual	*100.00*	*13.58*	*11.39*	*11.16*	*9.92*	*8.24*	*54.30*
% Overall	*84.77*	*11.51*	*9.66*	*9.46*	*8.41*	*6.99*	*46.03*

Note. ME = Main Effect. The variance accounted for by the external analysis and each component extracted in the internal analysis is listed in rows labelled in regular font. The percentages of variance accounted for by the external analysis and each component extracted in the internal analysis are listed in rows labelled in italic font. *% Overall*: percentage of overall variance attributable to the source identified in each column. *% ME Group*: percentage of predictable variance attributable to the source identified in each column and *% Residual*: percentage of residual variance attributable to the source identified in each column. Values can be computed by dividing the appropriate variance values listed in regular font. All internal analyses were separately rotated using varimax with Kaiser normalization. Order of components corresponds to the magnitude of variance explained.

The component loadings for the single component extracted from the predicted solution are listed in Table 6. As can be seen in Table 5 ("ME Group"), this component accounted for 15.10% of the overall variance and 99.15% of the variance accounted for by group differences. WAIS-R Digit Symbol and WMS-R Logical Memory I and II showed the highest loadings, with lesser contributions from Category Fluency and CPT Distractibility; however, this component also showed decreasing contributions from the remaining variables. The predictor loadings for this solution (i.e., correlations between component scores and the group contrasts) revealed that schizophrenia patients had lower scores than their siblings (r = 0.72), who in turn scored lower than controls (r = 0.25), on this component. Moreover, the magnitude of the correlations suggests that the difference between schizophrenia patients and siblings was greater than the difference between siblings and controls, suggesting that siblings scored closer to controls. This is reflected in the groups' mean component scores, displayed in Figure 4. The pattern of these component scores across the groups also reflected the pattern demonstrated for each of the overall components using ANOVAs (see Figure 2).

Table 7 presents the component loadings for the five components extracted from the residual solution. These cognitive domains resembled those components that emerged from the PCA on the overall data (i.e., Working Memory/Attention, Verbal Memory, Perceptual Organization, Digit Span, and Visual Memory); however, differences in the strengths of the component loadings and amounts of variance explained by each component reflect the variance removed by partialling out group differences. For example, the loadings in the residual solution (listed in Table 7) are smaller overall compared to those in the overall solution (see Table 4). This is because the components in the residual solution account for a smaller portion of the overall variance (mean percentage explained = 9.21% for the residual solution versus 12.16% for the

overall solution). In addition, those variables with high loadings on the component extracted from the predicted solution contributed less strongly to each component in the residual solution relative to the overall solution (e.g., WAIS-R Digit Symbol's loading on Working Memory/Attention is almost two times larger in the overall solution than in the residual solution).

Table 6. Study 1: Component Loadings for the Portion of Variance Predicted by Group Membership

Variables	*General Cognitive Ability*
WAIS-R Digit Symbol	**0.54**
WMS-R Logical Memory II	**0.50**
WMS-R Logical Memory I	**0.50**
Category Fluency	**0.48**
CPT Distractibility	**0.45**
Nback Two	**0.43**
WCST Categories over Trials	**0.42**
WMS-R Visual Reproduction II	0.39
Nback One	0.38
WAIS-R Arithmetic	0.38
WCST Perseverative Errors	-0.38
WMS-R Verbal Paired Associates I	0.38
Nback Three	0.37
Trails B-A	-0.35
CPT Vigilance	0.35
WMS-R Visual Reproduction I	0.34
Letter Fluency	0.32
WMS-R Verbal Paired Associates II	0.29
WMS-R Digit Span Backward	0.28
WAIS-R Picture Completion	0.27
WMS-R Digit Span Forward	0.22
WAIS-R Similarities	0.21

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.40 are set in bold.

Figure 4. Study 1: Group Mean Component Scores (Error Bars are Standard Errors) for the Component Extracted from the Portion of Variance Predicted by Group Membership

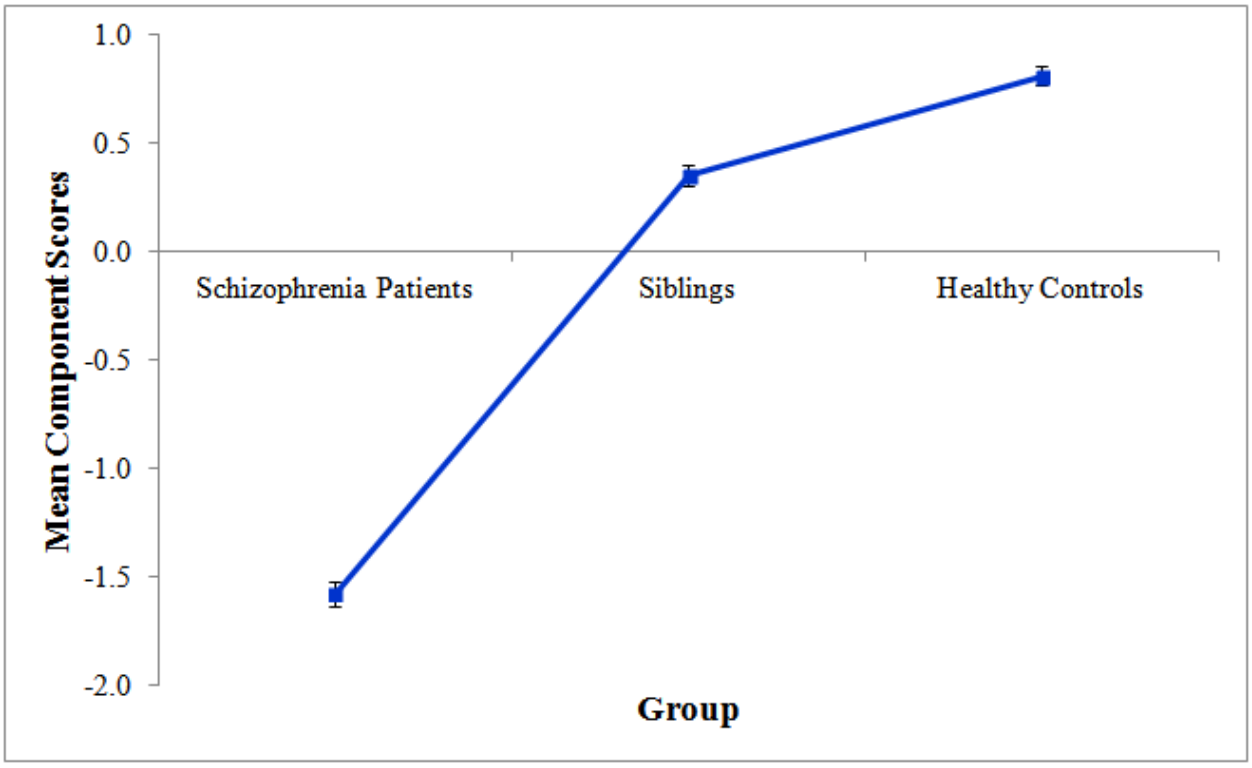


Table 7. Study 1: Component Loadings for the Residual Solution

Variables	*Working Memory/ Attention*	*Verbal Memory*	*Perceptual Organization*	*Digit Span*	*Visual Memory*
Nback Two	**0.76**	0.10	0.06	0.11	0.04
Nback One	**0.73**	0.14	0.04	0.09	0.03
Nback Three	**0.73**	0.10	0.07	0.08	0.13
CPT Distractibility	**0.42**	-0.02	0.16	0.09	0.15
CPT Vigilance	0.33	0.03	0.23	0.02	0.11
WMS-R Logical Memory II	0.13	**0.73**	0.07	0.07	-0.01
WMS-R Logical Memory I	0.13	**0.72**	0.08	0.08	-0.04
WMS-R Verbal Paired Associates I	0.04	**0.62**	0.12	0.10	0.26
WMS-R Verbal Paired Associates II	0.07	**0.52**	0.12	0.02	0.31
WCST Perseverative Errors	-0.27	0.02	**-0.67**	0.05	-0.23
WCST Categories over Trials	0.26	0.03	**0.63**	-0.03	0.22
WAIS-R Picture Completion	0.07	0.09	**0.55**	0.10	0.10
WAIS-R Similarities	-0.07	0.29	**0.55**	0.18	-0.05
Category Fluency	0.04	0.31	0.35	0.28	-0.09
WAIS-R Digit Symbol	0.28	0.10	0.28	0.14	0.02
WMS-R Digit Span Forward	0.05	0.02	-0.07	**0.79**	0.16
WMS-R Digit Span Backward	0.16	0.07	0.12	**0.71**	0.15
Letter Fluency	0.10	0.16	0.36	**0.43**	-0.23
WAIS-R Arithmetic	0.29	0.27	0.28	**0.40**	0.00
Trails B-A	-0.36	-0.12	-0.13	-0.37	-0.07
WMS-R Visual Reproduction I	0.19	0.10	0.14	0.13	**0.75**
WMS-R Visual Reproduction II	0.18	0.17	0.15	0.11	**0.72**

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.40 are set in bold.

Potential Moderating/Confounding Variables

Given that age, sex, education and IQ were significantly different between the groups (see above), subsequent analyses were conducted to examine whether these variables demonstrated moderating or confounding effects.

IQ

Although both WRAT-R Reading and WAIS-R FSIQ differed significantly between the groups, they were not included in the following analyses assessing potential moderating/confounding variables. IQ is likely an important contributor to the variability in the cognitive measures included in the current study. Moreover, the WAIS-R FSIQ score is derived from the WAIS-R subscales (Arithmetic, Similarities, Picture Completion, and Digit Symbol), which were included as dependent variables. Finally, the component resulting from the portion of variance in cognitive performance attributable to group differences, which is hypothesized to reflect a generalized cognitive domain, is likely partially influenced by IQ. For these reasons, examining potential moderating and confounding effects of IQ was not feasible in the current study.

Potential Moderating Variables

In order to examine the potential moderating effects of age, sex, and education, higher-order interactions between group and each variable were computed separately. Three CPCAs with hierarchical regression similar to that conducted in Study 2 (see "Study 2. Two Sets of Predictor Variables" in the Appendix for mathematical calculations) were used to investigate the potential two-way interactions between group and age, sex, and education. The percentages of variance accounted for by the combination of, and interaction between, group and each of these

three variables are listed in Table 8. Given that the interactions between group and age, sex, and education each accounted for less than 0.20% of the overall variance, it was concluded that diagnosis-dependent group differences on cognitive performance were not dependent on any of these potential moderating variables; therefore, no further analyses were conducted.

Table 8. External Sources of Variance Accounted for by the Interactions Between Group and Age, Sex, or Education for Study 1

Source	Variance	% Overall
Overall	22.00	100%
Study 1: Two-way Interactions between Group and Age		
[Age, Group]	3.46	15.72%
Age X Group	0.04	0.19%
Study 1: Two-way Interactions between Group and Sex		
[Sex, Group]	3.42	15.53%
Sex X Group	0.04	0.17%
Study 1: Two-way Interactions between Group and Education		
[Education, Group]	3.38	15.34%
Education X Group	0.03	0.14%

Potential Confounding Variables

In order to examine whether age, sex, and education were associated with cognitive performance in addition to differing between the groups, these variables were correlated with performance on an aggregate score of the five cognitive measures with the highest component loadings on the domain resulting from the variance attributable to group (i.e., WAIS-R Digit Symbol, WMS-R Logical Memory I & II, Category Fluency, and CPT Distractibility; see Table 6). These correlations revealed higher performance for older ($r = 0.08$, $p < .05$, $\eta^2 = 0.00$), female ($r = 0.28$, $p < .001$, $\eta^2 = 0.08$), and more highly educated ($r = 0.43$, $p < .001$, $\eta^2 = 0.19$) individuals. Given that age, sex and education showed associations with both group (see the Results section above) and the cognitive measures, these variables may have contributed to the group differences found in the CPCA above.

As described by Miller & Chapman (2001), using ANCOVA to partial potential confounding variables out of the sets of independent and dependent variables is not suitable when dealing with variables that code for group. Partialling potential confounding variables out of group removes potentially meaningful variance from group, leading to a fragmented variable and issues with interpretation. Therefore, we examined the potential confounding effects of age, sex, and education by partialling these variables out of the dependent variables (i.e., cognitive measures) only. By doing so, we were able to control for potential confounding effects while maintaining the integrity of the independent variable (group). This analysis proceeded in identical fashion to the original CPCA analysis, except that age, sex, and education were first partialled out of the set of dependent variables. This was achieved through multivariate multiple regression, with the 22 cognitive measures as the dependent variables, and age, sex, and education (combined in a single matrix) as the independent variables. The residual scores were then used as the new dependent variables for the CPCA.

Partialling out the potential confounding variables (age, sex, and education) led to a slight reduction in the amount of variance accounted for by group (from 15.23% to 13.52%), as well as small modifications in the amounts of variance accounted for by the components emerging from all three solutions (overall, predicted, and residual; see Tables 5 vs. 9). Importantly, the component structure remained unchanged from the original CPCA. Specifically, Working Memory/Attention, Verbal Memory, Perceptual Organization, Digit Span, and Visual Memory components emerged from both the overall and residual solutions, and only a single component emerged from the predicted solution. This component was almost identical to the original analysis, with the strongest contributions from WAIS-R Digit Symbol, WMS-R Logical Memory I and II, Category Fluency, and CPT Distractibility. These results suggest that although age, sex,

and education are associated with the generalized cognitive domain that characterizes group differences in cognitive performance, this general cognitive impairment in schizophrenia patients and siblings exists over and above the effects of age, sex, and education.

Table 9. Study 1: Variance (Cell Values in Regular Font) and Percentage of Variance (Cell Values in Italics) Accounted for by the Constrained Principal Component Analysis with Confounds Removed

	External	Internal					All Comps
		1	2	3	4	5	
Overall	22.00	3.41	3.21	2.65	2.25	1.71	13.24
% Overall	*100.00*	*15.50*	*14.60*	*12.05*	*10.23*	*7.79*	*60.18*
ME Group	2.97	2.95	-	-	-	-	2.95
% ME Group	*100.00*	*99.35*	-	-	-	-	*99.35*
% Overall	*13.52*	*13.43*	-	-	-	-	*13.43*
Residual	19.03	2.46	2.34	2.05	1.98	1.53	10.36
% Residual	*100.00*	*12.94*	*12.29*	*10.77*	*10.39*	*8.04*	*54.43*
% Overall	*86.48*	*11.19*	*10.63*	*9.32*	*8.98*	*6.96*	*47.08*

Note. ME = Main Effect. The variance accounted for by the external analysis and each component extracted in the internal analysis is listed in rows labelled in regular font. The percentages of variance accounted for by the external analysis and each component extracted in the internal analysis are listed in rows labelled in italic font. *% Overall*: percentage of overall variance attributable to the source identified in each column. *% ME Group*: percentage of predictable variance attributable to the source identified in each column and *% Residual*: percentage of residual variance attributable to the source identified in each column. Values can be computed by dividing the appropriate variance values listed in regular font. All internal analyses were separately rotated using varimax with Kaiser normalization. Order of components corresponds to the magnitude of variance explained.

Discussion

With regard to the primary aim, we investigated the cognitive domains underlying group differences among schizophrenia patients, their unaffected siblings, and healthy controls using two different statistical analysis techniques: PCA-ANOVA; and CPCA.

Principal Component Analysis Followed by Analysis of Variance

The PCA revealed five components that reflected common cognitive domains identified in the literature (i.e., Working Memory/Attention, Verbal Memory, Perceptual Organization, Digit Span, and Visual Memory; Dickinson et al., 2010; Nuechterlein et al., 2004). The results of the ANOVA demonstrated that schizophrenia patients were significantly impaired on each of these domains relative to controls, and the groups showed similar patterns of performance across each domain (i.e., controls were superior to siblings, who were superior to schizophrenia patients). Siblings scored significantly lower than controls on Working Memory/Attention, with a trend towards significance for Verbal Memory, but were not distinguishable from controls on Perceptual Organization, Digit Span, or Visual Memory. These results are in line with previous research finding impairment in schizophrenia on a wide range of cognitive domains (Dickinson et al., 2007; Heinrichs & Zakzanis, 1998), and suggest that schizophrenia patients are impaired on the domains of working memory/attention, verbal memory, perceptual organization, digit span, and visual memory. Given that siblings had less widespread impairment (i.e., only on Working Memory/Attention and Verbal Memory), the PCA-ANOVA findings suggest that siblings are impaired in the domains of working memory/attention and verbal memory, with relatively intact performance on perceptual organization, digit span, and visual memory. These findings are in line with previous research reporting more reliable performance deficits in

siblings of patients with schizophrenia on those domains that are most severely impaired in the patients themselves (Sitskoorn et al., 2004). However, examining group differences within pre-defined cognitive domains simply establishes that cognitive impairment in schizophrenia is present in these domains. Such a method may not accurately reflect the nature of these deficits or whether they are due to domain-specific impairment and/or to a general cognitive impairment that manifests as deficits in these more specific cognitive domains.

Constrained Principal Component Analysis

In contrast to the PCA-ANOVA, in which group differences were examined *after* the identification of cognitive domains, CPCA separates the overall variance into that which can be accounted for by a set of independent variables and that which cannot *prior to* performing PCA to identify underlying structure. Utilization of this method in the current study allowed for the identification of cognitive domains that differed between groups rather simply determining whether group differences exist within certain cognitive domains, whose configuration is likely influenced by a host of other factors. In this way, it was possible to examine whether group differences in cognitive ability are best characterized as a general cognitive impairment or as a range of deficits in more specific cognitive domains.

The results identified a single cognitive domain accounting for group differences in cognitive performance that included similar contributions from all variables, with WAIS-R Digit Symbol and WMS-R Verbal Memory immediate and delayed subscales showing the highest loadings (see Table 6). Schizophrenia patients performed significantly worse than their siblings, who in turn performed significantly worse than controls, on this domain. These findings are in line with previous research reporting that the majority of diagnosis-based group differences in

cognitive performance are the result of a generalized cognitive impairment in schizophrenia patients and, to a lesser extent, their siblings (Dickinson et al., 2008). Moreover, this generalized cognitive domain did not match any of the five cognitive dimensions in the overall solution (which are identical to the components from the PCA described above), as these were retained in the residual solution (see Table 7). This supports the notion of a generalized cognitive deficit in schizophrenia rather than of domain-specific deficits. The single component that was extracted from the predicted solution accounted for virtually all of the variance that was predictable from group differences, and also showed relatively similar contributions from all of the included variables. If group differences were the result of multiple deficits in specific cognitive domains in schizophrenia rather than of a general cognitive impairment, one would expect these to emerge as separate components in the predicted solution, approximating the results from the PCA-ANOVA. Similarly, if group differences were the result of both a generalized cognitive deficit and of domain-specific deficits in schizophrenia, one might expect the emergence of a large component showing relatively similar contributions from all variables as was the case in the current CPCA results; however, other components reflecting those domain-specific impairments should also emerge and the larger component signifying general cognitive ability should account for much less than is the case in the present study. (Note that a two-component solution showed substantial cross-loadings for all variables that were of similar magnitude across components, suggesting no more than a single cognitive domain could account for group differences in cognitive performance.) Therefore, these results suggest not only that a generalized cognitive impairment underlies schizophrenia patients' poor performance on these more specific cognitive domains seen through the PCA-ANOVA, but also that any domain-specific deficits (if they exist) are likely secondary to a deficit in general cognitive ability.

The pattern of performance on this component (as reflected by mean component scores) is strikingly similar to the pattern underlying each of the five components extracted from the overall solution (see Figures 2 and 4 for comparison). Moreover, the cognitive measures that showed the strongest contributions to the general cognitive ability domain that emerged from the CPCA (i.e., WAIS-R Digit Symbol and WMS-R Logical Memory) were major contributors to the most severely impaired domains in the PCA-ANOVA analysis (i.e., Working Memory/Attention and Verbal Memory, respectively). In this case, the technique employed by PCA-ANOVA, namely, performing dimension reduction prior to examining group differences, is misleading. The CPCA results show that the pattern of cognitive deficits apparent in the PCA-ANOVA results is best explained by a single cognitive domain, rather than by similar patterns across multiple domains. This major discrepancy in interpretations highlights the necessity of understanding the research question that is truly being addressed with a given statistical analysis technique. When constraining the overall variance on the basis of group differences prior to identifying underlying structure (as is the case in CPCA), one is asking "*what cognitive domains account for differences in cognitive performance between schizophrenia patients, their unaffected siblings, and healthy controls?"*, whereas in the case of a PCA-ANOVA (or any statistical analysis technique in which dimension reduction is carried out prior to examining group differences), the question is more akin to *"does the performance of schizophrenia patients, their unaffected siblings, and controls differ on the cognitive domains reflected by standard clinical tests?"*. Although the latter question may be of interest, it does not provide insight into whether cognitive impairment in schizophrenia is fundamentally the result of domain-specific deficits or of a general cognitive impairment, given that the results are limited to pre-defined cognitive domains which may not necessarily reflect group differences in cognitive performance.

Although the generalized cognitive domain that emerged from the predicted solution showed relatively similar contributions from all of the variables included, it was noticeably dominated by WAIS-R Digit Symbol and WMS-R Logical Memory immediate and delayed subscales (see Table 6). This finding is in line with the literature, which reports the highest deficits on measures of information processing speed and verbal memory (Dickinson et al., 2008; Heinrichs & Zakzanis, 1998; Knowles et al., 2010). That these variables were subsumed under a component reflecting generalized cognitive ability in the current study, rather than emerging as separate components, suggests that measures of information processing speed and verbal memory are likely particularly good indicators of general cognitive ability, as has been suggested by previous researchers (Dickinson et al., 2007; González-Blanch et al., 2011), rather than primarily deficient in schizophrenia *per se*. As described above, measures of information processing speed and verbal memory generally recruit a wide range of cognitive sub-processes, a characteristic that makes them particularly suitable as indices of general cognitive ability (Dickinson, 2008; Dickinson et al., 2007; González-Blanch et al., 2011). Digit symbol coding in particular also recruits cognitive processes that rely on more psychophysiological-based functions; for example, the use of visual scanning to consult the code key for each test item during this task recruits smooth eye pursuit movements, which are known to be deficient in schizophrenia (O'Driscoll & Callahan, 2008). Potential delays caused by this deficiency in patients could increase the time required to consult and copy each symbol and, therefore, lead to a reduction in the number of items correctly reproduced in the allotted time.

It should be noted that another potential measure of processing speed included in the current study (i.e., the TMT) did not show the same high loadings on this component as did WAIS-R Digit Symbol. While it is possible that Digit Symbol is a better measure of general cognitive

ability than TMT, the latter has been shown to distinguish well between schizophrenia patients and controls (González-Blanch et al., 2011), as well as between siblings and controls (Egan, Goldberg, Gscheidle, et al., 2001), and would, therefore, be expected to load highly onto a component emerging from the portion of variance in cognitive performance that is predictable from the differences between these three groups. The reason for the discrepancy in the current study is likely based on the score used; the two TMT variables, A and B, were transformed into a single difference score as is often recommended (Lezak, 1995). By subtracting TMT-A (which is a relatively pure measure of processing speed) from TMT-B (which includes both processing speed and set-shifting), the resulting variable (Trails B-A) is thought to become a more pure measure of set-shifting. It is possible that including these as separate variables would have led each of them to display higher loadings on this component, TMT-A as a measure of processing speed and TMT-B as a more comprehensive measure of processing speed and set shifting. Indeed, a re-analysis including TMT-A and -B as separate variables led to a very similar general cognitive ability component to the one described above, but with TMT-B and -A occupying the 5^{th} and 9^{th} highest loadings, respectively, relative to the 14^{th} place rank for Trails B-A displayed in Table 6.

Overall, the results from Study 1 suggest that (1) group differences in cognitive performance between schizophrenia patients, their unaffected siblings and controls are best captured by a single cognitive domain reflecting general cognitive ability; (2) WAIS-R Digit Symbol and WMS-R Logical Memory are particularly good measures of this dimension; and that (3) schizophrenia patients as a group do not show domain-specific deficits over and above this generalized cognitive impairment. Given that this interpretation is in contrast with that resulting from the PCA-ANOVA described previously (which indicated that schizophrenia patients were

impaired on all cognitive domains), the current set of results also highlights the importance of fully understanding the research questions that different statistical methods are designed to address.

Study 2: Family Membership Patterns of Cognitive Deficits in Schizophrenia

Results

T-tests on the demographic variables listed in Table 2 revealed that schizophrenia patients were more likely to be less educated, $t(416) = -7.39$, $p < .001$, $\eta^2 = 0.12$, and have lower premorbid IQ (WRAT-R Reading, $t(416) = -2.73$, $p < .01$, $\eta^2 = 0.02$), and WAIS-R FSIQ, $t(416) = -11.37$, $p < .001$, $\eta^2 = 0.24$) scores compared to siblings. The schizophrenia patient group also consisted of more males than the sibling group, χ^2 (1, $N = 418$) = 57.31, $p < .001$, $\varphi = 0.37$.

Constrained Principal Component Analysis

Table 10 shows the distribution of variance for each of the elements of the CPCA with hierarchical regression (solutions for overall, main, interaction, and residual effects). The external analysis shows that the combination of group and family membership accounted for 59.12% of the overall variance, while the internal analysis shows the percentage of variance accounted for by the components extracted from each solution. Of the 59.12% of the overall variance predictable from the combination of group and family membership, 12.76% was due to the main effect of group independent of family membership, 46.29% was due to the main effect of family membership independent of group, and the remaining 0.07% was due to the main effect of the overlap between group and family membership. Given that the overlap only accounted for 0.07% of the overall variance, it was not analyzed further. The interaction between group and family membership accounted for 27.60% of the overall variance over and above the main effects, leaving 13.28% of the overall variance that was not predicted by the main effects of group, family membership, their overlap, or by the interaction between the two. As can be seen in the internal analysis portion of Table 10, five components were extracted from the overall

solution, one component was extracted from the main effect of group, two components were extracted from the main effect of family, and three components were extracted from the interaction between group and family.

Table 10. Study 2: Variance (Cell Values in Regular Font) and Percentage of Variance (Cell Values in Italics) Accounted for by the Constrained Principal Component Analysis

	External	Internal					All
		1	2	3	4	5	Comps
Overall	22.00	3.24	3.13	3.09	2.25	1.98	13.69
% Overall	*100.00*	*14.74*	*14.21*	*14.06*	*10.23*	*9.01*	*62.24*
[Group Family]	13.01	3.75	3.47	-	-	-	7.22
% [Group Family]	*100.00*	*28.86*	*26.67*	-	-	-	*55.53*
% Overall	*59.12*	*17.06*	*15.77*	-	-	-	*32.83*
ME Group	2.81	2.81	-	-	-	-	2.81
% ME Group	*100.00*	*100.00*	-	-	-	-	*100.00*
% Overall	*12.76*	*12.76*	-	-	-	-	*12.76*
ME Family	10.18	2.37	2.18	-	-	-	4.56
% ME Family	*100.00*	*23.30*	*21.45*	-	-	-	*44.75*
% Overall	*46.29*	*10.79*	*9.93*	-	-	-	*20.72*
ME Group/Family	0.02	0.02	-	-	-	-	0.02
% Group/Family	*100.00*	*100.00*	-	-	-	-	*100.00*
% Overall	*0.07*	*0.07*	-	-	-	-	*0.07*
INT Group x Family	6.07	1.10	0.79	0.78	-	-	2.67
% INT Group x Family	*100.00*	*18.06*	*13.05*	*12.83*	-	-	*43.94*
% Overall	*27.60*	*4.98*	*3.60*	*3.54*	-	-	*12.13*
Residual	2.92	0.67	-	-	-	-	0.67
% Residual	*100.00*	*23.09*	-	-	-	-	*23.09*
% Overall	*13.28*	*3.07*	-	-	-	-	*3.07*

Note. ME = Main Effect; INT = Interaction Effect. The variance accounted for by the external analysis and each component extracted in the internal analysis is listed in rows labelled in regular font. The percentages of variance accounted for by the external analysis and each component extracted in the internal analysis are listed in rows labelled in italic font. *% Overall*: percentage of overall variance attributable to the source identified in each column. *% [Group Family]*: percentage of combined predictable variance attributable to the source identified in each column. *% ME Group:* percentage of independent group variance attributable to the source identified in each column. *% ME Family:* Percentage of independent family variance attributable to the source identified in each column. *% ME Group/Family:* Percentage of overlapping variance attributable to the source identified in each column. *% INT Group x Family:* Percentage of group x family interaction variance attributable to the source identified in each column. *% Residual*: percentage of residual variance attributable to the source identified in each column. Values can be computed by dividing the appropriate variance values listed in regular font. All internal analyses were separately rotated using varimax with Kaiser normalization. Order of components corresponds to the magnitude of variance explained.

The five components that emerged from the overall data, which consisted of schizophrenia patients' and siblings' difference-from-control scores on the 22 variables of interest (see above), were Working Memory/Attention, Visual Memory, Verbal Memory, Fluency, and Digit Span. Table 11 displays the loadings for each of these five components, which showed strong similarities to those described in the "Principal Component Analysis Followed by Analysis of Variance" section of Study 1; differences included the Perceptual Organization and Visual Memory components combining to form an overall Visual Memory component, and the emergence of a Fluency component. As can be seen in Table 10 ("Overall"), each component accounted for over 9% of the overall variance.

The main effect of group was characterized by a single component that accounted for all of the variance predicted by group independent of family membership (i.e., 12.76% of the overall variance). This component was strikingly similar to the component described in the "Constrained Principal Component Analysis" section of Study 1 (see Table 6), with strong contributions from WAIS-R Digit Symbol, Category Fluency, CPT Distractibility, and WMS-R Logical Memory I and II (see Table 12). The predictor weight (the correlation between the patient versus sibling contrast and the component scores) reflected the fact that this portion of variance was designed to include only that predictable by group differences; thus, the contrast between schizophrenia patients and siblings correlated perfectly with their component scores (r = 1.00).

Table 11. Study 2: Component Loadings for the Overall Solution

Variables	*Working Memory/Attention*	*Visual Memory*	*Verbal Memory*	*Fluency*	*Digit Span*
Nback Two	**0.81**	0.17	0.11	0.17	0.20
Nback Three	**0.77**	0.22	0.06	0.02	0.21
Nback One	**0.76**	0.10	0.19	0.17	0.19
CPT Distractibility	**0.43**	0.32	0.20	0.36	-0.03
Trails B-A	**-0.42**	-0.20	-0.30	-0.26	-0.21
CPT Vigilance	**0.42**	0.28	0.14	0.30	-0.02
WCST Perseverative Errors	-0.33	**-0.71**	-0.08	-0.29	-0.01
WCST Categories over Trials	0.30	**0.69**	0.13	0.29	0.04
WMS-R Visual Reproduction I	0.36	**0.66**	0.23	-0.12	0.13
WMS-R Visual Reproduction II	0.34	**0.65**	0.34	-0.07	0.08
WAIS-R Picture Completion	-0.01	**0.59**	0.11	0.16	0.28
WAIS-R Similarities	-0.10	**0.48**	0.28	0.25	0.34
WMS-R Logical Memory II	0.05	0.22	**0.82**	0.21	0.23
WMS-R Logical Memory I	0.04	0.21	**0.80**	0.21	0.26
WMS-R Verbal Paired Associates I	0.18	0.16	**0.76**	0.14	0.06
WMS-R Verbal Paired Associates II	0.33	0.09	**0.64**	0.07	-0.04
Letter Fluency	0.13	0.00	0.09	**0.77**	0.24
Category Fluency	0.14	0.24	0.35	**0.68**	0.21
WAIS-R Digit Symbol	0.31	0.30	0.25	**0.59**	-0.02
WMS-R Digit Span Forward	0.16	0.05	0.05	0.01	**0.77**
WMS-R Digit Span Backward	0.26	0.15	0.15	0.22	**0.68**
WAIS-R Arithmetic	0.21	0.32	0.32	0.24	**0.56**

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.40 are set in bold.

Table 12. Study 2: Component Loadings for the Main Effect of Group Independent of Family Membership

Variables	*General Cognitive Ability*
WAIS-R Digit Symbol	**0.57**
Category Fluency	**0.49**
CPT Distractibility	**0.47**
WMS-R Logical Memory II	**0.45**
WMS-R Logical Memory I	**0.44**
WCST Categories over Trials	**0.43**
WCST Perseverative Errors	-0.39
CPT Vigilance	0.37
WMS-R Visual Reproduction II	0.36
Nback Two	0.35
Trails B-A	-0.35
WMS-R Verbal Paired Associates I	0.33
Nback One	0.32
WAIS-R Arithmetic	0.31
WMS-R Visual Reproduction I	0.29
Nback Three	0.27
WMS-R Digit Span Backward	0.27
Letter Fluency	0.25
WMS-R Verbal Paired Associates II	0.23
WAIS-R Picture Completion	0.22
WMS-R Digit Span Forward	0.20
WAIS-R Similarities	0.20

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.40 are set in bold.

The main effect of family membership independent of group revealed a two-component solution, and component loadings for each are presented in Table 13. These components reflected Verbal Memory and Non-Verbal Memory, and accounted for 10.79% and 9.93% of the overall variance, respectively. Verbal Memory showed high contributions from WMS-R Logical Memory I and II, and decreasing contributions from WAIS-R Arithmetic and Similarities,

Category Fluency, and WMS-R VPA I. Non-Verbal Memory showed high contributions from Nback One, Two, and Three, as well as decreasing contributions from WCST Perseverative Errors (reversed) and Categories over Trials, and WMS-R Visual Reproduction I and II. The predictor loadings (listed in Table 19 of the Appendix) identify the contributions of each family to each component (positive loadings = high performance; negative loadings = low performance).

Table 14 displays the component loadings for the three components that emerged from the interaction between group and family membership. These components reflected Working Memory/Attention, Visual Memory, and Verbal Memory, and accounted for 4.98%, 3.60%, and 3.54% of the overall variance, and 18.06%, 13.05%, and 12.83% of the variance predictable by the interaction, respectively (see Table 10). Working Memory/Attention was similar to the component of the same name in the overall solution, with strong contributions from WMS-R Digit Span Backward, Trails B-A (reversed), Nback One and Two, as well as CPT Vigilance and Distractibility. The Visual Memory component reflected the component of the same name in the overall solution, with strong contributions from WMS-R Visual Reproduction I and II, WCST Categories over Trials and Perseverative Errors (reversed) with a smaller contribution from WAIS-R Picture Completion. Verbal Memory reflected the component of the same name in the overall solution, with contributions from WMS-R Verbal Paired Associates I and II and Logical Memory I and II, and smaller contributions from the WAIS-R Arithmetic and Similarities subscales. The predictor loadings, listed in Table 20 of the Appendix, identify the contributions of each family to each of these components (see the "Study 2: Component Score Analysis" section in the Appendix and the Cluster Analysis section below for a more in-depth analysis of these results).

Table 13. Study 2: Component Loadings for the Main Effect of Family Membership Independent of Group

Variables	*Verbal Memory*	*Non-Verbal Memory*
WMS-R Logical Memory I	**0.64**	-0.01
WMS-R Logical Memory II	**0.62**	-0.02
WAIS-R Arithmetic	**0.47**	0.23
WAIS-R Similarities	**0.47**	0.12
Category Fluency	**0.45**	0.13
WMS-R Verbal Paired Associates I	**0.43**	0.12
WMS-R Digit Span Backward	**0.36**	0.24
WAIS-R Picture Completion	**0.33**	0.24
Letter Fluency	0.28	0.11
WMS-R Verbal Paired Associates II	0.27	0.23
WMS-R Digit Span Forward	0.25	0.14
Nback Three	-0.02	**0.58**
Nback Two	0.07	**0.56**
Nback One	0.08	**0.51**
WCST Perseverative Errors	-0.13	**-0.47**
WMS-R Visual Reproduction I	0.20	**0.41**
WCST Categories over Trials	0.20	**0.39**
WMS-R Visual Reproduction II	0.20	**0.36**
CPT Distractibility	0.11	**0.30**
CPT Vigilance	0.08	0.29
Trails B-A	-0.29	-0.29
WAIS-R Digit Symbol	0.18	0.23

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised; Values greater than or equal to 0.30 are set in bold.

Table 14. Study 2: Component Loadings for the Effect of the Interaction Between Group and Family Membership

Variables	*Working Memory/ Attention*	*Visual Memory*	*Verbal Memory*
WMS-R Digit Span Backward	**0.34**	0.02	0.08
Trails B-A	**-0.34**	-0.05	-0.12
Nback One	**0.33**	0.11	0.07
Nback Two	**0.33**	0.11	0.02
CPT Vigilance	**0.30**	0.12	-0.04
CPT Distractibility	**0.30**	0.06	0.07
Letter Fluency	0.27	-0.05	0.13
Nback Three	0.27	0.20	0.07
WAIS-R Arithmetic	0.26	0.15	0.22
Category Fluency	0.25	0.02	0.10
WAIS-R Digit Symbol	0.20	0.10	0.05
WMS-R Digit Span Forward	0.19	0.03	0.04
WMS-R Visual Reproduction II	0.04	**0.41**	0.22
WMS-R Visual Reproduction I	0.10	**0.36**	0.18
WCST Categories over Trials	0.18	**0.35**	0.01
WCST Perseverative Errors	-0.18	**-0.34**	-0.05
WAIS-R Picture Completion	-0.02	0.27	0.01
WMS-R Verbal Paired Associates II	0.09	-0.03	**0.47**
WMS-R Verbal Paired Associates I	0.09	0.05	**0.43**
WMS-R Logical Memory II	0.08	0.13	0.26
WMS-R Logical Memory I	0.07	0.11	0.25
WAIS-R Similarities	0.10	0.17	0.18

Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised.

A single component was extracted from the residual solution, which included only the variance not predictable from the main effects of group or family membership, their overlap, or the interaction between the two variables. This component reflected a more pure Working Memory domain, with the strongest loadings from Nback One, Two and Three, and decreasing contributions from the remaining variables. The low component loadings, displayed in Table 15, reflect that this component only accounted for 3.07% of the overall solution.

Table 15. Study 2: Component Loadings for the Residual Solution

Variables	*Working Memory*
Nback One	0.27
Nback Three	0.26
Nback Two	0.25
WAIS-R Picture Completion	0.24
Letter Fluency	0.23
WMS-R Verbal Paired Associates II	0.20
WMS-R Visual Reproduction II	0.19
WAIS-R Arithmetic	0.18
WAIS-R Similarities	0.17
WCST Perseverative Errors	-0.16
WMS-R Digit Span Backward	0.16
WMS-R Verbal Paired Associates I	0.15
WMS-R Visual Reproduction I	0.15
WMS-R Logical Memory II	0.15
WCST Categories over Trials	0.13
WMS-R Logical Memory I	0.13
Category Fluency	0.13
WAIS-R Digit Symbol	0.13
CPT Vigilance	0.12
CPT Distractibility	0.11
WMS-R Digit Span Forward	0.08
Trails B-A	-0.05

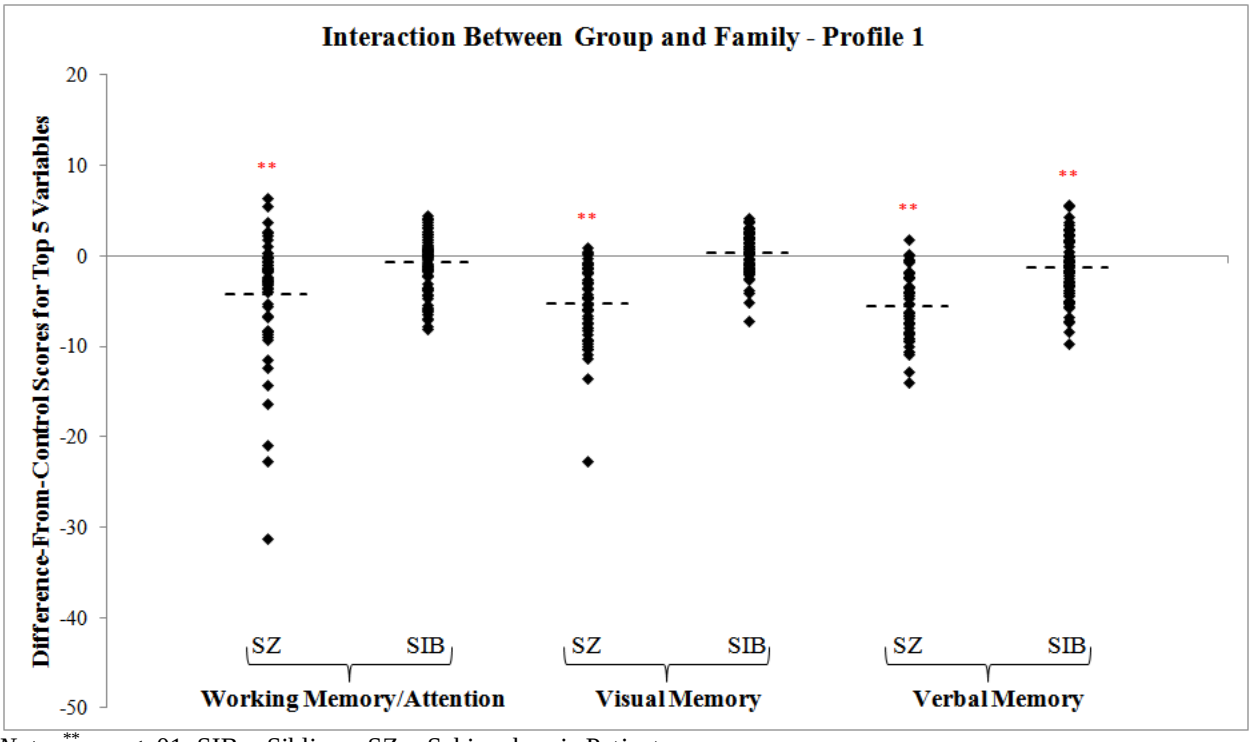
Note. CPT = Continuous Performance Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WCST = Wisconsin Card Sorting Test; WMS-R = Wechsler Memory Scale – Revised.

Cluster Analysis

Cognitive domains that represent endophenotypes in schizophrenia would be expected to show impairment in both schizophrenia patients and, to a lesser degree, in siblings from the same families. In line with the notion of multiple cognitive endophenotypes in schizophrenia, we also expected that patterns of cognitive domain impairment would differ across families. In order to investigate these hypotheses, we conducted a cluster analysis on the predictor loadings of the portion of variance attributable to the interaction between group and family membership. A nine-cluster solution was chosen (see Figure 12 in the Appendix). Only clusters with 30 or more siblings were further analyzed, leading to three clusters consisting of 54, 53, and 19 families respectively. Figures 4-6 illustrate patients' and siblings' mean difference-from control scores on the five variables with the highest component loadings on each component for these three family profiles. The dotted line crossing each collection of points represents the mean difference-from-control score for that group and component, and a value of 0 on the y-axis represents scores that are equal to controls, given that these are difference-from-control scores. In order to examine whether these subgroups were significantly impaired on each cognitive domain, one-sample t-tests were conducted for each of the patient/sibling groups' difference-from-control scores within each family profile on all three components, and alpha criteria were corrected for multiple comparisons.

Profile 1 consisted of 54 families (54 patients and 80 siblings). Patients were significantly impaired on Working Memory/Attention, $t(53) = -4.57$, $p < .001$, $\eta^2 = 0.28$, Visual Memory, $t(53) = -8.72$, $p < .001$, $\eta^2 = 0.60$, and Verbal Memory, $t(53) = -11.31$, $p < .001$, $\eta^2 = 0.71$. Siblings were significantly impaired on Verbal Memory, $t(79) = -3.57$, $p > .005$, $\eta^2 = 0.14$, but showed intact performance on Working Memory/Attention and Visual Memory (all *p*s > .01), which suggests that this family cluster showed familial impairment on verbal memory.

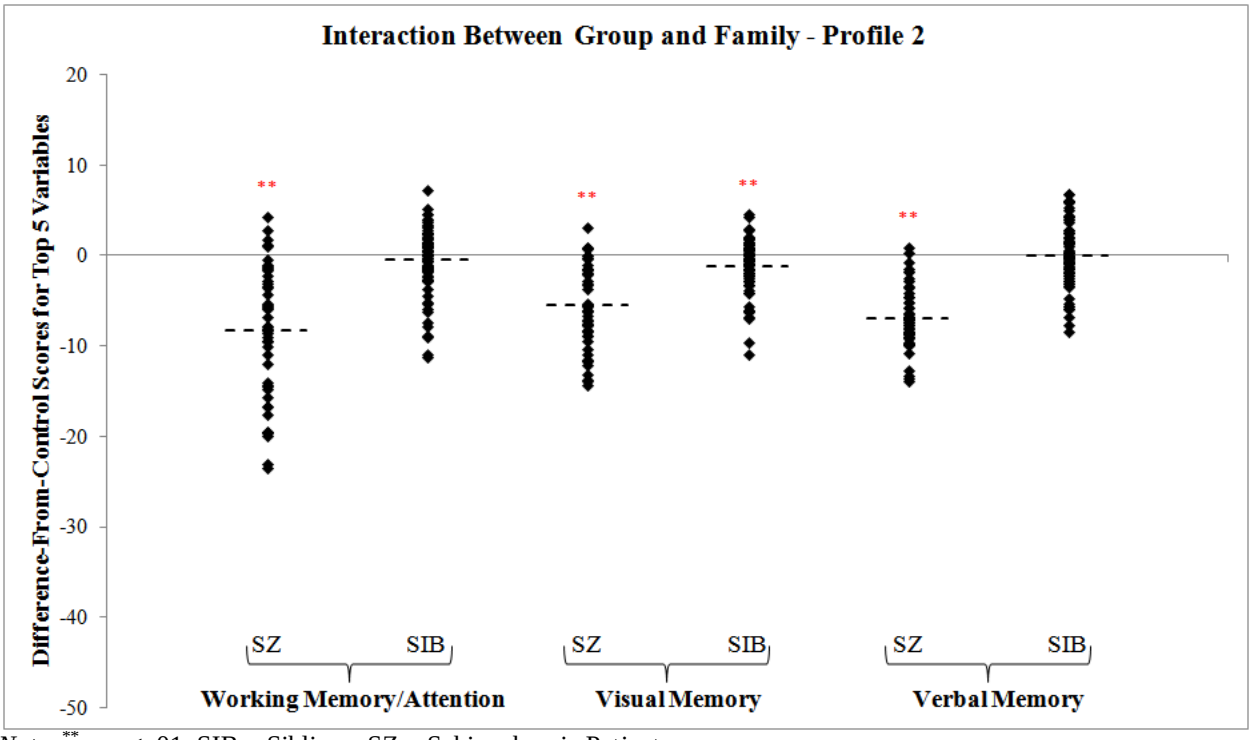
Figure 5. Study 2: Mean Difference-From-Control Scores (on the Five Variables with the Highest Component Loadings for each Component) by Participant for Family Profile 1 from the Cluster Analysis of the Interaction Between Group and Family Membership



Note. ** = $p < .01$; SIB = Siblings; SZ = Schizophrenia Patients.

Profile 2 consisted of 53 families (53 patients and 83 siblings). Patients were significantly impaired on Working Memory/Attention, $t(52) = -8.05$, $p < .001$, $\eta^2 = 0.56$, Visual Memory, $t(52) = -8.89$, $p < .001$, $\eta^2 = 0.60$, and Verbal Memory, $t(52) = -15.05$, $p < .001$, $\eta^2 = 0.81$. Siblings were significantly impaired on Visual Memory, $t(82) = -3.87$, $p > .001$, $\eta^2 = 0.15$, but showed intact performance on Working Memory/Attention and Verbal Memory (all ps > .05), which suggests that this family cluster showed familial impairment on visual memory.

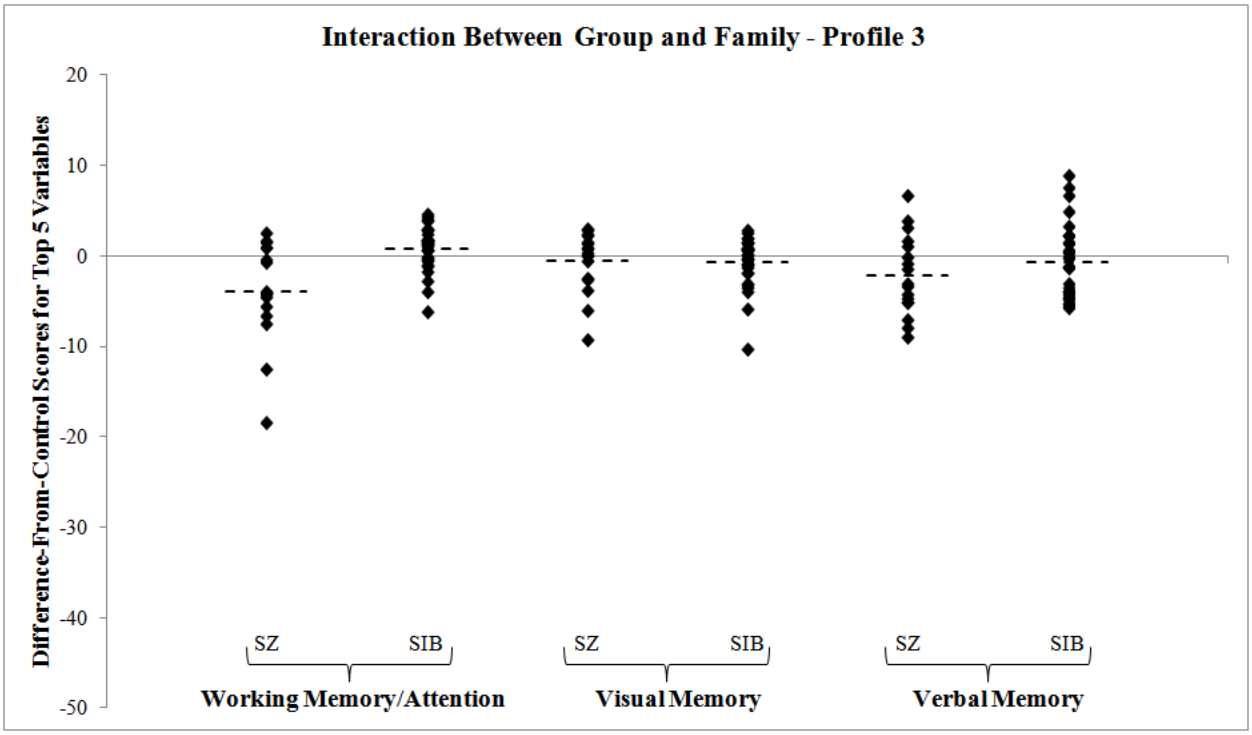
Figure 6. Study 2: Mean Difference-From-Control Scores (on the Five Variables with the Highest Component Loadings for each Component) by Participant for Family Profile 2 from the Cluster Analysis of the Interaction Between Group and Family Membership



Note. ** = $p < .01$; SIB = Siblings; SZ = Schizophrenia Patients.

Profile 3 consisted of 17 families (17 patients and 31 siblings). Both patients and siblings showed intact performance on Working Memory/Attention, Visual Memory, and Verbal Memory (all *p*s > .01).

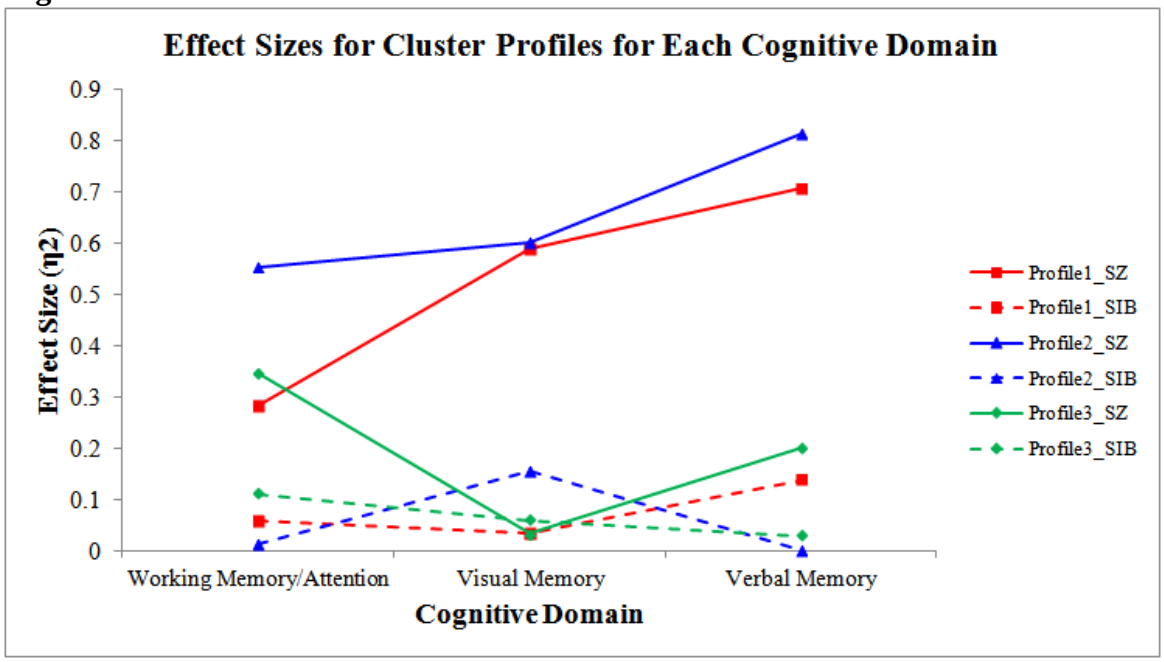
Figure 7. Study 2: Mean Difference-From-Control Scores (on the Five Variables with the Highest Component Loadings for each Component) by Participant for Family Profile 3 from the Cluster Analysis of the Interaction Between Group and Family Membership



Note. ** = $p < .01$; SIB = Siblings; SZ = Schizophrenia Patients.

A summary of the effect sizes for each profile is depicted in Figure 8. Taken together with the impairment patterns described above, these results suggest familial impairment in verbal memory for family cluster 1, and familial impairment in visual memory for family cluster 2. Although the effect size for Working Memory/Attention was large in patients from family cluster 3, these patients did not show significant impairment on this domain. Moreover, siblings from this family cluster were not impaired on any of the three cognitive domains; therefore, family cluster 3 did not demonstrate any pattern of familial impairment. Given that patients showed more widespread impairment than their siblings in the first two family clusters, this suggests that factors other than family-specific genotypes may be contributing to the family-membership-dependent group differences identified in this analysis. This is likely due to the fact that the family-membership variable, which simply coded whether a given individual was part of a given family, necessarily included both genetic and environmental factors attributable to family.

Figure 8. Study 2: Effect Sizes for Each Cognitive Domain for Schizophrenia Patients and Siblings from Each Cluster Profile



Note. SIB = Siblings; SZ = Schizophrenia Patients.

Potential demographic differences between the three family profiles described above were examined by performing ANOVAs (or Chi Square as appropriate) on the schizophrenia patients and siblings separately. Age, sex, education, WRAT-R Reading, and WAIS-R FSIQ were included as the dependent variables, and profile membership was the independent variable. For schizophrenia patients, the 30 PANSS subscales were also included as dependent variables. Patients from the three profiles differed significantly on WAIS-R FSIQ, $F(2,120) = 4.29$, $p < .05$, $\eta^2 = 0.07$, and the Somatic Concern subscale of the PANSS, $F(2,120) = 3.29$, $p < .05$, $\eta^2 = 0.05$ (see Table 16 for profile comparisons on all demographic/symptom measures for patients). LSD post-hoc comparisons indicated that patients from profile 3 had higher WAIS-R FSIQ scores than patients from profile 2, and less Somatic Concern than patients from profile 1, with no other significant differences between the groups. Siblings from the three family profiles did not differ significantly on any of the demographic variables (all *p*s > .20).

Table 16. Comparisons Between Family Clusters (Patients Only) on Demographic & Symptom Measures

Variable	*Statistic*
Demographic Variables	
Sex	χ^2 (1, N = 1201) = .12
Age	F(2,122) = 0.53
Education	F(2,122) = 0.64
WAIS-R FSIQ	F(2,122) = 4.29*
WRAT-R Reading	F(2,122) = 0.77
PANSS Positive Subscale	
Delusions	F(2,111) = 2.50
Conceptual Disorganisation	F(2,114) = 0.48
Hallucinatory Behaviour	F(2,116) = 0.17
Excitement	F(2,115) = 0.76
Grandiosity	F(2,113) = 1.52
Suspiciousness/Persecution	F(2,116) = 0.54
Hostility	F(2,117) = 0.84
PANSS Negative Subscale	
Blunted Affect	F(2,115) = 0.84
Emotional Withdrawal	F(2,116) = 1.03
Poor Rapport	F(2,116) = 0.15
Passive/Apathetic Social Withdrawal	F(2,116) = 1.07
Difficulty in Abstract Thinking	F(2,115) = 0.32
Lack of Spontaneity	F(2,115) = 0.16
Stereotyped Thinking	F(2,107) = 1.66
PANSS General Psychopathology Subscale	
Somatic Concern	F(2,114) = 3.29*
Anxiety	F(2,117) = 0.06
Guilt Feelings	F(2,112) = 0.21
Tension	F(2,116) = 0.47
Mannerisms & Posturing	F(2,117) = 0.04
Depression	F(2,116) = 0.64
Motor Retardation	F(2,117) = 3.03
Uncooperativeness	F(2,117) = 0.02
Unusual Thought Content	F(2,113) = 2.29
Disorientation	F(2,116) = 0.59
Poor Attention	F(2,114) = 0.19
Lack of Judgment & Insight	F(2,112) = 0.72
Disturbance of Volition	F(2,114) = 1.09
Poor Impulse Control	F(2,114) = 1.34
Preoccupation	F(2,113) = 1.63
Active Social Avoidance	F(2,115) = 2.68

Note. * = $p < .05$, PANSS = Positive and Negative Syndrome Scale; WAIS-R FSIQ = Wechsler Adult Intelligence Scale – Revised Full Scale IQ; WRAT-R = Wide Range Achievement Test – Revised.

Potential Moderating/Confounding Variables

Given that sex, education and IQ were significantly different between the groups (see above), subsequent analyses were conducted to examine whether these variables demonstrated moderating or confounding effects.

IQ

Although both WRAT-R Reading and WAIS-R FSIQ differed significantly between the groups, they were not included in the following analyses assessing potential moderating/confounding variables (refer to the IQ section from Study 1 above for justification).

Potential Moderating Variables

In order to examine the potential moderating effects of sex and education, higher-order three-way interactions between group, family membership and each variable were computed separately. Two CPCAs with hierarchical regression using one higher level of complexity were conducted in order to examine the potential three-way interactions of group and family membership with sex and education. Age was not included as it did not differ significantly between the groups. These analyses were performed in the following steps: (1) calculating the portion of variance in the overall data that was predictable from the main effects of group, family membership, and sex or education; (2) calculating the portions of variance in the residual variance from (1) that was predictable from the two-way interactions between group and family membership, group and sex or education, and family membership and sex or education hierarchically (the residuals from the first two-way interaction were used to examine the second two-way interaction, and the residuals from the second two-way interaction were used to examine the final two-way interaction); and (3) calculating the portion of variance from the

preceding set of residuals in (2) (i.e., with the main effects and two-way interactions partialled out) that were predictable from the three-way interaction between group, family membership, and sex or education. Each interaction term was first regressed onto a matrix combining the variables that were included in the three-way interaction (e.g., the matrix reflecting the interaction between group, family membership and sex was regressed onto a matrix that consisted of the group, family membership and sex variables side-by-side) as was done for Study 2 (see equation 8 in the "Partialling Out the Interaction" section of the Appendix) in order to ensure that the interactions were distinct from the main effects. The percentages of variance accounted for by the combination of, and interaction between, group, family membership and each potential confounding variable are listed in Table 17. The interactions between group, family membership and sex or education accounted for less than 1.5% of the overall variance, and it was concluded that they would not provide any additional information over and above the main effects of, and the interaction between, group and family membership. Therefore, no further analyses were conducted.

Table 17. External Sources of Variance Accounted for by the Interactions Between Group, Family Membership, and Sex or Education for Study 2.

Source	Variance	% Overall
Overall	22.00	100%
Study 2: Three-way Interactions between Group, Family, and Sex		
[Sex, Group, Family]	13.08	59.48%
Group X Family	5.65	25.66%
Sex X Group	0.00	0.00%
Sex X Family	0.32	1.48%
Sex X Group X Family	0.22	1.02%
Study 2: Three-way Interactions between Group, Family, and Education		
[Education, Group, Family]	13.21	60.03%
Group X Family	5.50	25.00%
Education X Group	0.00	0.00%
Education X Family	0.30	1.38%
Education X Group X Family	0.16	0.73%

Potential Confounding Variables

In order to examine whether sex and education were associated with cognitive performance in addition to differing between the groups, these variables were correlated with performance on an aggregate score of the five cognitive measures with the highest component loadings on the cognitive domains resulting from: (1) the portion of variance attributable to the main effect of group (1 component); (2) the portion of variance attributable to the main effect of family membership (2 components); and (3) the portion of variance attributable to the interaction between group and family membership over and above the main effects (3 components). These correlations revealed that females outperformed males on: (1) the general cognitive ability component from the main effect of group, $r = 0.166$, $p < .005$, $\eta^2 = 0.03$; (2) the Verbal Memory component from the main effect of family membership, $r = 0.17$, $p < .005$, $\eta^2 = 0.03$; and (3) the Visual Memory, $r = 0.11$, $p < .05$, $\eta^2 = 0.01$, and Verbal Memory, $r = 0.11$, $p < .05$, $\eta^2 = 0.01$, components from the interaction between group and family membership. In addition, participants with a higher level of education outperformed those with less education on all components: (1) general cognitive ability from the main effect of group, $r = 0.28$, $p < .001$, $\eta^2 = 0.08$; (2) Verbal and Non-Verbal Memory from the main effect of family membership, $r = 0.29$, $p < .001$, $\eta^2 = 0.08$, and $r = 0.22$, $p < .001$, $\eta^2 = 0.05$, respectively; and (3) the Working Memory/Attention, $r = 0.15$, $p < .01$, $\eta^2 = 0.02$, Visual Memory, $r = 0.23$, $p < .001$, $\eta^2 = 0.05$, and Verbal Memory, $r = 0.25$, $p < .001$, $\eta^2 = 0.06$, components from the interaction between group and family membership. Given that sex and education showed associations with both group and the cognitive measures in the same directions, these variables may have contributed to the group differences found in the above CPCA with hierarchical regression.

As described above for Study 1, using ANCOVA to control for potential confounds is not suitable when dealing with variables that code for group (and, in the current study, family membership) because it removes potentially meaningful variance from group (Miller & Chapman, 2001). Therefore, we examined the potential confounding effects of sex and education by partialling these variables out of the dependent variables (i.e., cognitive measures) only. By doing so, we were able to control for potential confounding effects while maintaining the integrity of the independent variables (group and family membership). This analysis proceeded in identical fashion to the original CPCA with hierarchical regression above, except that sex and education were first partialled out of the set of dependent variables. This was achieved through multivariate multiple regression, with the 22 cognitive measures as the dependent variables, and sex and education (combined in a single matrix) as the independent variables. The residual scores were then used as the new dependent variables for the CPCA.

Partialling out sex and education led to a slight reduction in the amounts of variance accounted for by the main effect of group (12.76% to 12.56%), and no change for the amounts of variance accounted for by the main effect of family membership and the overlap between group and family membership (see Tables 10 vs. 18). Because these changes led the residual variance (which was used as the new dependent variable for the interaction step) to account for a greater portion of the overall variance, the interaction between group and family membership accounted for slightly more than in the original analysis (from 27.60% to 27.83%), with no change to the residual variance, which consists of the portion of variance not accounted for by the main effects of, or interaction between, group and family membership. These slight changes in variance did not affect any of the underlying component structures: the components from the overall solution were Working Memory/Attention, Visual Memory, Verbal Memory, Fluency, and Digit Span;

the component resulting from the main effect still reflected general cognitive ability, with highest contributions from WAIS-R Digit Symbol and WMS-R Logical Memory; the two components resulting from the main effect of family membership remained Verbal and Non-Verbal Memory; and the three components resulting from the interaction between group and family were Working Memory/Attention, Verbal Memory, and Visual Memory. Note that the order of the final three components (from the interaction between group and family membership) is changed from the original analysis, with Verbal Memory (3.63%) accounting for more of the overall variance than Visual Memory (3.60%). This change was caused by Verbal Memory accounting for a greater portion of variance than in the original analysis (3.54%), with Visual Memory remaining the same. Importantly, the components themselves were unchanged. These results suggest that although sex and education are associated with the components resulting from the main effects of, and interaction between, group and family membership, the results described above exist over and above any effects of sex and education.

Table 18. Study 2: Variance (Cell Values in Regular Font) and Percentage of Variance (Cell Values in Italics) Accounted for by the Constrained Principal Component Analysis with Confounds Removed

	External	Internal					All Comps
		1	2	3	4	5	
Overall	22.00	3.43	3.18	3.13	2.19	1.77	13.70
% Overall	*100.00*	*15.59*	*14.45*	*14.24*	*9.94*	*8.04*	*62.27*
[Group Family]	12.96	3.72	3.52	-	-	-	7.24
% [Group Family]	*100.00*	*28.73*	*27.17*	-	-	-	*55.89*
% Overall	*58.89*	*16.92*	*16.00*	-	-	-	*32.92*
ME Group	2.76	2.76	-	-	-	-	2.76
% ME Group	*100.00*	*100.00*	-	-	-	-	*100.00*
% Overall	*12.56*	*12.56*	-	-	-	-	*12.56*
ME Family	10.18	2.40	2.18	-	-	-	4.58
% ME Family	*100.00*	*23.56*	*21.43*	-	-	-	*44.99*
% Overall	*46.26*	*10.90*	*9.92*	-	-	-	*20.81*
ME Group/Family	0.02	0.02	-	-	-	-	0.02
% Group/Family	*100.00*	*100.00*	-	-	-	-	*100.00*
% Overall	*0.07*	*0.07*	-	-	-	-	*0.07*
INT Group x Family	6.12	1.10	0.80	0.79	-	-	2.69
% INT Group x Family	*100.00*	*17.97*	*13.04*	*12.94*	-	-	*43.95*
% Overall	*27.83*	*5.00*	*3.63*	*3.60*	-	-	*12.23*
Residual	2.92	0.67	-	-	-	-	0.67
% Residual	*100.00*	*22.96*	-	-	-	-	*22.96*
% Overall	*13.28*	*3.05*	-	-	-	-	*3.05*

Note. ME = Main Effect; INT = Interaction Effect. The variance accounted for by the external analysis and each component extracted in the internal analysis is listed in rows labelled in regular font. The percentages of variance accounted for by the external analysis and each component extracted in the internal analysis are listed in rows labelled in italic font. *% Overall*: percentage of overall variance attributable to the source identified in each column. *% [Group Family]*: percentage of combined predictable variance attributable to the source identified in each column. *% ME Group:* percentage of independent group variance attributable to the source identified in each column. *% ME Family:* Percentage of independent family variance attributable to the source identified in each column. *% ME Group/Family:* Percentage of overlapping variance attributable to the source identified in each column. *% INT Group x Family:* Percentage of group x family interaction variance attributable to the source identified in each column. *% Residual*: percentage of residual variance attributable to the source identified in each column. Values can be computed by dividing the appropriate variance values listed in regular font. All internal analyses were separately rotated using varimax with Kaiser normalization. Order of components corresponds to the magnitude of variance explained.

Discussion

Constrained Principal Component Analysis

With regard to the second aim of the present research, we investigated the cognitive domains that accounted for family membership-dependent and family membership-independent group differences on cognitive performance between schizophrenia patients and their unaffected siblings. In order to achieve this, a CPCA with hierarchical regression was used to examine the underlying component structures of cognitive performance (using difference-from-control scores) for the main effects of group and family membership, as well as for the interaction between group and family membership. The results identified family membership-independent group differences on cognitive performance, which resembled the cognitive domain that emerged from the main effect of group from Study 1; namely, a generalized cognitive deficit in schizophrenia patients, also apparent to a lesser degree in siblings, and was common to all families. Family membership-dependent group differences were characterized by three cognitive domains (Working Memory/Attention, Visual Memory, and Verbal Memory) on which performance differed across families, and together these domains accounted for over a quarter of the overall variance in cognitive performance. A cluster analysis identified three large family subgroups, two of which showed familial cognitive impairment patterns, in verbal memory for cluster 1 and in visual memory for cluster 2. However, it was not possible to determine whether these differences were uniquely the result of family-specific genotypes due to the nature of the family membership variable included in the current study. These results suggest that (1) family membership-independent group differences in cognitive performance can be described as impairment in a generalized cognitive domain in schizophrenia patients and, to a lesser degree, their siblings, and that (2) family membership-dependent group differences in Working

Memory/Attention, Visual Memory, and Verbal Memory are also present, and account for a large portion of the overall variance in cognitive performance. Future research would benefit from a more rigorous measure of family membership in order to isolate potential environmental and genetic influences and to identify potential genetic contributions to cognitive impairment in schizophrenia.

The results from the main effect of group independent of family membership showed striking similarities to the results from Study 1, which examined the cognitive domains that accounted for group differences among schizophrenia patients, siblings, and healthy controls. Specifically, a single domain emerged that accounted for all of the variance predictable by the main effect of group (see Table 10) and showed strong contributions from WAIS-R Digit Symbol and WMS-R Logical Memory immediate and delayed subscales. Given that this analysis was run on a subsample of the sample included in Study 1, these results might be expected. However, it is important to emphasize that the group variable in the current analysis reflected the main effect of group *independent* of family membership; therefore, this finding extends the results from Study 1 by highlighting that the general cognitive impairment that characterizes these group differences exists over and above differences attributable to family membership.

The two components resulting from the main effect of family membership reflect the cognitive domains that are best predicted by family membership independent of group, namely, Verbal Memory and Non-Verbal Memory. These two components reflected combinations of the components from the overall solution (i.e., components 3, 4, and 5 for Verbal Memory, and components 1 and 2 for Non-Verbal Memory; see Table 11), and accounted for a large portion of the overall variance. Both the immediate and delayed subscales of the WMS-R Logical Memory test showed strong contributions to the single domain accounting for the main effect of group as

well as to the first (and largest) domain accounting for the main effect of family membership, suggesting that these variables are strong predictors of performance dependent on both group and family membership. In contrast, WAIS-R Digit Symbol did not show a substantial contribution to either of the cognitive domains resulting from the main effect of family membership. This, coupled with its high contribution on the generalized cognitive ability domain resulting from the main effect of group, suggests that measures of information processing speed (or at least WAIS-R Digit Symbol) are much stronger predictors of group differences than of differences between families. Therefore, whereas WAIS-R Digit Symbol might be a particularly good measure of the general cognitive deficit that characterizes schizophrenia patients and, to a lesser extent, their unaffected siblings, WMS-R Logical Memory appears to capture differences that depend on both group and family membership.

It should be noted that the overall solution in Study 2 was not identical to that from the previous study reported here (Study 1), and comparisons between the two sets of results should, therefore, be made with caution. Although the sample for Study 2 consisted of a subsample of Study 1, the sample characteristics may have been modified based on the method by which it was selected. First, all of the control participants were removed and the remaining scores were transformed into difference-from-control scores (by subtracting them from the mean score of controls on their respective measures). It was necessary to exclude healthy controls from this analysis as they were not part of a family and would, therefore, have no values on the family membership variable. The transformation of patients' and siblings' scores into difference scores was carried out in order to allow for comparisons to be made to the control group and, thus, interpret the results in terms of relative impairment or lack thereof. Second, families that included a schizophrenia patient with no siblings, or families consisting of only siblings (likely

due to the exclusion criteria described above), were not included in the subsample for Study 2. This was necessary to preserve the integrity of the family membership variable, and to ensure that group differences within families could be accurately interpreted for the results emerging from the interaction between group and family membership. While it is likely that these changes produced the differences between the overall solutions in studies 1 and 2, the two component structures remained highly similar (see Tables 4 and 11 for comparison), with three of the five components essentially reflecting the same cognitive domain. Moreover, the component emerging from the main effect of group in Study 2 also showed strong similarities to the component that reflected the underlying structure of group differences in Study 1, as was described above.

Interaction Between Group and Family Membership

The three components that emerged from the interaction between group and family membership reflected specific cognitive domains that were similar to those seen in the overall solution (i.e., Working Memory/Attention, Visual Memory, and Verbal Memory), as well as to common cognitive domains identified in the literature (Dickinson et al., 2010; Nuechterlein et al., 2004). In support of the notion that cognitive impairment in more specific domains are dependent on family membership, we found that some families showed the expected pattern of performance, with schizophrenia patients performing substantially worse than, and siblings scoring closer to, controls (i.e., closer to 0, given that the scores used were difference-from-control scores). However, for other families, schizophrenia patients' performance on the measures that comprised these cognitive domains was left relatively intact. These results are in line with previous research finding familial impairment in specific cognitive domains (e.g., attention; Egan et al., 2000), and suggest that, while diagnosis-dependent group differences in

cognitive performance that are common to all families appear to be due to a general cognitive impairment in patients and, to a lesser extent, their siblings, diagnosis-dependent familial impairment in more specific cognitive domains is also present.

Although the three cognitive domains that emerged from the interaction between group and family membership were highly similar to those with the same names from the overall solution, there were some noticeable differences. This partially reflects the variance removed from the main effects of group and family membership, but might also signify cognitive domains that are more complex than those generally assessed by standardized cognitive tests. For example, the Working Memory/Attention domain in the interaction solution included stronger contributions from Letter and Category Fluency, Digit Span Backwards, and WAIS-R Arithmetic than was the case in the overall solution. Noticeably, the Fluency and Digit Span domains present in the overall solution did not emerge to form distinct domains in any of the subsequent portions of variance, likely because the variables that comprised those domains were subsumed under the Verbal Memory domain resulting from the main effect of group, and under the Working Memory/Attention domain resulting from the interaction.

In order to identify subgroups of families who demonstrated different patterns of performance across these three cognitive domains, we conducted a cluster analysis on the predictor loadings resulting from the CPCA on the variance predicted from the interaction between group and family membership. Of the three cognitive domains identified in the interaction between group and family membership, Verbal Memory and Visual Memory showed some characteristics that might identify them as a potential candidate for a cognitive endophenotype of schizophrenia. As described by Gottesman and Gould (2003), putative endophenotypes should be associated with the disorder in question, heritable, state-independent,

co-segregate with the disorder within families, and be more prevalent in first-degree relatives than within the normal population. Both patients and siblings from family cluster 1 in the current study showed impairment on verbal memory relative to controls, and the same pattern was not observable in the other family profiles. Similarly, familial impairment in visual memory was observable in family cluster 2, a pattern that was not present in the other family clusters. A wealth of literature already supports the notion that both schizophrenia patients and their first-degree relatives are impaired in these domains (Heinrichs & Zakzanis, 1998; Sitskoorn et al., 2004). There is also evidence of the heritability of related cognitive processes (Greenwood et al., 2007), and that impairment is state-independent in schizophrenia patients (Mesholam-Gately et al., 2009). Although it was not possible in the current study to determine whether these cognitive domains reflected family-specific genotypes, the significance of the current study with regards to genetic research on endophenotypes is that the three cognitive domains identified here are specifically related to the interaction between diagnosis and family membership, rather than simply describing impairment characteristic of schizophrenia. Given the challenge of finding reliable evidence for and clearly replicating genetic links to cognitive endophenotypes in schizophrenia (Sullivan, 2005), it is possible that identifying such cognitive endophenotypes through the method used in the current study (i.e., emerging from the portion of variance in cognitive performance that is specifically related to the interaction between diagnosis and family membership) could lead to more focused and potentially more fruitful investigations into their underlying genetics.

It is important to note that the cognitive domains identified from portions of variance including family membership in the present study (i.e., the main effect of family membership and the interaction between group and family membership) include both genetic and

environmental influences. It is likely that some of the family membership-dependent group differences observed in the current study were influenced by factors other than genetic variability. Although genetic examination of the putative endophenotypes identified here was beyond the scope of the present study, previous research has supported the heritability of, and identified potential gene associations with, several cognitive endophenotypes in schizophrenia, some of which are similar to those identified in the current study (e.g., degraded-stimulus CPT and letter-number span for Working Memory/Attention; spatial memory for Visual Memory; and California Verbal Learning Test for Verbal Memory; Greenwood et al., 2007; Greenwood et al., 2013). Twin and adoption studies also indicate that schizophrenia is more strongly influenced by genetics than by environment, and that the most influential environmental factors occur in utero (Sullivan, 2005), which would likely vary between siblings of the same families. Although the influence of genetic factors on cognitive performance in schizophrenia has been well-documented, future behavioural investigations into the familial nature of cognitive deficits in schizophrenia would benefit from including a more rigorous measure than simple family membership.

Overall, the results from Study 2 suggest that (1) a single cognitive domain that is common to all families accounts for group differences in cognitive performance between schizophrenia patients and their unaffected siblings; (2) family membership-dependent group differences in Working Memory/Attention, Visual Memory and Verbal Memory are also present and account for a large portion of variance in cognitive performance; and that (3) some of these family membership-dependent group differences may reflect distinct cognitive endophenotypes in schizophrenia. These findings highlight the importance of examining both family membership-independent and family membership-independent group differences in cognitive performance in

schizophrenia, and emphasize that both genetic and environmental factors influence family membership-dependent group differences that must be taken into account for future investigations into the cognitive endophenotypes of schizophrenia.

Conclusions

The goals of the present research were two-fold: (1) to examine whether diagnosis-dependent group differences in cognitive performance among schizophrenia patients, their unaffected siblings and healthy controls are fundamentally the result of a general cognitive impairment and/or of domain-specific deficits in schizophrenia; and (2) to examine the cognitive domains that characterize family membership-dependent and family membership-independent group differences in cognitive performance between schizophrenia patients and their unaffected siblings.

Study 1 demonstrated that diagnosis-dependent group differences in cognitive performance were due to impairment in a generalized cognitive domain in schizophrenia patients and, to a lesser extent, their siblings. This general cognitive deficit was most strongly related to indices of speed of information processing (e.g., WAIS-R Digit Symbol) and verbal memory (e.g., WMS-R Logical Memory), which generally consist of measures that recruit a wide range of cognitive processes. Study 1 also highlighted the importance of being cognizant of the research questions that a particular statistical analysis technique is designed to address, since the two techniques showcased here led to seemingly conflicting results. Different techniques that do not address the same research questions may lead to opposite conclusions if they are assumed to be testing the same hypothesis. The finding of diagnosis-dependent group differences being characterized by a general cognitive impairment was supported in Study 2, in which we investigated the cognitive domains that accounted for the main effect of group, family membership, as well as the interaction between the two. Specifically, the main effect of group independent of family membership reflected a highly similar generalized cognitive domain, with strong contributions from indices of information processing speed and verbal memory. Given that this effect was

independent of the effect of family membership, these results suggest that diagnosis-dependent group differences in cognitive performance that are common to all families are the result of a fundamental impairment in a generalized cognitive domain in schizophrenia. Study 2 also indicated that the more specific cognitive domains on which patients showed impaired performance were dependent on family membership, such that, for some families, schizophrenia patients demonstrated intact performance on one or more cognitive domains reflecting working memory/attention, visual memory, or verbal memory.

While a general cognitive deficit appears to explain diagnosis-dependent group differences in cognitive performance common to all families, diagnosis-dependent familial impairment on more specific cognitive domains is also present. Although these family membership dependent group differences were influenced by both genetic and environmental factors in the current study, they reflected differences in performance that were specifically related to the interaction between group and family membership. This method, in combination with a more rigorous measure of family membership, has the potential to identify multiple cognitive endophenotypes in schizophrenia that may reflect distinct genetic profiles, and could be a fruitful avenue of future research in terms of identifying the factors that underlie differences in cognitive performance among families.

References

Aleman, A., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (1999). Memory impairment in schizophrenia: A meta-analysis. *American Journal of Psychiatry, 156*, 1358-1366.

Bachman, P., Reichenberg, A., Rice, P., Woolsey, M., Chaves, O., Martinez, D., et al. (2010). Deconstructing processing speed deficits in schizophrenia: Application of a parametric digit symbol coding test. *Schizophrenia Research, 118*(1-3), 6-11.

Benton, A., & Hamsher, D. (1989). *Multilingual aphasia examination*. Iowa City, IA: AJA Associates.

Bokat, C. E., & Goldberg, T. E. (2003). Letter and category fluency in schizophrenic patients: A meta-analysis. *Schizophrenia Research, 64*(1), 73-78.

Cannon, T. D., Huttunen, M. O., Lonnqvist, J., Tuulio-Henriksson, A., Pirkola, T., Glahn, D., et al. (2000). The Inheritance of Neuropsychological Dysfunction in Twins Discordant for Schizophrenia. *American Journal of Human Genetics, 67*(2), 369-382.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and kg criteria for determining the number of factors. *Multivariate Behavioral Research, 12*(3), 289-325.

Chkonia, E., Roinishvili, M., Herzog, M. H., & Brand, A. (2010). First-order relatives of schizophrenic patients are not impaired in the Continuous Performance Test. *Journal of Clinical & Experimental Neuropsychology, 32*(5), 481-486.

Davis, A. S., & Pierson, E. E. (2012). The relationship between the WAIS-III digit symbol coding and executive functioning. *Applied Neuropsychology: Adult, 19*(3), 192-197.

Dickinson, D. (2008). Digit symbol coding and general cognitive ability in schizophrenia: Worth another look? *British Journal of Psychiatry, 193*(5), 354-356.

Dickinson, D., Goldberg, T. E., Gold, J. M., Elvevag, B., & Weinberger, D. R. (2010). Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. *Schizophrenia Bulletin*.

Dickinson, D., Iannone, V. N., Wilk, C. M., & Gold, J. M. (2004). General and specific cognitive deficits in schizophrenia. *Biological Psychiatry, 55*(8), 826-833.

Dickinson, D., Ragland, J. D., Gold, J. M., & Gur, R. C. (2008). General and specific cognitive deficits in schizophrenia: Goliath defeats David? *Biological Psychiatry, 64*(9), 823-827.

Dickinson, D., Ramsey, M. E., & Gold, J. M. (2007). Overlooking the obvious: A meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Archives of General Psychiatry, 64*(5), 532-542.

Egan, M. F., Goldberg, T. E., Gscheidle, T., Weirich, M., Bigelow, L. B., & Weinberger, D. (2000). Relative risks of attention deficits in siblings of patients with schizophrenia. *American Journal of Psychiatry, 157*(8), 1309-1316.

Egan, M. F., Goldberg, T. E., Gscheidle, T., Weirich, M., Rawlings, R., Hyde, T. M., et al. (2001). Relative risk for cognitive impairments in siblings of patients with schizophrenia. *Biological Psychiatry, 50*(2), 98-107.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., et al. (2001). Effect of COMT Val[sup 108/158] Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America, 98*(12), 6917.

Emsley, R., Rabinowitz, J., & Torreman, M. (2003). The factor structure for the Positive and Negative Syndrome Scale (PANSS) in recent-onset psychosis. *Schizophrenia Research, 61*(1), 47-57.

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology Review, 15*(2), 73-95.

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1996a). *Structured Clinical Interview for DSM-IV Axis I Disorders Research Version (SCID-I)*. New York: Biometrics Research, New York State Psychiatric Institute.

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1996b). *Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II)*. New York: Biometrics Research, New York Psychiatric Institute.

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: A meta-analysis. *Psychological Medicine, 39*(6), 889-905.

Genderson, M. R., Dickinson, D., Diaz-Asper, C. M., Egan, M. F., Weinberger, D. R., & Goldberg, T. E. (2007). Factor analysis of neurocognitive tests in a large sample of schizophrenic probands, their siblings, and healthy controls. *Schizophrenia Research, 94*(1-3), 231-239.

Gladsjo, J. A., McAdams, L. A., Palmer, B. W., Moore, D. J., Jeste, D. V., & Heaton, R. K. (2004). A six-factor model of cognition in schizophrenia and related psychotic disorders: Relationships with clinical symptoms and functional capacity. *Schizophrenia Bulletin, 30*(4), 739-754.

Glahn, D. C., Almasy, L., Blangero, J., Burk, G. M., Estrada, J., Peralta, J. M., et al. (2007). Adjudicating neurocognitive endophenotypes for schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 144B*(2), 242-249.

Gold, J. M., & Dickinson, D. (2013). "Generalized Cognitive Deficit" in Schizophrenia: Overused or Underappreciated? *Schizophrenia Bulletin, 39*(2), 263-265.

González-Blanch, C., Pérez-Iglesias, R., Rodríguez-Sánchez, J. M., Pardo-García, G., Martínez-García, O., Vázquez-Barquero, J. L., et al. (2011). A digit symbol coding task as a screening instrument for cognitive impairment in first-episode psychosis. *Archives of Clinical Neuropsychology, 26*(1), 48-58.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.

Gordon, M. (1983). *Instruction manual for the Gordon Diagnostic System*. DeWitt, NY: Gordon Systems.

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *The American Journal Of Psychiatry, 160*(4), 636-645.

Green, M. F., Horan, W. P., & Sugar, C. A. (2013). Has the Generalized Deficit Become the Generalized Criticism? *Schizophrenia Bulletin, 39*(2), 257-262.

Green, M. F., Kern, R. S., Braff, D. L., & Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: Are we measuring the 'right stuff'? *Schizophrenia Bulletin, 26*(1), 119-136.

Greenwood, T. A., Braff, D. L., Light, G. A., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., et al. (2007). Initial heritability analyses of endophenotypic measures for schizophrenia: The consortium on the genetics of schizophrenia. *Archives of General Psychiatry, 64*(11), 1242-1250.

Greenwood, T. A., Swerdlow, N. R., Gur, R. E., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., et al. (2013). Genome-wide linkage analyses of 12 endophenotypes for schizophrenia from the consortium on the genetics of schizophrenia. *The American Journal Of Psychiatry, 170*(5), 521-532.

Heaton, R. (1981). *Wisconsin Card Sorting Test, Manual*. Odessa, FL: Psychological Assessment Resources.

Heinrichs, W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology, 12*, 426-445.

Hunter, M. A., & Takane, Y. (1998). CPCA: A program for principal component analysis with external information on subjects and variables. *Behavior Research Methods, 30*(3), 506-516.

Hunter, M. A., & Takane, Y. (2002). Constrained principal component analysis: Various applications. *Journal of Educational and Behavioral Statistics, 27*(2), 105-145.

Jastak, S., & Wilkinson, G. (1984). *The Wide Range Achievement Test-Revised.* Wilmington, Del.: Jastak Associates.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Heidelberg and New York: Springer.

Joy, S., Fein, D., & Kaplan, E. (2003). Decoding digit symbol: Speed, memory, and visual scanning. *Assessment, 10*(1), 56-65.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276.

Kay, S. R., Opler, L. A., & Lindenmayer, J. P. (1989). The Positive and Negative Syndrome Scale (PANSS): Rationale and Standardization. *British Journal of Psychiatry, 155*(suppl. 7), 59-65.

Keefe, R. S. E., Bilder, R. M., Harvey, P. D., Davis, S. M., Palmer, B. W., Gold, J. M., et al. (2006). Baseline neurocognitive deficits in the CATIE schizophrenia trial. *Neuropsychopharmacology, 31*(9), 2033-2046.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352-358.

Knowles, E. E. M., David, A. S., & Reichenberg, A. (2010). Processing speed deficits in schizophrenia: Reexamining the evidence. *American Journal of Psychiatry, 167*(7), 828-835.

Lançon, C., Auquier, P., Nayt, G., & Reine, G. (2000). Stability of the five-factor structure of the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research, 42*(3), 231-239.

Lavigne, K. M., Hofman, S., Ring, A. J., Ryder, A. G., & Woodward, T. S. (2013). The personality of meaning in life: Associations between dimensions of life meaning and the Big Five. *The Journal of Positive Psychology, 8*(1), 34-43.

Laws, K. R. (1999). A meta-analytic review of Wisconsin Card Sort studies in schizophrenia: General intellectual deficit in disguise? *Cognitive Neuropsychiatry, 4*, 1-35.

Lezak, M. D. (1995). *Neuropsychological assessment (3rd ed.)*. New York, NY US: Oxford University Press.

McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J., & Mueser, K. T. (2007). A meta-analysis of cognitive remediation in schizophrenia. *American Journal of Psychiatry, 164*(12), 1791-1802.

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: A meta-analytic review. *Neuropsychology, 23*, 315-336.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40.

Mohamed, S., Paulsen, J. S., O'Leary, D., Arndt, S., & Andreasen, N. (1999). Generalized cognitive deficits in schizophrenia: A study of first-episode patients. *Archives of General Psychiatry, 56*(8), 749-754.

Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., & Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research, 72*(1), 29-39.

O'Driscoll, G. A., & Callahan, B. L. (2008). Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. *Brain & Cognition, 68*(3), 359-370.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Tuscon, AZ: Neuropsychology Press.

Sitskoorn, M. M., Aleman, A., Ebisch, S. J. H., Appels, M. C. M., & Kahn, R. S. (2004). Cognitive deficits in relatives of patients with schizophrenia: A meta-analysis. *Schizophrenia Research, 71*(2/3), 285-295.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York, NY US: Oxford University Press.

Sullivan, P. F. (2005). The genetics of schizophrenia. *PLoS Medicine, 2*(7), 614-618.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education, Inc.

Takane, Y., & Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Applicable Algebra in Engineering, Communication and Computing, 12*, 391-419.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika, 56*(1), 97-120.

Tuulio-Henriksson, A., Haukka, J., Partonen, T., Varilo, T., Paunio, T., Ekelund, J., et al. (2002). Heritability and number of quantitative trait loci of neurocognitive functions in families with schizophrenia. *American Journal of Medical Genetics, 114*(5), 483-490.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. San Antonio: Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Weickert, T. W., Goldberg, T. E., Gold, J. M., Bigelow, L. B., Egan, M. F., & Weinberger, D. R. (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Archives of General Psychiatry*, *57*, 907-913.

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, 168*(5), 472-485.

Appendices

Mathematical Operations

Principal Component Analysis

PCA is a variable reduction technique used to identify the underlying structure of a set of variables by forming a smaller number of latent variables (components) that reflect the most possible variance in the original dataset (Jolliffe, 2002). These components are created from a weighted sum of the included variables based on their patterns of intercorrelation, and are ordered by the amount of variance they explain in the original dataset, such that the first component accounts for the highest amount of variance in the overall data, the second component accounts for the second highest amount of variance, and so on. Although the total number of components could mathematically equal the number of variables input into the analysis, the first few generally capture a substantial portion of the total variance. As such, the original dataset can be expressed by a smaller number of latent variables with minimal information loss, which allows for the examination of underlying structure while still maintaining the integrity of the data.

One of the most common methods of performing a PCA is through singular value decomposition (SVD). Given a mean-centered matrix *X* of *n* rows by *m* columns, where *n* is equal to the number of participants and *m* is equal to the number of variables, the SVD proceeds in the following manner:

$$ {}_{\mathrm{n}}X_{\mathrm{m}} = {}_{\mathrm{n}}U_{\mathrm{n}} * {}_{\mathrm{n}}D_{\mathrm{m}} * {}_{\mathrm{m}}V_{\mathrm{m}}^{\mathrm{T}} \quad (1) $$

where *U* is an *n* by *n* matrix that contains the eigenvectors of the *n* by *n* covariance matrix of *X* ($\mathrm{cov}(X) = X^{\mathrm{T}} * X$), *D* is an *n* by *m* diagonal matrix whose elements are the square roots of the eigenvalues (λ) of the corresponding eigenvectors in *U* (such that $\sum\lambda = \mathrm{SS_X}$, where SS is the sum of squares), and *V* is an *m* by *m* loading matrix that serves to transform the variables to and from component and variable space.

The percentage of variance accounted for in *X* by each component can be calculated with the following formula:

$$\mathrm{VarC_i} = \left(\frac{\lambda_\mathrm{i}}{\sum\lambda\mathrm{r}}\right)/(\mathrm{n}-1) \tag{2}$$

The component loadings, which identify the contribution of each of the *m* variables to each component, are calculated as follows:

$$\mathrm{CompLoadings} = (V * D)/\sqrt{(\mathrm{n}-1)} \tag{3}$$

which is equivalent to the correlation between *U* and *X*.

Constrained Principal Component Analysis

As was briefly described above, CPCA is a statistical analysis technique that combines multivariate multiple regression and principal component analysis in order to examine the underlying structure of the portion of a dataset that is predictable from a number of variables of interest. Briefly, CPCA involves two steps: the external analysis, in which the variance in a set of dependent variables is separated into that which is predictable from a set of independent variables and that which is not; and the internal analysis, in which the three resulting sources of variance (i.e., overall, predictable, and residual) are submitted to a principal component analysis to derive their underlying structures. CPCA allows for the identification and comparison of the

underlying structures inherent in the unconstrained data, in the portion of the data that is predictable by a set of independent variables of interest, and in the portion of the data that is not predictable by those independent variables. As such, it is possible to examine not only which independent variables predict a set of dependent variables, but also how these variables interrelate and form higher-order components specifically within the portion of variance that they account for.

There are several forms of CPCA, the simplest of which includes a single set of independent variables (as is the case for Study 1 below); however, it is possible to include more than a single set of independent variables and examine the main effects of, and interaction between, the two, as will be demonstrated below for Study 2.

Study 1. One Set of Predictor Variables

The simplest form of the CPCA model, which involves a single set of independent variables, is as follows:

$$ {}_{\mathrm{n}}Z_{\mathrm{m}} = ({}_{\mathrm{n}}G_{\mathrm{v}} * {}_{\mathrm{v}}C_{\mathrm{m}}) + {}_{\mathrm{n}}E_{\mathrm{m}} \quad (4) $$

where *Z* is the criterion data, *G* is a set of independent variables, $C = (G^{\mathrm{T}} * G)^{-1} * G^{\mathrm{T}} * Z$ (where $^{\mathrm{T}}$ indicates the preceeding matrix is transposed and where the inverse is computed as the Moore-Penrose pseudoinverse) are the regression coefficients (i.e., betas) that are applied to *G* to produce the predicted scores (GC), and *E* is the residual data. The criterion data, *Z*, would consist of participants' scores on a set of measures, with one row per participant and one column per measure, leading to a matrix of *n* participants by *m* measures. *G* can include any variables that are expected to relate to *Z* (e.g., participants' scores on other measures of interest or on demographic variables), and would take the form of a matrix of *n* participants by *v* variables. The

external analysis would then be conducted via multivariate multiple regression, thus dividing *Z* into the portion of variance that can be predicted by *G* (*GC*) and that which cannot (*E*).

The internal analysis involves computing a PCA on each of the sources of variance resulting from the external analysis using the method described above. Three PCAs would be used for the simple model described in this section: on the unconstrained criterion data (*Z*); on the variance in the criterion data that is predicted by the set of independent variables (the predicted data, *GC*); and on the variance in the criterion data that is not predicted by the set of independent variables (the residual data, *E*). The component-solutions that emerge (overall, predicted, and residual solutions, respectively) can then be examined to determine which components (or combinations of components) in *Z* are predictable by the independent variables (*G*) and which are not. In order to determine specific associations between the independent variables and those components resulting from the PCA on *GC* (i.e., predictor loadings), correlations are computed between participants' scores on each component (*U* from equation 1) and the independent variables, *G*.

Study 2. Two Sets of Predictor Variables

More complex forms of CPCA can also be implemented. For example, a CPCA with hierarchical regression can be used to investigate the potential interaction between two sets of predictor variables, such as that of group (*GRP*) and family membership (*FAM*) on cognitive performance in schizophrenia patients, their siblings, and healthy controls, as was done in Study 2. In this variation of the simple CPCA described above, the main effects of *GRP* and *FAM* are partialled out of the unconstrained data prior to examining the effect of the interaction between the two. A CPCA of this kind is performed in several steps:

Partialling Out the Main Effects

First, the main effects of *GRP* and *FAM* must be partialled out of the unconstrained data using multivariate multiple regression. In order to ensure the main effects of *GRP* and *FAM* are independent of one another, each is first partialled out of the other. This is achieved through multivariate multiple regression in the following steps:

$$GRP = (FAM * C_{FAM}) + GRP_{.FAM} \quad (5a)$$

where *GRP* is the first set of independent variables, $C_{FAM} = (FAM^{\mathrm{T}} * FAM)^{-1} * FAM^{\mathrm{T}} * GRP$ are the regression coefficients (i.e., betas) that are applied to *FAM* to produce the predicted scores ($FAM{*}C_{FAM}$), and $GRP_{.FAM}$ is the residual data. $FAM{*}C_{FAM}$ reflects the portion of variance in *GRP* that is predictable by *FAM* and $GRP_{.FAM}$ reflects the portion of variance in *GRP* that is independent of *FAM* (i.e., *GRP* with *FAM* partialled out).

Next, the variance in *FAM* is divided into that which is predictable by *GRP* and that which is not:

$$FAM = (GRP * C_{GRP}) + FAM_{.GRP} \quad (5b)$$

where FAM is the second set of independent variables, $C_{GRP} = (GRP^{\mathrm{T}} * GRP)^{-1} * GRP^{\mathrm{T}} * FAM$ are the regression coefficients (i.e., betas) that are applied to *GRP* to produce the predicted scores ($GRP{*}C_{GRP}$), and $FAM_{.GRP}$ is the residual data. $GRP{*}C_{GRP}$ reflects the portion of variance in *FAM* that is predictable by *GRP* and $FAM_{.GRP}$ reflects the portion of variance in *FAM* that is independent of *GRP* (i.e., *FAM* with *GRP* partialled out).

Once *GRP* and *FAM* have been constrained to their independent sources of variance ($GRP_{.FAM}$ and $FAM_{.GRP}$, respectively), these are partialled out of the combined variance in *GRP* and *FAM*:

$$[GRP\ FAM] = (GRP_{.FAM} * C_{GRP.FAM}) + (FAM_{.GRP} * C_{FAM.GRP}) + (GF_{OV}) \quad (6a)$$

where [*GRP FAM*] includes both sets of predictor variables (*GRP* and *FAM*) side-by-side in a single matrix, $GRP_{.FAM}$ is the first independent predictor (from equation 5a), $FAM_{.GRP}$ is the second independent predictor (from equation 5b), and GF_{OV} is the residual variance, which reflects the variance overlapping between *GRP* and *FAM*. $C_{GRP.FAM}$ and $C_{FAM.GRP}$ are the beta weights that, when applied to *GRP* and *FAM*, respectively, produce the predicted scores, ($GRP_{.FAM}$*$C_{GRP.FAM}$) and ($FAM_{.GRP}$*$C_{FAM.GRP}$). These computations are carried out in two steps:

$$[GRP\ FAM] = (GRP_{.FAM} * C_{GRP.FAM}) + E \quad (6b)$$

where [*GRP FAM*] includes both sets of predictor variables (*GRP* and *FAM*) side-by-side in a single matrix, $GRP_{.FAM}$ is the first independent predictor (from equation 5a), $C_{GRP.FAM} = (GRP_{.FAM}^{T} * GRP_{.FAM})^{-1} * GRP_{.FAM}^{T} * [GRP\ FAM]$ are the beta weights that, when applied to $GRP_{.FAM}$ produce the predicted scores, $GRP_{.FAM}C_{GRP.FAM}$, and *E* is the residual data, which reflects the variance in *GRP* and *FAM* combined that is not predictable by the variance in *GRP* that is independent of *FAM*. The portion of variance in [*GRP FAM*] that is predictable by the variance in *FAM* that is independent of group is then separated as follows:

$$E = (FAM_{.GRP} * C_{FAM.GRP}) + GF_{OV} \quad (6c)$$

where *E* is the residual variance from equation 6b, $FAM_{.GRP}$ is the second independent predictor from equation 5b, $C_{FAM.GRP} = (FAM_{.GRP}^{T} * FAM_{.GRP})^{-1} * FAM_{.GRP} * E$ are the beta weights that,

when applied to $FAM_{.GRP}$ produce the predicted scores, $FAM_{.GRP}C_{FAM.GRP}$, and GF_{OV} is the residual variance, which reflects the variance overlapping between *GRP* and *FAM*. The three elements of the model in equation 6a above will be referred to as GRP_{IND}, FAM_{IND}, and GF_{OV}, respectively.

Once computation of the non-overlapping independent variables GRP_{IND}, FAM_{IND}, and GF_{OV} is complete, these matrices are input consecutively in a CPCA with hierarchical regression model:

$$Z = (GRP_{IND} * C_{GRP}) + E_1 \tag{7a}$$

where *Z* is the criterion data, GRP_{IND} is the independent predictor *GRP* ($GRP_{.FAM}C_{GRP.FAM}$ from equation 6b), $C_{GRP} = (GRP_{IND}^T * GRP_{IND})^{-1} * GRP_{IND}^T * Z$ consists of the beta weights that, when applied to GRP_{IND}, produce the predicted scores $GRP_{IND}C_{GRP}$ that reflect the main effect of *GRP* independent of *FAM*, and E_1 is the residuals.

This first step separates the variance accounted for by the main effect of *GRP* independent of *FAM* from the unconstrained variance in *Z*. In order to determine the main effect of *FAM* independent of *GRP*, the residuals in E_1 must be used as the set of criterion variables in the next step:

$$E_1 = (FAM_{IND} * C_{FAM}) + E_2 \tag{7b}$$

where E_1 consists of the residuals from equation 7a, FAM_{IND} is the independent predictor *FAM* ($FAM_{.GRP}C_{FAM.GRP}$ from equation 6c), $C_{FAM} = (FAM_{IND}^T * FAM_{IND})^{-1} * FAM_{IND}^T * E_1$ includes the beta weights that, when applied to FAM_{IND} produce the predicted scores $FAM_{IND}C_{FAM}$ that reflect the main effect of *FAM* independent of *GRP*, and E_2 are the residuals.

This second step separates the variance accounted for by the main effect of *FAM* from the residual variance, E_1 (from equation 7a), which is the overall variance in *Z* with the main effect of *GRP* partialled out. The next step involves partialling out the main effect of the variance that is shared between *GRP* and *FAM* (i.e., GF_{OV} from equation 6c) from the residual variance E_2 (from equation 7b), which is the overall variance in *Z* with the main effects of *GRP* and *FAM* partialled out:

$$E_2 = \left(GF_{OV} * C_{GF_{OV}}\right) + E_3 \tag{7c}$$

where E_2 consists of the residuals from equation 7b, GF_{OV} is the overlapping variance between *GRP* and *FAM* (from equation 6c) with the main effects of *GRP* and *FAM* partialled out, $C_{GF_{OV}} = (GF_{OV}^T * GF_{OV})^{-1} * GF_{OV}^T * E_2$ includes the beta weights that, when applied to GF_{OV} form the predicted scores that reflect the main effect of the overlap between *GRP* and *FAM*, and E_3 are the residuals.

This third step separates the variance shared between *GRP* and *FAM* from the residual variance, E_2, which is the overall variance in *Z* with the main effects of *GRP* and *FAM* partialled out. The matrix E_3 consists of the variance in the overall data that is not predictable from either the main effects of *GRP* and *FAM* or the overlapping variance between *GRP* and *FAM*. Note that regressing *Z* onto [*GRP FAM*] from equation 6a would produce the same residual results as the matrix E_3 that is calculated hierarchically from equations 7a-7c. This is because$[GRP\ FAM] = (GRP_{.FAM} * C_{GRP.FAM}) + (FAM_{.GRP} * C_{FAM.GRP}) + GF_{OV}$, as is shown in equation 6a. It was necessary to proceed in the hierarchical manner illustrated in equations 7a-7c in order to calculate the portions of variance that accounted for the main effects of *GRP*, *FAM*, and the overlap, rather than simply the portion of variance that accounted for all three combined.

Partialling Out the Interaction

Once the main effects have been partialled out of the overall variance in *Z*, it is possible to do a similar analysis to determine the portion of variance that can be explained by the interaction between *GRP* and *FAM* over and above their main effects. Before doing so, however, it is important to partial these main effects out of the interaction term, which is simply the product of *GRP* and *FAM*. This is achieved using the following equation:

$$(GRP * FAM) = \left([GRP\ FAM] * C_{[GRP\ FAM]}\right) + GF_{INT} \tag{8}$$

where (*GRP * FAM*) is the product of *GRP* and *FAM*, [*GRP FAM*] is a matrix including both *GRP* and *FAM* side-by-side (as in equations 6a and 6b), $C_{[GRP\ FAM]} = ([GRP\ FAM]^{\mathrm{T}} * [GRP\ FAM])^{-1} * [GRP\ FAM] * (GRP * FAM)$ are the beta weights that, when applied to [*GRP FAM*], produce the predicted scores [*GRP FAM*]$C_{[GRP\ FAM]}$, and GF_{INT} consists of the residual scores that reflect the interaction between *GRP* and *FAM* with the main effects of *GRP*, *FAM*, and the overlap between *GRP* and *FAM* partialled out.

This step separates the variance accounted for by the main effects of *GRP* and FAM as well as their overlap from the interaction term, (*GRP*FAM*). The matrix GF_{INT} reflects the interaction between *GRP* and *FAM* over and above the main effects, that is, the variance in the interaction term that is not predictable from either the main effects of *GRP* and *FAM* or the overlap between *GRP* and *FAM*. This interaction can then be examined using the following model:

$$E_3 = \left(GF_{INT} * C_{GF\,INT}\right) + E_4 \tag{9}$$

where E_3 are the residual scores from equation 7c, GF_{INT} is the interaction term independent of the main effects (from equation 8), $C_{GF\,INT} = (GF_{INT}^{T} * GF_{INT})^{-1} * GF_{INT}^{T} * E_3$ are the beta weights that, when applied to GF_{INT}, produce the predicted scores $GF_{INT}C_{GFINT}$ that reflect the interaction effect, and E_4 consists of the residuals that are not predictable by either the main effects of *GRP* and *FAM*, their overlap, or the interaction between the two variables.

Principal Component Analyses on Unconstrained Data, Main Effects, and Interaction

Once the main effects and interaction have been statistically separated, PCAs are computed on each as was done above for the simplest CPCA model, in the manner described in the Principal Component Analysis section above. In this hierarchical model, PCAs are computed on: the overall data (*Z*) – the results of which will be identical to those in a simple CPCA model with the same criterion data; the main effects of *GRP*, *FAM* and their overlap [(GRP_{IND}*C_{GRP}), (FAM_{IND}*C_{FAM}), and (GF_{OV}*C_{GFOV}), from equations 7a, 7b, and 7c, respectively]; the interaction between *GRP* and *FAM* (GF_{INT}*C_{GFINT} from equation 9); and the residuals (E_4 from equation 9). Predictor loadings (i.e., correlations between the partialled independent variables - $GRP_{.FAM}$, $FAM_{.GRP}$, GF_{OV}, and GF_{INT} from equations 5a, 5b, and 8, respectively – and their respective component scores) are also computed to relate the predicted scores from the main effects and interaction effect back to the independent variables.

Study 2: Component Score Analysis

In order to further investigate the meaning of the interaction between group and family membership, we examined the component scores of those families who demonstrated the ten highest positive and negative predictor loadings (total = 20 families) for each of the three components (Working Memory/Attention: positive families = 178, 185, 195, 292, 411, 467, 528,

648, 649, 744; negative families = 144, 147, 152, 271, 289, 295, 400, 478, 590, 694; Visual Memory: positive families = 198, 210, 424, 519, 527, 535, 589, 656, 730, 814; negative families = 140, 155, 166, 184, 195, 221, 230, 336, 376, 775; Verbal Memory: positive families = 139, 162, 169, 191, 195, 233, 235, 298, 375, 472; negative families = 129, 135, 185, 208, 210, 224, 265, 404, 579, 699; see Table 20). Appendix Figures 8-10 illustrate the component scores for the patients and siblings of these families for Working Memory/Attention, Visual Memory, and Verbal Memory, respectively. Taken together with Table 20, one can see that for those families with positive predictor loadings (e.g., families 178, 185, 195, 292, 411, 467, 528, 648, 649, and 744 for Working Memory/Attention), siblings had positive and patients had negative component scores. In contrast, for those families with negative predictor loadings (e.g., families 144, 147, 152, 271, 289, 295, 400, 478, 590, and 694 for Working Memory/Attention), siblings had negative and patients had positive component scores. This pattern of positive predictor loadings leading to positive component scores for siblings and negative component scores for patients (and the inverse for negative predictor loadings) is due to schizophrenia patients being coded as -1 and siblings being coded as 1 in the matrix reflecting the interaction between group and family membership. These different patterns were related back to performance by submitting the predictor loadings for each component for each family to a cluster analysis (see the Cluster Analysis section in the main text).

Table 19. Study 2: Predictor Loadings for the Main Effect of Family Membership Independent of Group (Positive Predictor Loadings = High Performance; Negative Predictor Loadings = Low Performance)

Variables	*Verbal Memory*	*Variables*	*Non-Verbal Memory*
Family 138	0.22	Family 152	-0.18
Family 289	-0.20	Family 190	-0.18
Family 210	0.19	Family 570	0.17
Family 130	0.19	Family 217	-0.17
Family 143	0.18	Family 140	-0.16
Family 329	-0.17	Family 325	0.16
Family 391	-0.17	Family 400	-0.16
Family 207	0.16	Family 345	0.16
Family 318	0.15	Family 775	0.15
Family 179	-0.14	Family 189	-0.15
Family 262	0.14	Family 795	0.15
Family 478	-0.14	Family 579	0.14
Family 579	-0.14	Family 373	0.14
Family 265	0.14	Family 179	-0.14
Family 313	0.13	Family 823	0.13
Family 158	0.13	Family 360	-0.13
Family 202	0.13	Family 699	0.13
Family 238	0.13	Family 418	-0.13
Family 290	-0.13	Family 212	-0.12
Family 528	-0.12	Family 401	0.12
Family 159	-0.12	Family 527	-0.12
Family 519	-0.12	Family 210	-0.12
Family 319	-0.12	Family 351	0.12
Family 435	-0.12	Family 132	-0.12
Family 726	-0.11	Family 423	0.11
Family 541	-0.11	Family 221	0.11
Family 388	-0.11	Family 541	0.11
Family 211	0.11	Family 441	0.11
Family 489	-0.11	Family 499	0.11
Family 305	-0.11	Family 135	-0.11
Family 336	-0.11	Family 295	0.11
Family 314	-0.10	Family 829	0.11
Family 198	0.10	Family 160	-0.11
Family 448	0.10	Family 606	0.11
Family 383	0.10	Family 262	0.11
Family 184	0.10	Family 226	0.11

Variables	*Verbal Memory*	*Variables*	*Non-Verbal Memory*
Family 169	0.10	Family 680	0.11
Family 227	0.09	Family 264	-0.11
Family 583	-0.09	Family 233	-0.10
Family 590	0.09	Family 194	-0.10
Family 775	-0.09	Family 403	-0.10
Family 180	0.09	Family 136	-0.10
Family 197	0.09	Family 298	0.10
Family 212	-0.09	Family 359	0.10
Family 150	0.09	Family 339	-0.10
Family 466	-0.09	Family 215	-0.10
Family 359	-0.09	Family 350	0.10
Family 190	0.09	Family 519	0.09
Family 187	0.09	Family 694	-0.09
Family 175	-0.08	Family 535	0.09
Family 271	0.08	Family 191	-0.09
Family 594	-0.08	Family 448	0.08
Family 135	0.08	Family 336	-0.08
Family 274	0.08	Family 540	-0.08
Family 606	0.08	Family 230	-0.08
Family 730	-0.07	Family 138	0.08
Family 680	-0.07	Family 231	0.07
Family 286	-0.07	Family 424	0.07
Family 136	0.07	Family 162	-0.07
Family 540	-0.07	Family 235	-0.07
Family 220	0.07	Family 153	0.07
Family 350	-0.07	Family 814	-0.07
Family 215	0.07	Family 161	-0.07
Family 779	-0.07	Family 147	-0.07
Family 129	-0.07	Family 144	-0.07
Family 404	-0.07	Family 273	0.07
Family 177	0.07	Family 139	-0.07
Family 411	-0.06	Family 388	-0.06
Family 337	-0.06	Family 392	0.06
Family 392	0.06	Family 375	0.06
Family 296	0.06	Family 173	-0.06
Family 376	-0.06	Family 710	0.06
Family 386	0.06	Family 228	0.06
Family 194	-0.06	Family 726	-0.06
Family 257	-0.06	Family 159	-0.06
Family 375	-0.06	Family 466	-0.06

Variables	*Verbal Memory*	*Variables*	*Non-Verbal Memory*
Family 518	0.05	Family 199	-0.06
Family 400	-0.05	Family 185	-0.06
Family 465	0.05	Family 289	-0.06
Family 226	0.05	Family 478	0.06
Family 140	-0.05	Family 197	0.06
Family 554	-0.05	Family 383	0.05
Family 527	-0.05	Family 354	-0.05
Family 139	0.05	Family 465	0.05
Family 161	-0.05	Family 313	-0.05
Family 151	0.05	Family 472	0.05
Family 814	-0.05	Family 288	0.05
Family 292	-0.05	Family 296	-0.05
Family 643	-0.05	Family 583	0.05
Family 293	-0.04	Family 238	-0.05
Family 403	-0.04	Family 730	0.05
Family 224	0.04	Family 589	-0.05
Family 185	-0.04	Family 286	-0.05
Family 418	-0.04	Family 193	0.05
Family 339	-0.04	Family 489	0.04
Family 589	-0.04	Family 195	-0.04
Family 233	0.04	Family 656	0.04
Family 379	-0.04	Family 314	-0.04
Family 472	-0.04	Family 169	-0.04
Family 829	0.04	Family 292	0.04
Family 141	-0.04	Family 204	0.04
Family 320	-0.04	Family 467	0.04
Family 189	-0.04	Family 779	-0.04
Family 351	-0.04	Family 205	-0.04
Family 178	0.03	Family 224	-0.04
Family 656	-0.03	Family 320	0.03
Family 699	-0.03	Family 220	-0.03
Family 204	0.03	Family 528	0.03
Family 688	0.03	Family 283	0.03
Family 354	-0.03	Family 202	-0.03
Family 373	-0.03	Family 329	0.03
Family 199	0.03	Family 518	-0.03
Family 412	-0.03	Family 412	0.03
Family 280	-0.03	Family 305	-0.03
Family 217	0.03	Family 391	0.03
Family 162	0.03	Family 274	0.03

Variables	*Verbal Memory*	*Variables*	*Non-Verbal Memory*
Family 535	0.03	Family 376	-0.03
Family 195	-0.03	Family 155	-0.03
Family 360	-0.03	Family 744	0.03
Family 242	-0.03	Family 257	-0.03
Family 166	0.02	Family 590	0.03
Family 264	0.02	Family 386	-0.03
Family 160	-0.02	Family 334	0.03
Family 401	-0.02	Family 269	0.03
Family 441	-0.02	Family 242	0.02
Family 155	0.02	Family 688	0.02
Family 295	0.02	Family 537	-0.02
Family 144	0.02	Family 177	-0.02
Family 499	-0.02	Family 594	-0.02
Family 283	-0.02	Family 554	-0.02
Family 424	-0.02	Family 127	-0.02
Family 273	0.02	Family 293	-0.02
Family 147	-0.02	Family 271	-0.02
Family 325	-0.02	Family 187	-0.02
Family 235	0.02	Family 158	-0.02
Family 230	0.02	Family 208	0.02
Family 710	0.02	Family 175	0.02
Family 153	0.01	Family 319	0.02
Family 795	0.01	Family 318	0.01
Family 570	-0.01	Family 211	0.01
Family 191	-0.01	Family 265	0.01
Family 648	-0.01	Family 290	0.01
Family 208	0.01	Family 151	0.01
Family 127	-0.01	Family 150	0.01
Family 228	0.01	Family 180	-0.01
Family 649	-0.01	Family 166	-0.01
Family 231	-0.01	Family 141	-0.01
Family 823	0.01	Family 649	0.01
Family 269	-0.01	Family 404	0.01
Family 188	-0.01	Family 198	0.01
Family 334	0.01	Family 188	-0.01
Family 298	-0.01	Family 130	0.01
Family 193	0.01	Family 643	-0.01
Family 694	0.01	Family 435	-0.01
Family 152	-0.01	Family 184	-0.01
Family 423	0.01	Family 337	0.01

Variables	*Verbal Memory*	*Variables*	*Non-Verbal Memory*
Family 537	0.00	Family 379	-0.01
Family 345	0.00	Family 129	0.01
Family 467	0.00	Family 227	0.00
Family 288	0.00	Family 143	0.00
Family 744	0.00	Family 207	0.00
Family 132	0.00	Family 280	0.00
Family 173	0.00	Family 648	0.00
Family 221	0.00	Family 178	0.00
Family 205	0.00	Family 411	0.00

Table 20. Study 2: Predictor Loadings for the Effect of the Interaction Between Group and Family Membership (Positive Predictor Loadings = Siblings' Component Scores > Patients; Negative Predictor Loadings = Patients' Component Scores > Siblings)

Variables	*Working Memory/Attention*	*Variables*	*Visual Memory*	*Variables*	*Verbal Memory*
Family 178	0.20	Family 519	0.18	Family 139	0.18
Family 694	-0.20	Family 195	-0.17	Family 298	0.18
Family 648	0.19	Family 589	0.17	Family 699	-0.17
Family 411	0.19	Family 775	-0.17	Family 233	0.17
Family 144	-0.18	Family 336	-0.16	Family 210	-0.16
Family 295	-0.17	Family 535	0.15	Family 404	-0.16
Family 649	0.17	Family 814	0.14	Family 162	0.16
Family 152	-0.16	Family 656	0.14	Family 472	0.15
Family 400	-0.16	Family 198	0.14	Family 265	-0.14
Family 271	-0.15	Family 527	0.14	Family 375	0.14
Family 147	-0.15	Family 166	-0.13	Family 129	-0.13
Family 289	-0.15	Family 221	-0.13	Family 235	0.13
Family 467	0.14	Family 730	0.13	Family 185	-0.13
Family 528	0.13	Family 376	-0.12	Family 191	0.12
Family 195	0.13	Family 155	-0.12	Family 195	0.12
Family 292	0.13	Family 210	0.12	Family 169	0.11
Family 744	0.13	Family 140	-0.12	Family 814	0.11
Family 185	0.12	Family 424	0.12	Family 224	-0.11
Family 354	0.12	Family 230	-0.11	Family 135	-0.11
Family 478	-0.12	Family 184	-0.11	Family 208	-0.11
Family 590	-0.11	Family 392	0.11	Family 217	0.11
Family 274	0.11	Family 318	-0.11	Family 221	0.11
Family 699	-0.11	Family 158	-0.11	Family 288	0.11
Family 155	-0.11	Family 153	-0.11	Family 579	-0.10
Family 594	0.10	Family 175	-0.11	Family 177	0.10
Family 199	-0.10	Family 290	0.11	Family 161	0.10
Family 424	0.10	Family 579	-0.11	Family 334	-0.10
Family 226	-0.10	Family 141	0.11	Family 656	0.10
Family 418	-0.10	Family 139	0.10	Family 465	0.10
Family 423	-0.10	Family 185	0.10	Family 606	-0.10
Family 296	0.10	Family 441	-0.10	Family 730	0.10
Family 376	-0.09	Family 208	-0.10	Family 130	-0.10
Family 224	-0.09	Family 373	-0.10	Family 242	0.10
Family 293	0.09	Family 205	-0.10	Family 400	-0.10
Family 329	-0.09	Family 401	-0.10	Family 319	0.09

Variables	*Working Memory/Attention*	*Variables*	*Visual Memory*	*Variables*	*Verbal Memory*
Family 298	-0.09	Family 151	0.10	Family 141	0.09
Family 173	0.09	Family 649	0.10	Family 144	-0.09
Family 499	-0.09	Family 143	-0.10	Family 147	-0.09
Family 159	0.09	Family 320	-0.10	Family 313	0.09
Family 392	-0.09	Family 351	0.09	Family 238	0.09
Family 320	0.09	Family 375	0.09	Family 274	-0.09
Family 215	-0.09	Family 386	-0.09	Family 166	-0.09
Family 489	-0.09	Family 412	0.09	Family 350	-0.09
Family 814	-0.09	Family 305	0.09	Family 528	-0.08
Family 656	-0.09	Family 193	0.09	Family 649	-0.08
Family 166	-0.08	Family 215	0.09	Family 360	0.08
Family 262	-0.08	Family 224	0.09	Family 726	-0.08
Family 435	0.08	Family 231	-0.08	Family 262	-0.08
Family 305	0.08	Family 286	-0.08	Family 540	0.08
Family 541	-0.08	Family 448	0.08	Family 231	0.08
Family 290	0.07	Family 169	0.08	Family 392	0.08
Family 375	-0.07	Family 188	-0.08	Family 271	0.08
Family 472	0.07	Family 829	-0.08	Family 403	-0.08
Family 730	0.07	Family 187	0.08	Family 643	0.08
Family 465	-0.07	Family 337	-0.08	Family 227	0.08
Family 175	0.07	Family 162	-0.08	Family 478	-0.08
Family 383	-0.07	Family 345	-0.08	Family 138	-0.08
Family 180	0.07	Family 227	-0.08	Family 710	0.08
Family 269	0.06	Family 202	-0.08	Family 314	-0.08
Family 205	0.06	Family 135	0.08	Family 286	-0.07
Family 314	0.06	Family 478	-0.08	Family 441	0.07
Family 139	-0.06	Family 220	0.08	Family 589	-0.07
Family 351	-0.06	Family 744	0.07	Family 290	-0.07
Family 184	0.06	Family 177	-0.07	Family 694	0.07
Family 162	0.06	Family 325	0.07	Family 158	-0.07
Family 339	0.06	Family 262	-0.07	Family 383	-0.07
Family 179	0.06	Family 313	0.07	Family 197	-0.07
Family 554	0.06	Family 528	0.07	Family 594	0.07
Family 688	-0.06	Family 298	0.07	Family 339	0.07
Family 153	-0.06	Family 271	-0.07	Family 775	0.06
Family 160	-0.06	Family 423	0.07	Family 345	0.06
Family 350	-0.06	Family 314	-0.06	Family 151	-0.06
Family 265	-0.06	Family 296	0.06	Family 140	0.06
Family 518	-0.06	Family 710	0.06	Family 205	0.06

Variables	*Working Memory/Attention*	*Variables*	*Visual Memory*	*Variables*	*Verbal Memory*
Family 337	0.06	Family 403	0.06	Family 194	-0.06
Family 129	0.06	Family 226	0.06	Family 318	-0.06
Family 264	0.05	Family 197	-0.06	Family 537	0.06
Family 193	-0.05	Family 329	-0.06	Family 212	0.06
Family 135	-0.05	Family 160	0.06	Family 829	-0.06
Family 589	0.05	Family 180	-0.06	Family 590	-0.06
Family 570	-0.05	Family 293	0.06	Family 189	-0.06
Family 150	0.05	Family 541	0.06	Family 132	0.06
Family 132	0.05	Family 295	-0.06	Family 292	0.06
Family 325	-0.05	Family 138	-0.06	Family 401	-0.06
Family 379	0.05	Family 354	0.05	Family 155	-0.06
Family 187	-0.05	Family 178	-0.05	Family 411	-0.05
Family 388	-0.05	Family 726	0.05	Family 379	-0.05
Family 207	0.05	Family 540	-0.05	Family 541	0.05
Family 313	0.05	Family 795	-0.05	Family 152	0.05
Family 208	0.05	Family 273	0.05	Family 199	0.05
Family 141	0.05	Family 211	0.05	Family 273	-0.05
Family 795	0.04	Family 238	-0.05	Family 527	0.05
Family 403	0.04	Family 127	0.05	Family 136	0.05
Family 336	0.04	Family 235	-0.05	Family 187	0.05
Family 130	0.04	Family 233	-0.05	Family 202	-0.05
Family 319	0.04	Family 289	-0.05	Family 211	-0.05
Family 466	0.04	Family 537	0.05	Family 173	0.05
Family 197	0.04	Family 383	0.05	Family 648	-0.05
Family 579	-0.04	Family 680	0.04	Family 583	0.04
Family 288	0.04	Family 283	0.04	Family 160	-0.04
Family 194	0.04	Family 359	-0.04	Family 570	-0.04
Family 535	-0.04	Family 274	-0.04	Family 359	-0.04
Family 360	-0.04	Family 472	0.04	Family 373	-0.04
Family 190	-0.04	Family 489	0.04	Family 220	-0.04
Family 227	-0.04	Family 779	0.04	Family 320	-0.04
Family 238	0.04	Family 319	0.04	Family 424	-0.04
Family 158	0.04	Family 379	-0.04	Family 179	-0.04
Family 228	-0.04	Family 194	0.04	Family 296	-0.04
Family 169	0.04	Family 204	-0.04	Family 175	-0.04
Family 345	0.04	Family 643	0.04	Family 228	-0.04
Family 140	-0.03	Family 144	-0.04	Family 376	0.04
Family 221	0.03	Family 465	0.04	Family 744	0.04
Family 412	0.03	Family 694	-0.04	Family 423	0.04

Variables	*Working Memory/Attention*	*Variables*	*Visual Memory*	*Variables*	*Verbal Memory*
Family 138	0.03	Family 292	-0.04	Family 391	-0.03
Family 231	0.03	Family 583	0.03	Family 269	0.03
Family 217	0.03	Family 217	-0.03	Family 193	-0.03
Family 177	-0.03	Family 280	-0.03	Family 127	-0.03
Family 519	-0.03	Family 699	0.03	Family 448	0.03
Family 643	-0.03	Family 518	0.03	Family 535	-0.03
Family 143	0.03	Family 189	0.03	Family 215	-0.03
Family 136	-0.03	Family 688	-0.03	Family 190	0.03
Family 151	0.03	Family 173	0.03	Family 336	-0.03
Family 404	0.03	Family 590	0.03	Family 198	0.03
Family 233	-0.03	Family 130	0.03	Family 143	0.03
Family 220	0.03	Family 228	0.03	Family 280	0.03
Family 286	-0.02	Family 570	-0.03	Family 386	0.03
Family 680	0.02	Family 152	-0.03	Family 325	-0.03
Family 189	0.02	Family 499	0.03	Family 204	-0.03
Family 211	0.02	Family 265	0.02	Family 289	-0.03
Family 726	0.02	Family 147	0.02	Family 489	-0.03
Family 606	-0.02	Family 418	0.02	Family 518	-0.03
Family 280	-0.02	Family 435	-0.02	Family 180	0.03
Family 127	0.02	Family 159	-0.02	Family 554	0.03
Family 273	-0.02	Family 334	-0.02	Family 688	-0.02
Family 257	-0.02	Family 136	-0.02	Family 295	-0.02
Family 583	-0.02	Family 264	-0.02	Family 823	-0.02
Family 775	-0.02	Family 411	0.02	Family 207	-0.02
Family 161	0.01	Family 191	-0.02	Family 305	0.02
Family 391	0.01	Family 129	0.02	Family 779	-0.02
Family 448	-0.01	Family 339	0.02	Family 337	0.02
Family 441	-0.01	Family 400	-0.02	Family 283	-0.02
Family 829	0.01	Family 257	-0.02	Family 230	0.02
Family 318	-0.01	Family 242	-0.02	Family 388	-0.02
Family 823	-0.01	Family 554	0.02	Family 418	0.01
Family 235	-0.01	Family 388	0.02	Family 329	-0.01
Family 359	0.01	Family 190	0.01	Family 435	-0.01
Family 283	-0.01	Family 207	-0.01	Family 153	-0.01
Family 386	0.01	Family 594	0.01	Family 412	-0.01
Family 212	-0.01	Family 150	0.01	Family 188	0.01
Family 188	-0.01	Family 391	0.01	Family 467	-0.01
Family 198	-0.01	Family 269	-0.01	Family 150	-0.01
Family 401	-0.01	Family 179	-0.01	Family 184	0.01

Variables	*Working Memory/Attention*	*Variables*	*Visual Memory*	*Variables*	*Verbal Memory*
Family 210	0.01	Family 212	-0.01	Family 264	0.01
Family 191	-0.01	Family 199	-0.01	Family 159	0.01
Family 373	0.00	Family 360	-0.01	Family 178	0.01
Family 527	0.00	Family 648	-0.01	Family 466	0.01
Family 537	0.00	Family 404	-0.01	Family 293	0.01
Family 710	0.00	Family 606	0.01	Family 351	0.01
Family 334	0.00	Family 161	-0.01	Family 795	0.01
Family 540	0.00	Family 823	0.00	Family 226	0.01
Family 242	0.00	Family 288	0.00	Family 354	0.01
Family 204	0.00	Family 466	0.00	Family 499	-0.01
Family 202	0.00	Family 350	0.00	Family 680	0.00
Family 779	0.00	Family 132	0.00	Family 257	0.00
Family 230	0.00	Family 467	0.00	Family 519	0.00

Figure 9. Study 2: Component Scores for Families with the Ten Highest Positive and Negative Predictor Loadings on Working Memory/Attention from the Interaction Between Group and Family Membership

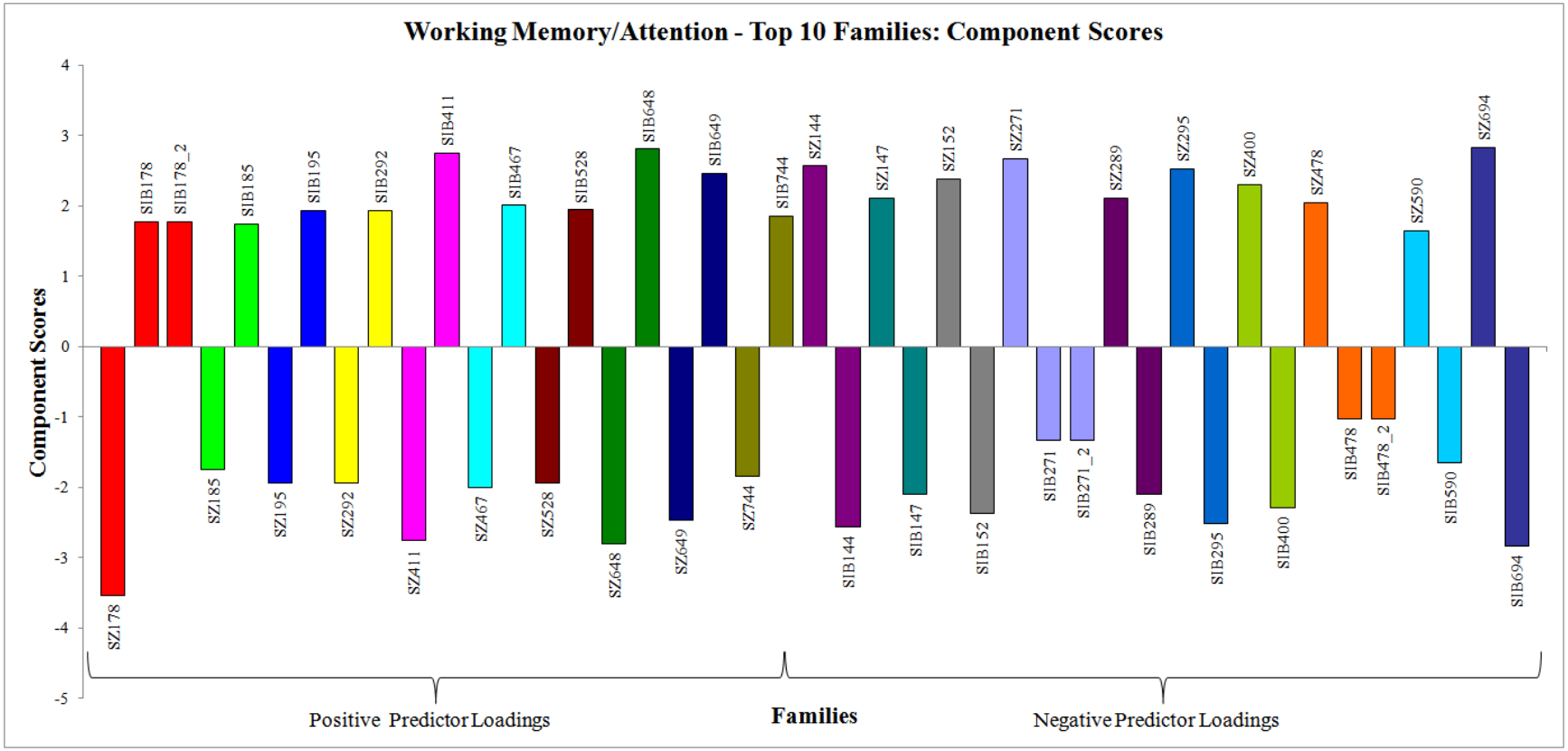


Figure 10. Study 2: Component Scores for Families with the Ten Highest Positive and Negative Predictor Loadings on Visual Memory from the Interaction Between Group and Family Membership

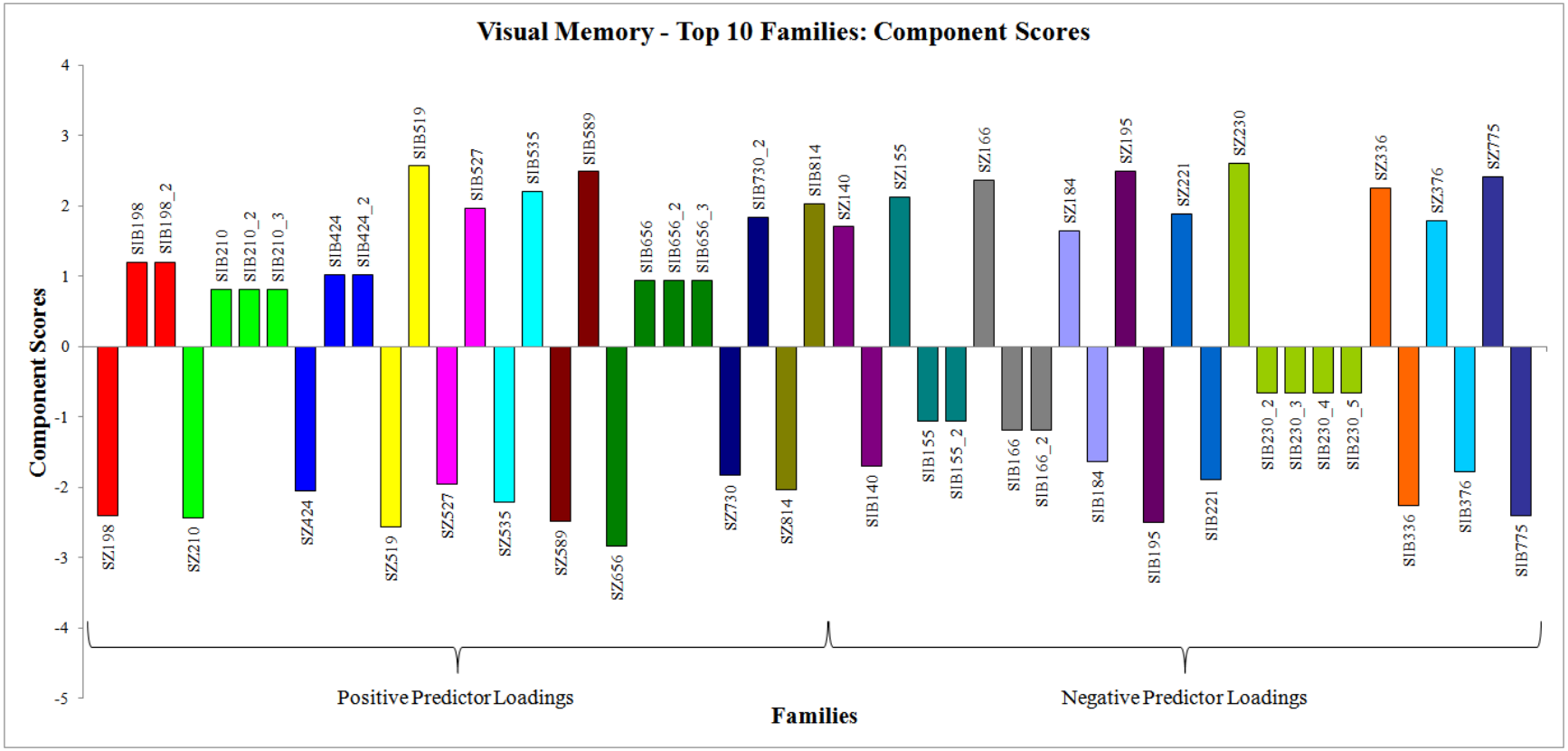


Figure 11. Study 2: Component Scores for Families with the Ten Highest Positive and Negative Predictor Loadings on Verbal Memory from the Interaction Between Group and Family Membership

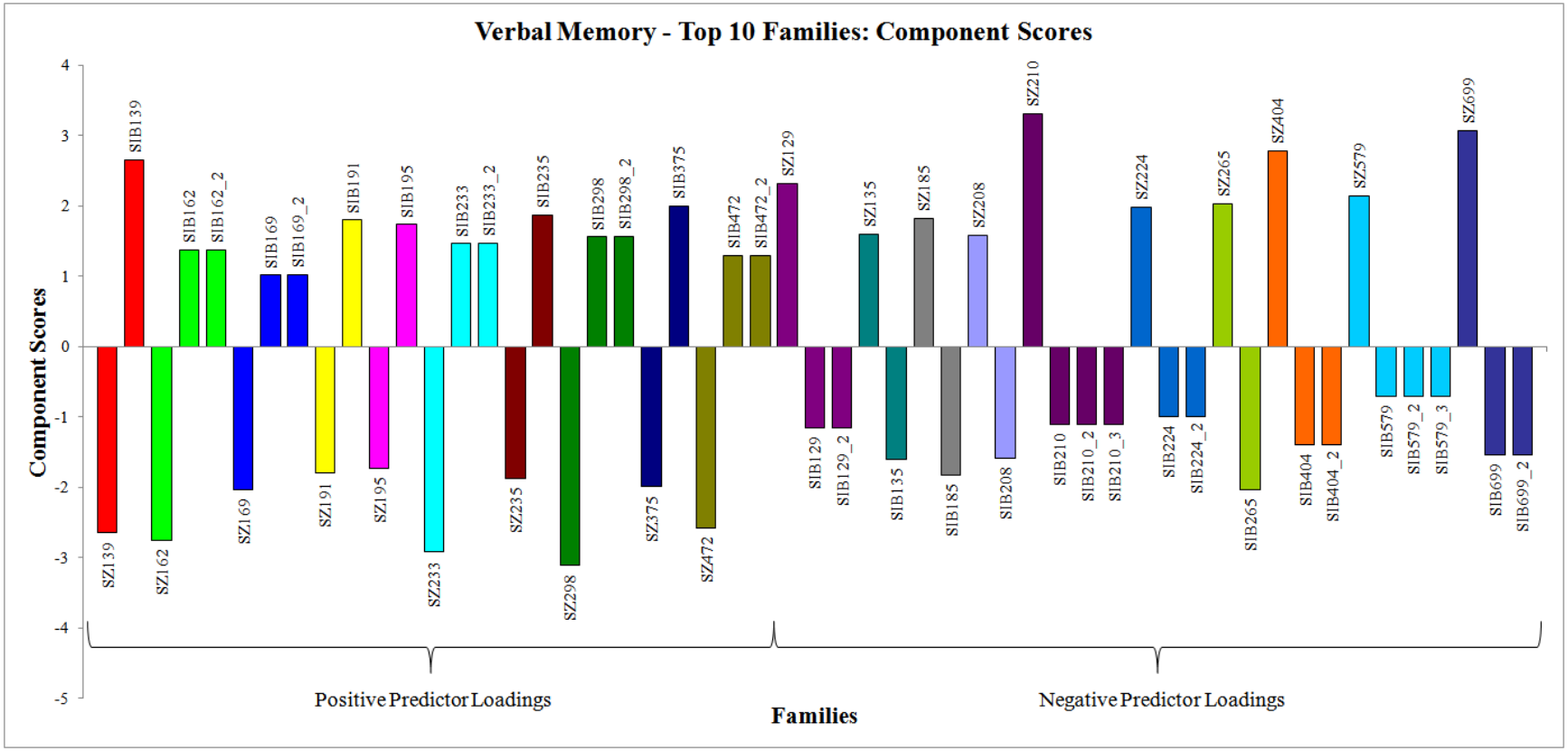
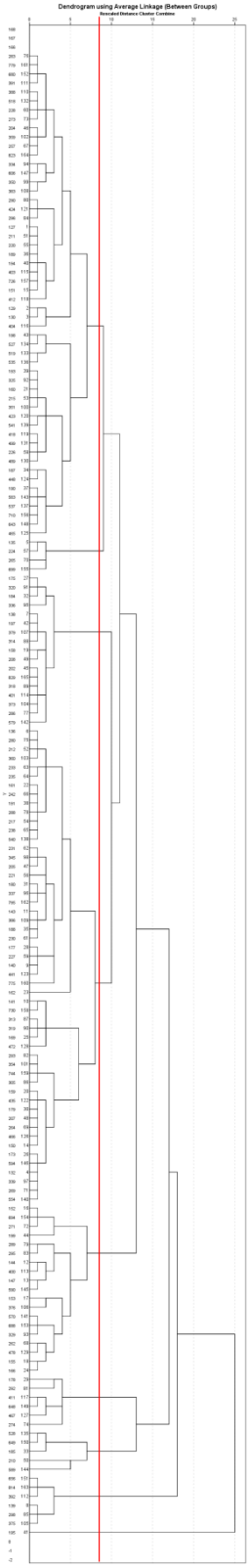


Figure 12. Study 2: Dendrogram for the Cluster Analysis on the Predictor Loadings from the Interaction (Red Line Indicates Cluster Cut-Off Point)



Study 1 – Eigenvalues and Scree Plot Script

```
load Zrank_notpartialled15SIBsout.mat
Z=zscore(Z);

[m n]=size(Z);
[U D V]=svd(Z, 'econ');

%Eigenvalues for Z
EigsZ=diag(D^2./(m-1));
VARtotal=sum(EigsZ) %#ok<NOPTS> %Overall variance
Zscree=plot(EigsZ,'+:');

load G_lincontrasts.mat

%G solution
C=pinv((G'*G))*G'*Z;
GC=G*C;
[p q]=size(GC);
[U_G D_G V_G]=svd(GC, 'econ');

%Eigenvalues for GC
EigsGC=diag(D_G^2./(p-1));
VARtotal_GC=sum(EigsGC) %#ok<NOPTS> %GC variance
EigsGC_new=zeros(size(EigsGC,1),1);
for i=1:size(EigsGC,1)
    temp=EigsGC(i,:)./VARtotal_GC;
    EigsGC_new(i,:)=temp*VARtotal;
end
GCscree=plot(EigsGC_new,'+:');

%Creating E
E=Z-GC;
[r s]=size(E);
[U_E D_E V_E]=svd(E, 'econ');

%Eigenvalues for E
EigsE=diag(D_E^2./(r-1));
VARtotal_E=sum(EigsE) %#ok<NOPTS> %E variance
EigsE_new=zeros(size(EigsE,1),1);
for i=1:size(EigsE,1)
    temp=EigsE(i,:)./VARtotal_E;
    EigsE_new(i,:)=temp*VARtotal;
end
Escree=plot(EigsE_new,'+:');
```

Study 1 – Constrained Principal Component Analysis Script

```
load Zrank_notpartialled15SIBsout.mat
Z=zscore(Z);

[m n]=size(Z);
[U D V]=svd(Z, 'econ');
Znum_comps=input('Please enter the number of components for Z:\n');

%Creating rotated Z loadings using V
A_Z=(V*D)./sqrt(m-1);
A_Z=A_Z(:,1:Znum_comps);
if Znum_comps<=1, Zrotatedsq=A_Z.^2;
else
    [ZrotatedbyV,ZrotatedT]=varimkn_beh(A_Z);
    Zrotatedsq=ZrotatedbyV.^2;
end
%Using U
UofZ=U(:,1:Znum_comps);
UofZ=zscore(UofZ);
if Znum_comps<=1, UofZ_rot=UofZ;
else
    UofZ_rot=UofZ*inv(ZrotatedT');
end

for k=1:size(Z,2)
    for l=1:Znum_comps
    ctmp=corrcoef(Z(:,k),UofZ_rot(:,l) );
    Zloadings(k,l)=ctmp(1,2);
    end
end

VARtotal=sum(diag(D^2./(m-1))) % Overall variance

%Creating GC
load G_lincontrasts.mat
C=pinv((G'*G))*G'*Z;
GC=G*C;

[p q]=size(GC);
[U_G D_G V_G]=svd(GC, 'econ');
GCnum_comps=input('Please enter the number of components for GC:\n');

%Creating rotated G loadings using V
A_G=(V_G*D_G)./sqrt(p-1);
A_G=A_G(:,1:GCnum_comps);
if GCnum_comps<=1
    GCrotatedsq=A_G.^2;
else
    [GCrotatedbyV,GCrotatedT]=varimkn_beh(A_G);
    GCrotatedsq=GCrotatedbyV.^2;
end
%Using U
UofGC=U_G(:,1:GCnum_comps);
```

```
UofGC=zscore(UofGC);
if GCnum_comps<=1
    UofGC_rot=UofGC;
else
    UofGC_rot=UofGC*inv(GCrotatedT');
end

for k=1:size(GC,2)
    for l=1:GCnum_comps
        ctmp=cov(GC(:,k),UofGC_rot(:,l) );
        GCloadings(k,l)=ctmp(1,2);
    end
end

VARtotal_GC=sum(diag(D_G^2./(p-1))) % GC variance

%To get the rotated predictor loadings for GC
for k=1:size(G,2)
    for l=1:GCnum_comps
        ctmp=corrcoef(G(:,k),UofGC_rot(:,l) );
        Gloadings(k,l)=ctmp(1,2);
    end
end

%E solution
E=Z-GC;

[r s]=size(E);
[U_E D_E V_E]=svd(E, 'econ');
Enum_comps=input('Please enter the number of components for E:\n');

%Creating rotated E loadings using V
A_E=(V_E*D_E)./sqrt(r-1);
A_E=A_E(:,1:Enum_comps);
if Enum_comps<=1, Erotatedsq=A_E.^2;
else
    [ErotatedbyV,ErotatedT]=varimkn_beh(A_E);
    Erotatedsq=ErotatedbyV.^2;
end
%Using U
UofE=U_E(:,1:Enum_comps);
UofE=zscore(UofE);
if Enum_comps<=1
    UofE_rot=UofE;
else
    UofE_rot=UofE*inv(ErotatedT');
end
for k=1:size(E,2)
    for l=1:Enum_comps
    ctmp=cov(E(:,k),UofE_rot(:,l) );
    Eloadings(k,l)=ctmp(1,2);
    end
end

VARtotalE=sum(diag(D_E^2./(r-1))); % E variance
```

Study 2 – Eigenvalues and Scree Plot Script (Step 1)

```
%%                                CALCULATING VARIABLES

load Zrank_fams_notpartialled15SIBsout.mat
load G1.mat
load G2.mat
GG=[G1 G2];
Z=zscore(Z);

%G1=G2C+G1ind
C_G2=pinv((G2'*G2))*G2'*G1; %Partialling G2 out of G1
G2C=G2*C_G2;
G1ind=G1-G2C;

%G2=G1C+G2ind
C_G1=pinv((G1'*G1))*G1'*G2; %Partialling G1 out of G2
G1C=G1*C_G1;
G2ind=G2-G1C;

%GG=G1indC+E_GG
C_G1ind=pinv((G1ind'*G1ind))*G1ind'*GG;
G1indC=G1ind*C_G1ind;
E_GG=GG-G1indC;

%E_GG=G2indC+E2_GG
C_G2ind=pinv((G2ind'*G2ind))*G2ind'*E_GG;
G2indC=G2ind*C_G2ind;

%GG=G1ind+G2ind+G1G2overlap
G1G2overlap=GG-G1indC-G2indC;

%%                                      Z SCREE

[m n]=size(Z);
[U D V]=svd(Z, 'econ');

%Eigenvalues for Z
EigsZ=diag(D^2./(m-1));
VARtotal=sum(diag(D^2./(m-1))) % Overall variance
Zscree=plot(EigsZ,'+:');

%%                                  G1 (INDEPENDENT OF G2) SCREE

C_G1ind=pinv((G1indC'*G1indC))*G1indC'*Z;
ZG1indC=G1indC*C_G1ind;
E1=Z-ZG1indC;

[r s]=size(ZG1indC); %#ok<NASGU>
[U_G1ind D_G1ind V_G1ind]=svd(ZG1indC, 'econ');

%Eigenvalues for G1 independent of G2
EigsG1ind=diag(D_G1ind^2./(r-1));
```

```
VARtotal_G1indC=sum(diag(D_G1ind^2./(r-1))) % G1ind variance
EigsG1ind_new=zeros(size(EigsG1ind,1),1);
for i=1:size(EigsG1ind,1)
    temp=EigsG1ind(i,:)./VARtotal_G1indC;
    EigsG1ind_new(i,:)=temp*VARtotal;
end
G1indCscree=plot(EigsG1ind_new,'+:');

%%                              G2 (INDEPENDENT OF G1) SCREE

C_G2ind=pinv((G2indC'*G2indC))*G2indC'*E1;
ZG2indC=G2indC*C_G2ind;
E2=E1-ZG2indC;

[t u]=size(ZG2indC); %#ok<NASGU>
[U_G2ind D_G2ind V_G2ind]=svd(ZG2indC, 'econ');

%Eigenvalues for G2 independent of G1
EigsG2ind=diag(D_G2ind^2./(t-1));
VARtotal_G2indC=sum(diag(D_G2ind^2./(t-1))) % G2ind variance
EigsG2ind_new=zeros(size(EigsG2ind,1),1);
for i=1:size(EigsG2ind,1)
    temp=EigsG2ind(i,:)./VARtotal_G2indC;
    EigsG2ind_new(i,:)=temp*VARtotal;
end
G2indCscree=plot(EigsG2ind_new,'+:');

%%                           REDUNDANCY BETWEEN G1 & G2 SCREE

C_G1G2overlap=pinv((G1G2overlap'*G1G2overlap))*G1G2overlap'*E2;
G1G2overlapC=G1G2overlap*C_G1G2overlap;
E3=E2-G1G2overlapC;
[v w]=size(G1G2overlapC);
[U_G1G2overlap D_G1G2overlap V_G1G2overlap]=svd(G1G2overlapC, 'econ');

%Eigenvalues for G1G2overlapC
EigsG1G2overlap=diag(D_G1G2overlap^2./(v-1));
VARtotal_G1G2overlapC=sum(diag(D_G1G2overlap^2./(v-1))) % G1G2overlap
variance
EigsG1G2overlap_new=zeros(size(EigsG1G2overlap,1),1);
for i=1:size(EigsG1G2overlap,1)
    temp=EigsG1G2overlap(i,:)./VARtotal_G1G2overlapC;
    EigsG1G2overlap_new(i,:)=temp*VARtotal;
end
G1G2overlapCscree=plot(EigsG1G2overlap_new,'+:');
```

Study 2 – Eigenvalues and Scree Plot Script (Step 2)

```
%%                          CALCULATING VARIABLES

load Gint.mat
load G1.mat
load G2.mat
GG=[G1 G2];
C_GGint=pinv((GG'*GG))*GG'*Gint;
GGC_int=GG*C_GGint;
E_Gint=Gint-GGC_int;

%%                         Z SCREE
[m n]=size(Z);
[U D V]=svd(Z, 'econ');

EigsZ=diag(D^2./(m-1));
VARtotal=sum(diag(D^2./(m-1))) % Overall variance


%%                      INTERACTION BETWEEN G1 & G2 SCREE
C_Gint=pinv((E_Gint'*E_Gint))*E_Gint'*E3;
GintC=E_Gint*C_Gint;
E4=E3-GintC;
[p q]=size(GintC);
[U_Gint D_Gint V_Gint]=svd(GintC, 'econ');

%Eigenvalues for GintC
EigsGintC=diag(D_Gint^2./(p-1));
VARtotal_Gint=sum(diag(D_Gint^2./(p-1))) % Gint variance
EigsGint_new=zeros(size(EigsGintC,1),1);
for i=1:size(EigsGintC,1)
    temp=EigsGintC(i,:)./VARtotal_Gint;
    EigsGint_new(i,:)=temp*VARtotal;
end
GintCscree=plot(EigsGint_new,'+:');

%%                                      RESIDUAL SCREE
 [r s]=size(E4);
[U_E4 D_E4 V_E4]=svd(E4, 'econ');

%Eigenvalues for E4
EigsE=diag(D_E4^2./(r-1));
VARtotalE4=sum(diag(D_E4^2./(r-1))) % E4 variance
EigsE_new=zeros(size(EigsE,1),1);
for i=1:size(EigsE,1)
    temp=EigsE(i,:)./VARtotalE4;
    EigsE_new(i,:)=temp*VARtotal;
end
Escree=plot(EigsE_new,'+:');
```

Study 2 – Constrained Principal Component Analysis Script (Step 1)

```
%%                          CALCULATING VARIABLES

load Zrank_fams_notpartialled15SIBsout.mat
load G1.mat
load G2.mat
GG=[G1 G2];
Z=zscore(Z);

%G1=G2C +G1ind
C_G2=pinv((G2'*G2))*G2'*G1;
G2C=G2*C_G2;
G1ind=G1-G2C;
save G1ind G1ind

%G2=G1C +G2ind
C_G1=pinv((G1'*G1))*G1'*G2;
G1C=G1*C_G1;
G2ind=G2-G1C;
save G2ind G2ind

%GG=G1indC+E_GG
C_G1ind=pinv((G1ind'*G1ind))*G1ind'*GG;
G1indC=G1ind*C_G1ind;
E_GG=GG-G1indC;

%E_GG=G2indC+E2_GG
C_G2ind=pinv((G2ind'*G2ind))*G2ind'*E_GG;
G2indC=G2ind*C_G2ind;

%GG=G1ind+G2ind+G1G2overlap
G1G2overlap=GG-G1indC-G2indC;


%%                              Z SOLUTION
%%

[m n]=size(Z);
[U D V]=svd(Z, 'econ');
Znum_comps=input('Please enter the number of components for Z:\n');

%Calculating component loadings
A_Z=(V*D)./sqrt(m-1);
A_Z=A_Z(:,1:Znum_comps);
UofZ=U(:,1:Znum_comps);
UofZ=zscore(UofZ);
if Znum_comps<=1, Zrotatedsq=A_Z.^2;
else
    [ZrotatedbyV,ZrotatedT]=varimkn_beh(A_Z);
    Zrotatedsq=ZrotatedbyV.^2;
end
```

```
if Znum_comps<=1, UofZ_rot=UofZ;
else
    UofZ_rot=UofZ*inv(ZrotatedT');
end

for k=1:size(Z,2)
    for l=1:Znum_comps
        ctmp=corrcoef(Z(:,k),UofZ_rot(:,l) );
        Zloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

VARtotal=sum(diag(D^2./(m-1))) % Overall variance

%%                      G1 (INDEPENDENT OF G2) SOLUTION

C_G1ind=pinv((G1indC'*G1indC))*G1indC'*Z;
ZG1indC=G1indC*C_G1ind;
E1=Z-ZG1indC;

[r s]=size(ZG1indC); %#ok<NASGU>
[U_G1ind D_G1ind V_G1ind]=svd(ZG1indC, 'econ');
G1indCnum_comps=input('Please enter the number of components for G1
independent of G2:\n');

%Calculating component loadings
A_G1ind=(V_G1ind*D_G1ind)./sqrt(r-1);
A_G1ind=A_G1ind(:,1:G1indCnum_comps);
UofG1indC=U_G1ind(:,1:G1indCnum_comps);
UofG1indC=zscore(UofG1indC);
if G1indCnum_comps<=1  G1indCrotatedsq=A_G1ind.^2;
else
    [G1indCrotatedbyV,G1indCrotatedT]=varimkn_beh(A_G1ind);
    G1indCrotatedsq=G1indCrotatedbyV.^2;
end

if G1indCnum_comps<=1  UofG1indC_rot=UofG1indC;
else
    UofG1indC_rot=UofG1indC*inv(G1indCrotatedT');
end
for k=1:size(ZG1indC,2)
    for l=1:G1indCnum_comps
    ctmp=cov(ZG1indC(:,k),UofG1indC_rot(:,l) );
    G1indCloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

VARtotal_G1indC=sum(diag(D_G1ind^2./(r-1))) % G1ind variance

%Calculating predictor loadings
for k=1:size(G1ind,2)
    for l=1:G1indCnum_comps
        ctmp=corrcoef(G1ind(:,k),UofG1indC_rot(:,l) );
        G1indloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
```

```
end

%%                          G2 (INDEPENDENT OF G1) SOLUTION

C_G2ind=pinv((G2indC'*G2indC))*G2indC'*E1;
ZG2indC=G2indC*C_G2ind;
E2=E1-ZG2indC;

[t u]=size(ZG2indC); %#ok<NASGU>
[U_G2ind D_G2ind V_G2ind]=svd(ZG2indC, 'econ');
G2indCnum_comps=input('Please enter the number of components for G2
independent of G1:\n');

%Calculating component loadings
A_G2ind=(V_G2ind*D_G2ind)./sqrt(t-1);
A_G2ind=A_G2ind(:,1:G2indCnum_comps);
UofG2indC=U_G2ind(:,1:G2indCnum_comps);
UofG2indC=zscore(UofG2indC);
if G2indCnum_comps<=1 G2indCrotatedsq=A_G2ind.^2;
else
    [G2indCrotatedbyV,G2indCrotatedT]=varimkn_beh(A_G2ind);
    G2indCrotatedsq=G2indCrotatedbyV.^2;
end

if G2indCnum_comps<=1 UofG2indC_rot=UofG2indC;
else
    UofG2indC_rot=UofG2indC*inv(G2indCrotatedT');
end
for k=1:size(ZG2indC,2)
    for l=1:G2indCnum_comps
        ctmp=cov(ZG2indC(:,k),UofG2indC_rot(:,l) );
        G2indCloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

VARtotal_G2indC=sum(diag(D_G2ind^2./(t-1))) % G2ind variance

%Calculating predictor loadings
for k=1:size(G2ind,2)
    for l=1:G2indCnum_comps
        ctmp=corrcoef(G2ind(:,k),UofG2indC_rot(:,l) );
        G2indloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end
```

Study 2 – Constrained Principal Component Analysis Script (Step 2)

```
%%                                CALCULATING VARIABLES

load Gint.mat
load G1.mat
load G2.mat
GG=[G1 G2];

C_GGint=pinv((GG'*GG))*GG'*Gint;
GGC_int=GG*C_GGint;
E_Gint=Gint-GGC_int;

cd Matrices
load E3
cd ..
load Zrank_fams_notpartialled15SIBsout.mat
Z=zscore(Z);

[m n]=size(Z);
[U D V]=svd(Z, 'econ');
VARtotal=sum(diag(D^2./(m-1))); % Overall variance

%%                          GINTERACTION SOLUTION
%%

%Ginteraction solution
C_Gint=pinv((E_Gint'*E_Gint))*E_Gint'*E3;
GintC=E_Gint*C_Gint;
E4=E3-GintC;

[p q]=size(GintC);
[U_Gint D_Gint V_Gint]=svd(GintC, 'econ');
GintCnum_comps=input('Please enter the number of components for GintC:\n');

%Calculating component loadings
A_Gint=(V_Gint*D_Gint)./sqrt(p-1);
A_Gint=A_Gint(:,1:GintCnum_comps);
UofGintC=U_Gint(:,1:GintCnum_comps);
UofGintC=zscore(UofGintC);
if GintCnum_comps<=1 GintCrotatedsq=A_Gint.^2;
else
    [GintCrotatedbyV,GintCrotatedT]=varimkn_beh(A_Gint);
    GintCrotatedsq=GintCrotatedbyV.^2;
end

if GintCnum_comps<=1
    UofGintC_rot=UofGintC;
else
    UofGintC_rot=UofGintC*inv(GintCrotatedT');
end
for k=1:size(GintC,2)
    for l=1:GintCnum_comps
        ctmp=cov(GintC(:,k),UofGintC_rot(:,l) );
```

```
        GintCloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

VARtotal_GintC=sum(diag(D_Gint^2./(p-1))) % Gint variance

%To get the rotated predictor loadings for GintC
for k=1:size(Gint,2)
    for l=1:GintCnum_comps
        ctmp=corrcoef(Gint(:,k),UofGintC_rot(:,l) );
        Gintloadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

%%                                  E4 SOLUTION

[r s]=size(E4);
[U_E4 D_E4 V_E4]=svd(E4, 'econ');
E4num_comps=input('Please enter the number of components for E4:\n');

%Creating rotated E4 loadings using V
A_E4=(V_E4*D_E4)./sqrt(r-1);
A_E4=A_E4(:,1:E4num_comps);
UofE4=U_E4(:,1:E4num_comps);
UofE4=zscore(UofE4);
if E4num_comps<=1, E4rotatedsq=A_E4.^2;
else
    [E4rotatedbyV,E4rotatedT]=varimkn_beh(A_E4);
    E4rotatedsq=E4rotatedbyV.^2;
end

if E4num_comps<=1
    UofE4_rot=UofE4;
else
    UofE4_rot=UofE4*inv(E4rotatedT');
end
for k=1:size(E4,2)
    for l=1:E4num_comps
        ctmp=cov(E4(:,k),UofE4_rot(:,l) );
        E4loadings(k,l)=ctmp(1,2); %#ok<AGROW>
    end
end

VARtotalE4=sum(diag(D_E4^2./(r-1))) % E4 variance
```