2018-12-07

Neuroimaging of Cognitive Flexibility: Task Validity and Clinical Applications

Dina R. Dajani
University of Miami, d.dajani92@gmail.com

UNIVERSITY OF MIAMI

NEUROIMAGING OF COGNITIVE FLEXIBILITY: TASK VALIDITY AND CLINICAL APPLICATIONS

By

Dina R. Dajani

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2018

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

NEUROIMAGING OF COGNITIVE FLEXIBILITY: TASK VALIDITY AND CLINICAL APPLICATIONS

Dina R. Dajani

Approved:

_____
Lucina Q. Uddin, Ph.D.
Associate Professor of Psychology

_____
Jennifer C. Britton, Ph.D.
Associate Professor of Psychology

_____
Maria M. Llabre, Ph.D.
Professor of Psychology

_____
Aaron S. Heller, Ph.D.
Assistant Professor of Psychology

_____
Anthony S. Dick, Ph.D.
Associate Professor of Psychology
Florida International University

_____
Guillermo Prado, Ph.D.
Dean of the Graduate School

DAJANI, DINA R. (Ph.D., Psychology)
Neuroimaging of Cognitive Flexibility: Task Validity and Clinical Applications. (December 2018)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Associate Professor Lucina Q. Uddin.
No. of pages in text. (99)

A key challenge for cognitive neuroscience is the development and validation of tasks that capture real-world cognitive phenomena and render them suitable for investigation within an MRI environment. Cognitive flexibility is necessary for a range of adaptive behaviors and is associated with optimal life outcomes, but as a psychological construct it has been difficult to operationalize and validate. Crucially, psychometrically validated tasks must be developed prior to their application to clinical populations to investigate the neural substrates of cognitive *in*flexibility. Here, we address these limitations by adapting a well-validated laboratory measure of cognitive flexibility to the scanner environment (Study I) and applied the neuroimaging results to characterize directed functional connectome profiles supportive of cognitive flexibility in children (Study II). First, the neural substrates of cognitive flexibility in a sample of 32 neurotypical adults (19-46 years) were characterized using task-based functional magnetic resonance imaging (fMRI). Results demonstrated that the fMRI-adapted task is reliable and showed convergent validity with the laboratory-based version of the task, which has previously been shown to measure cognitive flexibility. In line with our hypotheses, we observed activation in prefrontal, posterior parietal, insular, basal ganglia and thalamic regions in response to engaging cognitive flexibility, over and above low-level visual and motor processes. In Study II, heterogeneity in cognitive flexibility in children with a range of

abilities (children with autism spectrum disorder [ASD], attention deficit/hyperactivity disorder [ADHD], and typically developing [TD] children) was parsed using directed functional connectivity profiles derived from resting-state fMRI data. Brain regions identified in Study I were used to guide region-of-interest (ROI) selection to estimate individual connectivity profiles in Study II. We expected to find at least three subgroups of children who differed in their network connectivity metrics and symptom measures. Unexpectedly, we did not find a stable or valid subgrouping solution, which suggests that categorical models of the neural substrates of cognitive flexibility in children may be invalid. Together, the results highlight the neurotypical correlates of cognitive flexibility in both children and adults within the frontoparietal and salience networks, and shed light on the validity of conceptualizing the neural substrates of cognitive flexibility categorically in children. Ultimately, this work may provide a foundation for the development of a revised nosology focused on neurobiological substrates of mental illness as an alternative to traditional symptom-based classification systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

Cognitive flexibility is a core component of executive functions which allows individuals to flexibly adapt to changes in environmental demands in the context of goal-directed behavior (Scott 1962). These skills emerge in early childhood and follow a protracted development through young adulthood (Anderson 2002). Intact cognitive flexibility is important across the lifespan, as these skills support social development and academic achievement in childhood and job success in adulthood (Bailey 2007, Diamond, Barnett et al. 2007, Hunter and Sparrow 2012, Chen, Yang et al. 2014, Engel de Abreu, Abreu et al. 2014). Although cognitive flexibility has received considerable attention in the psychological literature, at the level of observable behavior and self- and other-reports, the large-scale neural networks that support cognitive flexibility are still under investigation. Further, scant research has identified functional network alterations that may underlie difficulties in implementing flexible behavior (Dajani and Uddin 2015). To fill this gap, this dissertation sought to combine basic cognitive neuroscience approaches and their clinical applications to characterize brain network profiles that support cognitive flexibility skills in the mature brain, in adults, and in the developing brain, in children.

Going beyond behavior to characterize its neural substrates is necessary to determine the underlying large-scale brain network interactions that give rise to cognitive flexibility. This work is crucial in determining whether multiple biological pathways give rise to flexible behavior. There are numerous lines of evidence to suggest that brain-behavior relationships are not simply one-to-one (Pessoa 2014), emphasizing the importance of considering neurobiological mechanisms of behavioral constructs. This is

especially important in consideration of the biological substrates of cognitive *in*flexibility. As an alternative to symptom-based approaches to diagnose mental health disorders, the goal of the Research Domain Criteria (RDoC) is to identify individuals with a common biological pathway to abnormality, relying on neural circuits, instead of behavioral markers, as the substrate of mental illness (Insel, Cuthbert et al. 2010). Focusing on the level of neurobiology instead of behavior is crucial because varying types of functional network miswirings across development may manifest as a singular phenotype (Di Martino, Fair et al. 2014), suggesting that distinct brain abnormalities may appear behaviorally identical. Likewise, disparate genetic etiologies may lead to similar behavioral profiles (Dougherty, Evans et al. 2016, Pelphrey 2017). For example, autism spectrum disorder (ASD) encompasses a heterogeneous group of individuals who exhibit variability in cognitive flexibility (Gioia, Isquith et al. 2002, Blijd-Hoogewys, Bezemer et al. 2014). Heterogeneity in ASD extends beyond behavior to the disparate pathophysiological mechanisms that lead to diagnosis (Jeste and Geschwind 2014). This biological heterogeneity impedes researchers from identifying biomarkers for ASD (Cuthbert and Insel 2013). Diagnostic categories should necessarily define neurobiologically homogeneous groups to allow for the development of targeted treatments specific to a neurobiological signature of the disorder. This suggests that advances in mental health research necessarily rely on characterizing underlying neurobiological mechanisms of pathophysiology, and that psychiatry may be fundamentally limited as long as assessments are limited to observable phenomena (Pine 2017).

The National Institute of Mental Health, who developed the RDoC framework, (Insel, Cuthbert et al. 2010), proposed a radical shift in clinical neuroscience research in order to address the gap between neuroscience research and clinical translation to practical changes in diagnostic criteria for sensitive and specific nosologies for mental health disorders. Eight years after RDoC was released, the original guiding principles continue to be informative for translational neuroscience research. Expanding on the implications of RDoC, Dr. Bruce Cuthbert clarified an important deviation from traditional translational research approaches: instead of first defining a mental disorder based on observable symptoms, then trying to identify underlying pathophysiology, RDoC turns that model on its head (Cuthbert 2014). First, Dr. Cuthbert argues, we should characterize healthy variability in a particular cognitive construct, then identify its neural underpinnings, and finally, seek to understand what accounts for dysfunction of these neural systems along the normal-to-abnormal spectrum. To accomplish this, RDoC promotes a dimensional model of psychopathology, thus encouraging the inclusion of the full range of behavior in research samples from healthy to impaired. Importantly, the assumption of a continuum between mental health and illness is still an unsettled empirical question (Garvey, Avenevoli et al. 2016), with great implications for the design of future diagnostic systems and treatment development (Coghill and Sonuga-Barke 2012). Due to the promise of producing translatable findings, this dissertation used the RDoC framework in an initial investigation of cognitive flexibility which ultimately seeks to understand the neural substrates of normal to abnormal manifestations of this crucial skill.

In an effort to characterize the neural substrates of successful, developmentally mature cognitive flexibility, the first study aimed to identify the neural substrates of cognitive flexibility in neurotypical adults using a task that engages flexible behavior. Study II capitalized on the results of Study I, using the brain regions identified as regions-of interest (ROIs) for estimation of individual-level connectivity profiles in children. Study II aimed to define subgroups of children across a wide range of abilities (including children with autism spectrum disorder, attention-deficit/hyperactivity disorder, and typically developing children) with similar network connectivity profiles which support cognitive flexibility. This study was undertaken in an effort to parse heterogeneity inherent to these neurodevelopmental disorders as a first step towards developing an alternative nosology that may define neurobiologically homogeneous groups of children. In this way, targeted treatments may be developed that correspond to specific neurobiological alterations. By focusing on brain networks important for cognitive flexibility, we can target a skill that is known to be developmentally essential and accelerate improvements in this ability, which may translate into better quality of life for children with neurodevelopmental disorders and their caregivers.

CHAPTER 2. STUDY I: Neural correlates of cognitive flexibility in neurotypical adults

Cognitive flexibility enables an individual to work efficiently to disengage from a previous task, reconfigure to a new response set, and implement this new response set to the task at hand (Dajani and Uddin 2015). Cognitive flexibility supports the transition to adulthood, such as keeping and maintaining a job, navigating social relationships, and facilitating independent living (Bailey 2007, Kapp, Gantman et al. 2011, Burt and Paysnick 2012). In addition, greater cognitive flexibility is associated with greater resilience to negative life events and stress (Genet and Siemer 2011) and higher levels of creativity (Chen, Yang et al. 2014). Despite the extensive advantages of intact cognitive flexibility, rigorous examination of this construct has been elusive. Previous neuroimaging studies of cognitive flexibility have used a wide array of tasks (Konishi, Nakajima et al. 1998, Badre and Wagner 2006, Leber, Turk-Browne et al. 2008, Armbruster, Ueltzhoffer et al. 2012), some of which have never been psychometrically validated, leading to inconsistent results across studies. Crucially, experimental tasks must correspond to real-world behavior if we hope to glean insight into how the brain remains flexible in everyday environments, and not simply within a highly structured laboratory setting.

Traditional frameworks of validity in psychology focus on *construct validity* and *ecological validity* to assess how well a task measures what we purport it is measuring and how well that translates to performance in everyday life (Brunswik 1956, Campbell and Fiske 1959). At the most basic level, construct validity describes how well a task measures what it is designed to measure (Schimmack 2010). In the case of cognitive neuroscience, it is important to consider whether the fMRI tasks we are using are actually

measuring the constructs we intend to investigate (e.g. do task-switching paradigms truly measure cognitive flexibility?). To answer these questions, psychometric evaluation is necessary for novel tasks. Tasks used in fMRI studies are almost always novel in some respect, given that a laboratory-based task often needs to be changed to be implemented in the scanner environment. Many subtypes of validity contribute to construct validity, including convergent and divergent (i.e., discriminant) validity (Campbell and Fiske 1959). While convergent validity establishes which construct(s) a given task may be tapping, divergent validity ensures tasks tap specific constructs (e.g., cognitive flexibility) and not general abilities (e.g., processing speed). Ecological validity is comprised of both *representativeness*, which describes the "naturalness or artificiality" of the experimental environment, materials, and stimuli, and *generalizability*, or the degree to which the measured phenomenon is able to explain similar processes in everyday life (Kvavilashvili and Ellis 2004). Although most cognitive neuroscience research may be described as unrepresentative because it takes place within the confines of an fMRI scanner, generalizability may be achieved by demonstrating that the relationships under study hold under similar circumstances in everyday life. Kvavilashvili and Ellis (2004) argue that generalizability is more important for establishing ecological validity, and the absence of representativeness does not preclude ecological validity. Here, we focus on assessing the convergent, divergent, and ecological validity of a laboratory task of cognitive flexibility (Dick 2014) that was adapted to the fMRI scanning environment.

Of paramount importance in accurately identifying the neural correlates of cognitive flexibility is using a valid task, and ensuring that the task is still valid once it has been adapted for the fMRI environment. In the laboratory, cognitive flexibility is

typically measured with set-shifting or task-switching behavioral paradigms (Dajani and Uddin 2015). These tasks require switches in either "sets" (e.g., rule or perceptual feature) or "tasks", requiring individuals to first identify the appropriate rule to be implemented, inhibit any incorrect prepotent responses, and switch to the new rule for successful completion. In addition, cognitive flexibility tasks can be inductive or explicit. Inductive tasks require more effort to identify the appropriate rule, as these tasks require participants to choose the new response set to switch to, instead of being explicitly provided the relevant rule to complete the task (Yerys, Wolff et al. 2012). Consequently, inductive tasks engage participants in endogenous problem-solving strategies, requiring the internal generation of response sets. Prior task-based neuroimaging studies of cognitive flexibility have used both explicit (e.g., stimulus-response reversal paradigms, Cubillo, Halari et al. 2010); (dimensional change card sort, Zelazo 2006) and inductive set-shifting tasks (e.g., modified Wisconsin Card Sort Tasks [WCST], Konishi, Hayashi et al. 2002) to identify the neural correlates of cognitive flexibility.

Systematic reviews and meta-analyses that combine inductive and explicit tasks suggest that cognitive flexibility is instantiated in the brain via the interplay of many individual brain regions within two large-scale networks: the executive control (ECN) and salience networks (SN, Kim, Cilles et al. 2012, Dajani and Uddin 2015). Within the ECN, regions activated during cognitive flexibility tasks include the ventrolateral prefrontal cortex (vlPFC), dorsolateral prefrontal cortex (dlPFC), inferior frontal junction (IFJ) and posterior parietal cortex (PPC); within the SN, regions activated include anterior insula (AI) and dorsal anterior cingulate cortex (dACC). Subcortical regions such as the caudate and thalamus are also frequently reported (Kim, Cilles et al. 2012).

Previous task-based neuroimaging studies provide a foundation for our understanding of brain regions contributing to successful cognitive flexibility. However, these prior studies mainly rely on explicit cognitive flexibility tasks, which bear little resemblance to how cognitive flexibility is implemented in everyday life. We argue that inductive tasks are more ecologically valid than explicit tasks. Many situations outside the laboratory are "open-ended" or "ill-structured" (Burgess, Alderman et al. 2006), therefore tasks that include many possible options, requiring the individual to choose a course of action, will best resemble the open-endedness of everyday experiences. This feature imbues inductive tasks with greater ecological validity than explicit measures of cognitive flexibility. The neural correlates of inductive cognitive flexibility tasks are also separable from their explicit counterparts: inductive tasks not only recruit "canonical" cognitive flexibility regions (e.g., IFJ and PPC), but additional brain regions such as the frontal pole, thalamus and lentiform nucleus (Kim, Cilles et al. 2012), suggesting an important neurobiological distinction between inductive and explicit tasks. Consequently, neuroimaging studies using explicit cognitive flexibility tasks may be missing meaningful information about how cognitive flexibility is implemented in everyday scenarios.

On the other hand, past studies that *did* capitalize on inductive cognitive flexibility tasks relied on tasks that introduce an additional confound: reasoning associated with identifying that the "rule" has switched. For tasks like the WCST or intradimensional/ extradimensional tasks, rule switches are signaled by experimental feedback. Participants must realize that the rule that they are currently using, which *previously* generated correct responses, is *now* generating incorrect responses, and therefore a new rule should be used. In short, the ability to identify a rule switch depends

on how amenable participants are to experimenter feedback. Because these tasks do not explicitly signal that the sorting principle or task has changed, errors on these tasks might reflect difficulty in identifying a rule switch and not cognitive flexibility *per se*.

Prior neuroimaging research in cognitive flexibility is limited by the use of tasks with suboptimal construct validity and ecological validity, due to the disregard for assessing psychometric properties of newly adapted tasks and researchers' reliance on explicit tasks. In this study, we overcome these limitations by adapting a psychometrically validated laboratory test of cognitive flexibility called the Flexible Item Selection Task (FIST, Jacques and Zelazo 2001). The FIST is a set-shifting task, requiring participants to repeatedly update the dimension by which they choose two sets of stimuli that "go together in one way" (e.g., the two sets of stimuli are the same color, Jacques and Zelazo 2001). We chose this paradigm because it capitalizes on advantages of explicit tasks, in instructing participants to consistently respond based on one stimulus type for a single selection, thereby reducing potential confounds due to difficulty in identifying that the rule has switched, and inductive tasks, by allowing participants to choose which sorting principle with which to complete the task, rendering superior ecological validity. The laboratory version of the FIST has already undergone rigorous psychometric evaluation, shown to have high internal consistency ($R_I = .91$) and convergent validity with the WCST (Dick 2014). An additional benefit of the FIST is that it reduces demands on inhibition by limiting the influence of prepotent bias, because switching occurs right after establishing a set (i.e., dimension) rather than after maintaining the same set across several trials. Due to the numerous advantages of the FIST, the first aim of this study was to adapt the FIST to the fMRI environment.

The ultimate goal of adapting the FIST to the fMRI environment is to generate an accurate understanding of the brain regions important for cognitive flexibility in neurotypical adults. Participants completed a computer-based version to permit assessment of behavioral performance outside of the scanner, as well as self-report measures of executive function and repetitive/inflexible behaviors, followed by an fMRI-adapted version of the FIST designed to identify brain regions that are active while participants engage in cognitive flexibility. We expected to observe higher activation in vlPFC, dlPFC, AI, dACC, posterior parietal cortex, striatum and thalamus during flexibility trials compared with control trials (Kim, Cilles et al. 2012).

Methods

Participants

Participants were 32 adults ages 19-46 years (M_{age} = 25.29 years, SD = 6.42, 17 males) recruited from the University of Miami in Coral Gables, Florida and the wider Miami community. Eleven participants were Hispanic/Latino and 15 were not Hispanic/Latino (missing data for n=6); 16 participants reported their race as 'white' and 4 reported their race as 'other' (missing data for n=12). All participants were self-reported right handed with no reported history of psychological disorders. Informed consent was obtained for all participants and they received 50 dollars in compensation for their participation. The University of Miami Institutional Review Board approved the study.

Neuropsychological Measures

Participants completed multiple behavior rating scales prior to FIST administration: the Adult Behavioral Rating Inventory of Executive Function (BRIEF-A,

Roth, Isquith et al. 2005), Perseverative Thinking Questionnaire (PTQ, Ehring, Zetsche et al. 2011), Adult Repetitive Behaviors Questionnaire-2 (RBQ-2A, Barrett, Uljarević et al. 2015), and the Center for Epidemiologic Studies Depression Scale (CES-D, Eaton, Smith et al. 2004). The BRIEF-A and RBQ-2A were used in the present study.

BRIEF-A. The Adult Behavior Rating Inventory of Executive Function (BRIEF-A) is a 75-item self-report questionnaire used to assess executive function and self-regulation in adults' day-to-day lives. The BRIEF-A consists of nine subscales: Inhibit, Self-Monitor, Plan/Organize, Shift, Initiate, Task Monitor, Emotional Control, Working Memory, and Organization of Materials (Roth, Isquith et al. 2005). The BRIEF-A subscales show good internal consistency (α = .73-.90), test-retest reliability (r= .82-.93), and convergent validity with other subjective reports of executive function in a healthy adult sample (Roth, Isquith et al. 2005). Importantly, this measure of EF is also ecologically valid, as questions probe behavior in everyday, "real-world" situations (e.g., "I leave the bathroom a mess" and "I have problems completing my work"). T-scores for subscales that tap three core distinct, but related, EFs were used in this study based on a widely accepted model of EF (Friedman, Miyake et al. 2011)—shift, inhibit, and working memory subscales. The shift subscale corresponds to cognitive flexibility skills in adults' day-to-day lives (e.g., "I have trouble changing from one activity or task to another") and was used to assess the ecological validity of the FIST. The inhibit and working memory subscales were used to assess the divergent validity of the FIST. Higher T-scores indicate worse EF abilities and T-scores at or above 65 are considered clinically significant (Roth, Isquith et al. 2005).

RBQ-2A. The Adult Repetitive Behaviors Questionnaire-2 (RBQ-2A, Barrett, Uljarević et al. 2015) is a self-report measure of a core symptom of autism spectrum disorders (ASD), restricted and repetitive behaviors, designed for assessment in adults. Psychometric studies of neurotypical adults reveal two subscales that are internally consistent, correlate with other measures of autism traits, and distinguish between adults with and without ASD—repetitive motor behaviors (RMB) and insistence on sameness (IS, Barrett, Uljarević et al. 2015). These subscales and the total score were used as indices of ecological validity for the FIST because past studies have demonstrated that deficits in set-shifting are related to heightened RRBs in adults (Lopez, Lincoln et al. 2005, Miller, Ragozzino et al. 2015, Mostert, Hoogman et al. 2015).

Flexible Item Selection Task

The Flexible Item Selection Task (FIST) is an adapted version of the 4-Match FIST (Dick 2014). This task is designed to challenge abstraction and cognitive flexibility skills. For each trial, participants are presented with four cards oriented vertically on which various stimuli are presented. Each of the cards contained images that varied along four dimensions: color (blue, green, red), shape (boat, flower, rabbit), size (large [1.82”], medium [0.83”], small [0.38”]), and number of images (one, two, three). Stimuli were created in Microsoft PowerPoint 2010 and were presented on white cards with a light grey background. For a full description of the images used and selection of dimensions, see Dick (2014).

Participants were trained to choose a pair of cards that were “the same in one way” (i.e., a selection); participants were asked to make three selections per trial. Participants used a four-button rectangular button box to select cards based on their

location on the monitor (Figure 1). Although there were four possible matches, we only required that participants provide three matches to reduce task difficulty and to avoid an excessively long trial duration. Two versions of the 4-Match FIST were administered to participants: 1) a computer-based task, and 2) the fMRI-adapted task. Tasks were developed in E-Prime 2.0 for presentation and recording of behavioral responses.

Computer-based task. We sought to conceptually replicate the administration of the 4-Match FIST developed by Dick (2014). We used the same six trials used in Dick (2014), but switched the order of two successive trials in order to maximize differences in the correct button presses between consecutive trials. Additionally, our computer-based task differed from the Dick (2014) 4-Match in that we only required three selections per trial instead of four. For each trial, participants were prompted to "choose two cards that are the same in one way" (Jacques, 2001). Following each of the three selections, participants were asked "how are they [the cards] the same?" Participants were free to choose any three combinations of cards, and therefore were free to choose the dimension by which the cards matched. These trials were self-paced, and accuracy depended both on correct card selections (e.g., cards 1 and 2) *and* the respective dimension by which they match (e.g., "They are both blue"). In general, selections could be incorrect due to: 1) incorrect card pair selection (i.e., the cards did not match along any dimension), 2) choosing the same card pair selection more than once per trial, or 3) choosing a correct pair of cards, but specifying an incorrect dimension. Card selections were made by participants using the button box; an experimenter recorded the dimension by pressing a corresponding letter on a keyboard. Participants were asked to say aloud their card choices and dimension. For a trial to be considered correct, all three selections must have been

correct. E-Prime data was processed using an in-house MATLAB script to calculate selection-level and trial-level accuracy. Computer-based task trial-level accuracy was used for validity analyses.

fMRI-adapted task. To adapt the computer-based 4-Match to the fMRI environment, several changes were applied (Figures 1 and 2). First, for trials indexing cognitive flexibility ("flexibility trials"), participants were instructed to make their three selections one after the other, without indicating the dimension by which they chose their selections. Instead, participants were asked to simply think about why the cards are the same. In addition, participants were instructed to choose each pair of cards as quickly as possible, as these trials were of fixed duration (8s, based on pilot data with four adults). These changes were implemented to keep task length to a minimum while still maintaining an event-related design. Flexibility trials were indicated by a heading at the top of the screen: "Now you choose". Twenty flexibility trials were created using the same stimuli as for the computer-based task; all 20 trials for the fMRI task and the 6 trials for the behavioral task contained unique combinations of color, shape, size, and number of stimuli.

Novel control trials were created for the fMRI task. These visually resembled the flexibility trials exactly, but participants were provided the correct responses, indicated by a thick black border surrounding the correct cards (Figure 1). Each correct selection, indicated by two cards with a thick border, appeared for 2.6s. Each trial, consisting of three selections, totaled 8s to match the length of flexibility trials. Because we set each selection duration to 2.6s, response time for control trials was not used in any analyses of this study. For each flexibility trial, there was a respective control trial that contained the

same stimuli. This allowed us to control for the visual stimuli and button presses necessary for completing the flexibility trials. To create each control trial, three selections were systematically chosen out of the four possible answers, leaving one "excluded dimension" for each control trial. Each dimension was excluded about the same number of times across control trials used for task training and the fMRI task (6-8 exclusions per dimension). The respective control trials corresponding to their flexibility trial (with the same 4 cards) were not used in the same run. For both flexibility and control trials, the combination of two button presses for incorrect answers (6 total combinations) were distributed evenly (e.g., buttons 1 and 2 indicated an incorrect response 15% of the time).

Each participant completed 4 runs of the fMRI task, with each run consisting of 10 flexibility and 10 control trials. Runs 1 and 3 contained the same trials but in a randomized order. Runs 2 and 4 also contained the same trials but in a randomized order. Both flexibility and control trials were 8s in duration. Optseq2 was used to determine the order of trial presentation and jitter for the length of fixation trials (https://surfer.nmr.mgh.harvard.edu/optseq/).

Accuracy was calculated for flexibility and control trials based on recorded button presses; response time was also recorded for flexibility trials. For a flexibility trial to be accurate, a participant had to make six correct button presses (corresponding to 3 consecutive selections of pairs of cards) within the 8s interval. The button presses were divided into three consecutive pairs, and each pair was assumed to correspond to one selection. If the participant did not make six button presses, the trial was automatically scored as incorrect. If a participant selected a pair of cards that did not match along any of the four dimensions, the trial was incorrect. Control trial accuracy was calculated at the

selection-level. For a control trial to be accurate, all three selections must have been accurate. An in-house MATLAB script was developed to calculate trial-level accuracy for flexibility and control trials for each run.

A combined accuracy and response time metric was calculated to provide an index of flexibility for each run of the fMRI-adapted task. This metric was computed similar to that of the Dimensional Change Card Sort task, a measure of cognitive flexibility, implemented as part of the NIH Toolbox (Zelazo, Anderson et al. 2013). First, mean accuracy computed across flexibility trials was transformed into a 5-point scaled metric by dividing mean accuracy by 2 (i.e., a mean accuracy of 9 out of 10 trials translated into a score of 4.5). Next, response times for accurate trials were transformed to a 5-point scale. Across all runs and participants for correct trials only, median response times ranged from 3275 - 7427.5ms. Median response times were normally distributed, skew: -0.03-0.54 and kurtosis: -0.79-0.03, therefore no transformation was applied. The minimum RT value (3275 ms) was subtracted from original RT values such that new values ranged between 0 and 4152.5ms. Then, median response times were algebraically transformed to a 5-point scale and reverse-scored according to equation (1), where 4152.5 represents the sample-specific range of median response time values in ms:

(1) $$5 - \frac{medianRT * 5}{4152.5}$$

Based on the above equation, higher scores indicate faster response times. Finally, rescaled accuracy and response times were summed to constitute a singular variable ranging from 0 to 10, with higher scores indicating greater cognitive flexibility. This combined accuracy-response time metric was used as the dependent variable of interest in this study.

Experimental Procedure

All participants were trained on the FIST by an experimenter at a desktop computer before completing the computer-based task and the fMRI task. All demonstration and practice trials were unique from trials of the computer-based and fMRI task. To introduce participants to the FIST, they were first shown two trials of the 2-Match FIST (Dick 2014), an easier version of the 4-Match used in this study. Next, an experimenter demonstrated two trials of the 4-Match while providing a scripted verbal explanation of their selections to ensure the participant understood how to respond to the 4-Match version. During these demonstration trials, experimenters presented all four dimensions as possible answers. Following the demonstration, participants completed two practice 4-Match trials on their own and were given feedback if they answered incorrectly. Then, participants completed the computer-based task.

After the computer-based task administration, participants were shown examples of control and flexibility trials to prepare for the fMRI-adapted version. An experimenter first explained how to respond to a single self-paced control trial, which was followed by two timed practice control trials. Participants were introduced to the flexibility trials of the fMRI task by completing a shortened version of the fMRI task, with both flexibility and control trials, at a computer. Finally, participants completed the same practice fMRI task in the mock scanner. Prior to administering the first run of the fMRI task in the scanner, participants completed a "refresher" FIST task that consisted of 3 practice trials. This refresher was implemented midway through data collection to remind participants of task instructions and ensure acceptable accuracy levels for all runs of the fMRI task, therefore n=24 (75%) participants received this refresher.

MRI data acquisition

Task fMRI data were acquired for participants on a 3T GE Discovery 750 series scanner using a 32-channel head coil (TR=2s, TE=30s, flip angle=75°, 3.4mm slices, voxel size= 3.4 isotropic mm). The first 5 volumes were immediately discarded to account for magnet stabilization, resulting in 122 volumes per run. High-resolution T1-weighted FSPGR BRAVO scans were also acquired to facilitate registration of the functional image to standard space (TI = 650ms, flip angle=12°, FOV=25.6cm, 1mm isotropic voxels). Additional structural and functional images were acquired but were not analyzed in this study.

Preprocessing

Raw functional and structural images were quality checked prior to preprocessing using a standardized in-house coding scheme. Preprocessing was conducted in FSL 5.0.9. First, structural images were brain extracted using FSL's BET tool. Using FEAT, fMRI data underwent motion correction, slice time correction, smoothing with a 6mm kernel, high pass filtering (100s), coregistration to the structural image and normalization to the 2mm MNI template. Data were quality checked following structural brain extraction and normalization steps to ensure fidelity of preprocessing steps.

Analytic Plan

Task validation

All descriptive statistics, reliability, and validity analyses were computed using R 3.4.2 (Team 2017). Code for all R-based analyses and figures are publicly available (https://github.com/xDinaDajani/fMRI_FIST_adult.git).

Reliability. Internal consistency for accuracy was computed for each run of the fMRI task based on trial-level accuracy using Kuder-Richardson 20 formula (KR-20), which is suitable for dichotomous data (Kuder and Richardson 1937). Accuracy was used instead of the combined accuracy-RT metric because trial-level data is necessary for internal consistency calculations, and accuracy-RT was calculated only at the run-level. KR-20 was estimated using the DescTools R package (Signorell 2017). We used established guidelines to identify acceptable levels of internal consistency (>.70), further distinguished as fair (.70-.79), good (.80-.89), or excellent (>.90) (Cicchetti and Sparrow 1990).

Test-retest reliability was computed across all runs of the fMRI-adapted FIST, measured by the intraclass coefficient (ICC), using the run-level combined accuracy-RT metric as the dependent variable. To compute the ICC, a two-way random effects model was used (i.e., ICC(2, 4), where 4 represents the number of time points the variable of interest will be averaged across, Shrout and Fleiss 1979), which is best-suited for test-retest applications (Sainani 2017). We report both the ICC(2,4) formula and the ICC(2,1) formula, where the former represents performance averaged across all four runs and the latter considers reliability on any single run. The ICC was calculated using psych R package (Revelle 2017). Established guidelines were used to interpret ICC values: poor (<.40), fair (.40-.59), good (.60-.74), and excellent (.75-1.00) (Cicchetti 1994). To determine the minimum number of runs (m) needed to obtain a reliable mean estimate of task performance, the following formula was used (Shrout and Fleiss 1979), where ρ^* is the minimum acceptable reliability coefficient (here, 0.75) and ρ_L is the lower bound of the 95% confidence interval for the ICC(2,1) reliability estimate:

$$m = \frac{\rho^*(1-\rho_L)}{\rho_L(1-\rho^*)}$$

Validity. Convergent validity can be confirmed in the measure of interest by its coherence with independent measurements of similar constructs; divergent validity is demonstrated if the measure is not highly correlated with measures from which they were intended to differ. The computer-based task closely resembles the psychometrically validated 4-Match FIST (Dick 2014). Therefore, to test the convergent validity of our newly adapted fMRI version of the 4-Match FIST, we correlated the accuracy-RT metric averaged across Runs 1-4 with the computer-based trial-level accuracy. In addition, we assessed the fMRI 4-Match's ecological validity with a "real-world" measure of cognitive flexibility, the Shift subscale of the BRIEF-A, and with measures of restricted and repetitive behaviors (Barrett, Uljarević et al. 2015), which are associated with heightened deficits in cognitive flexibility (Lopez, Lincoln et al. 2005, Miller, Ragozzino et al. 2015, Mostert, Hoogman et al. 2015). Finally, to ensure the fMRI 4-Match is not simply a general measure of executive function, but specifically measures cognitive flexibility, we assessed the fMRI task's divergent validity by correlating task accuracy-RT with working memory and inhibition subscales of the BRIEF-A.

Task-based fMRI data analyses

General linear model (GLM). Flexibility and control trials (and their temporal derivatives) were modeled at the run-level for each subject using a gamma HRF using FSL. The six rigid motion estimates were included as nuisance regressors at the run-level. Runs were combined within-subjects using a fixed effects analysis. Finally, data were combined across subjects with a mixed effects design using FLAME 1 to identify brain regions activated for two contrasts of interest: Flexibility – Control and Control –

Flexibility trials. Main effects for Flexibility and Control trials were also modeled relative to fixation trials. Significant voxels were identified using a voxel-level threshold at an FWE-corrected $p < .05$, which adequately controls the false-positive rate (Eklund, Nichols et al. 2016).

Results

Task validation

Reliability. Participants completed four runs of the FIST in the scanner. Internal consistency for Runs 1 through 4, respectively, were 0.68, 0.81, 0.66, and 0.64. When reliability was assessed for Runs 1 and 2 combined, reliability reached .85 (95% CI [.84-.87]); reliability for Runs 3 and 4 was .79 (95% CI [.76-.81]). These values indicate "good" and "fair" reliability, respectively. The combined accuracy-RT metric averaged across runs exhibited excellent test-retest reliability as indexed by the intraclass correlation, ICC=.89, 95% CI [.76-.95]. Individual runs still demonstrated fair to excellent test-retest reliability, ICC=.67, 95% CI [.44-.82]. But, to achieve excellent reliability for the accuracy-RT metric (i.e., ICC≥.75), it is necessary to obtain a mean score across at least 4 runs (m=3.84).

Validity. The fMRI-adapted FIST combined accuracy-RT metric trended towards a large positive correlation with computer-based task accuracy ($r(30)$=.33, p = .07, Figure 3), suggesting that the fMRI-adapted version displays convergent validity with the original task developed by (Dick 2014). The fMRI-adapted FIST accuracy-RT metric did not correlate with any of the three subscales of a self-report of "real world" executive functions, a measure of repetitive motor behaviors, or a measure of insistence on sameness (p's > .55, Table 1, Figure 4).

Behavioral data

Computer-based task. Participants performed near ceiling levels on the computer-based task (trial-level accuracy [proportion of correct trials]: M=0.88, SD=0.17, Figures 5 and 6). A repeated-measures ANOVA demonstrated that participants performed equally well on all selections, $F(2, 62)=1.67$, p=.20 (selection 1: M=0.97, SD=0.08; selection 2: M=0.95, SD=0.15; selection 3: M=0.93, SD=0.09). When participants committed errors, they were most likely due to misidentifying the dimension by which cards match (52% of errors). Participants also identified cards that did not match along any dimension (31% of errors) and repeated their card choice (17% of errors).

Participants identified card pairs by "color" on nearly every trial (proportion of trials: M=.97, SD=.06), with "shape" being identified often as well (M=.83, SD=.31, Figure 7). The "size" dimension was identified less frequently (M=.68, SD=.38) and the "number" dimension was identified on the fewest proportion of trials (M=.52, SD=.38). Across participants, "color" was most frequently chosen on the first selection, whereas "size" and "number" were more frequently chosen on the third selections (Figure 8). The majority of participants tended to vary the order in which they used a particular dimension to choose card pairs. Most participants (56%) repeated a particular pattern for two out of the six trials (e.g., repeating the following pattern: "color" for selection 1, "size" for selection 2, and "shape" for selection 3); fewer participants never repeated a particular selection pattern (9%); even fewer repeated a pattern for four out of the six trials (6%).

fMRI-adapted task. Across all runs, participants exhibited high performance on both flexibility and control trials (Accuracy: M_{flex}=0.82, SD_{flex}=.20, $M_{control}$=0.90,

$SD_{control}$=.16; RT for correct trials (ms): $Median_{flex}$=5403, SD_{flex}=798.37, Table 2). A two-way repeated measures ANOVA demonstrated that for all runs, accuracy was higher for control trials compared with flexibility trials ($F(1, 31)=7.43$, $p=.01$, Figure 9). Further, there was a significant cubic trend in the accuracy-RT metric across runs, demonstrating an improvement in performance over time, $F(1, 93)=10.61$, $p=.002$ (Figure 10). Post-hoc contrasts, adjusted using Tukey's HSD, revealed that accuracy-RT was lowest for Run 1 (p's < .001), similar between Runs 2 and 3 ($p=.91$), and highest for Run 4 (p's < .03).

fMRI data

General linear model. As hypothesized, during both Flexibility and Control trials, activation was observed in canonical "task-positive" regions, including the posterior parietal cortex, lateral prefrontal cortices, dACC and anterior insula. Deactivation was observed in "task-negative" regions comprising the DMN (e.g., angular gyrus, anterior temporal lobe, posterior cingulate [PCC] and ventromedial PFC, Figure 11). The Flexibility – Control contrast revealed stronger and more widespread activation across the cortex for Flexibility trials relative to Control trials, including canonical areas of the dorsal attention network (DAN), ECN and SN (Table 3, Figure 12). Specifically, there was stronger activation in bilateral dlPFC, left IFJ, bilateral frontal eye fields, bilateral anterior insula, dACC/pre-SMA and bilateral inferior parietal lobule during Flexibility vs. Control trials. In addition, stronger activation during Flexibility trials was present bilaterally in the lower- and higher-order visual areas, posterior inferior temporal gyrus, precuneous, cerebellum, thalamus, and globus pallidus. There were also small clusters that showed weaker deactivation for Flexibility trials compared with Control trials in the precuneous, cuneus, and lingual gyrus. The Control – Flexibility contrast revealed

primarily weaker deactivation (i.e., stronger deactivation for Flexibility trials) in regions comprising the DMN including medial prefrontal, posterior cingulate, angular gyrus and anterior temporal regions (Table 4, Figure 12). The Control – Flexibility contrast also identified small clusters that had higher activation for Control relative to Flexibility trials in the right lateral occipital cortex, left parietal/central opercular cortex, and the right postcentral gyrus.

Discussion

The development and validation of tasks that capture real-world cognitive phenomena is nontrivial. Cognitive flexibility is difficult to operationalize in an MRI environment, and previous work has had limited success in identifying specific brain regions and networks that underlie this critical ability. We present an fMRI-adapted version of the Flexible Item Selection Task (FIST) that has been well validated in the behavioral literature. We confirm that this task is a reliable and valid measure of cognitive flexibility, despite the changes made to adapt the task to the scanner environment. As expected, we identified regions in canonical "task-positive" networks that were more active in response to trials engaging cognitive flexibility compared with a task controlling for visual and motor responses.

Cognitive neuroscience seeks to understand the neurobiological mechanisms underlying a variety of behaviors by asking individuals to perform highly circumscribed tasks within an unfamiliar environment – the bore of an fMRI scanner – while managing their arousal levels, response speed, and head movement. Yet, it is rarely acknowledged whether these tasks measure meaningful behaviors. Tests of reliability and validity can be used to formalize and ascertain the extent to which our artificial tasks map onto real-

world behaviors. Here, we made several changes to a task originally designed for in-person administration at a computer to render it suitable for the fMRI environment. Changes from the original behavioral task included requiring timed responses, removing the requirement to identify the dimension by which to choose pairs of cards that match, including novel control trials, and introducing a randomized jitter to the presentation and length of fixation trials. These changes were necessary for valid interpretation of neuroimaging findings, but may alter the psychometrics of the original computer-based task. It was therefore essential to reconfirm the reliability and validity of the task following fMRI-adaptation. We were able to confirm the reliability of the fMRI-adapted FIST, which, demonstrated fair to good internal consistency and excellent test-retest reliability across four runs completed on the same day. Further, a positive correlation was demonstrated between the computer-based administration, which had previously been validated (Dick 2014), and the novel fMRI-adapted version, suggesting our fMRI-adapted version converges with an established measure of cognitive flexibility. Although this correlation did not reach significance in the current sample (p=.07), it is likely that a larger sample size and more variability in performance on the computer-based measure would yield the expected correlation. Given our sample size (N=32), a correlation of r=.33, which corresponds to a medium effect size (Cohen 1988), indicates that the fMRI-adapted FIST displays convergent validity with the computer-based version. Future studies should attempt to replicate this result using a sample with at least 70 participants, which would be sufficient to detect a correlation of r=.33 with 80% power.

Contrary to our hypotheses, our fMRI-adapted FIST did not correlate with a self-reported measure of "real-world" executive functions or repetitive/inflexible behaviors,

precluding evidence for ecological validity of the FIST. Importantly, this was the first study to assess the ecological validity of the FIST. Several factors may have led to these null associations. First, the measure of ecologically valid EF used in this study, the BRIEF-A, has been shown in previous studies to correlate with other questionnaires (i.e., subjective measures) but *not* with laboratory-based (i.e., objective) measures of executive function (Toplak, West et al. 2013). This may not reflect an inherent defect in either the BRIEF nor in objective measures such as the FIST. In fact, researchers acknowledge that subjective and objective measures of executive function measure distinct constructs (Toplak, West et al. 2013, Silver 2014). Using the FIST, we captured information at the level of efficiency of cognitive flexibility available to an individual within the context of a structured setting (a laboratory environment where the goal was explicitly provided to the individual to "choose two cards that go together in one way"). Therefore, tasks such as the FIST focus mainly on performance at the level of neurobiological function and behavioral impairment (Chan, Shum et al. 2008). This is on par with the goal of cognitive neuroscience, which seeks to understand the neurobiological underpinnings of successful cognitive flexibility. On the other hand, rating scales such as the BRIEF most likely reflect information about success in rational goal pursuit in complex environments where many executive functions may be required in multi-step tasks. For this reason, individuals may perform well on an objective measure despite known executive dysfunction because they are being tested in an optimally structured environment. Based on these data, it may be beneficial to use objective measures of executive function to disentangle the neural correlates of executive subdomains, whereas subjective measures may be more beneficial

in clinical settings where disability or social participation levels are of interest (Chan, Shum et al. 2008).

Further validating our novel adaptation of the FIST, we observed increased activation in regions comprising the superordinate fronto-cingulo-parietal network during cognitive flexibility over and above basic visual and motor processes. Replicating findings from a previous meta-analysis of cognitive flexibility tasks, we identified greater activation in the IFJ, dlPFC, AI, dACC, IPL/IPS, posterior temporal cortex, extrastriate cortex and thalamus (Kim, Cilles et al. 2012). As is expected during externally oriented cognition (Anticevic, Cole et al. 2012), regions of the DMN decreased in activation compared with control trials.

The IFJ and IPL have been implicated in domain-general switching processes (across perceptual, response/rule, and endogenous switching paradigms), and are suggested to contribute to representing and updating task sets (Derrfuss, Brass et al. 2005, Kim, Cilles et al. 2012). Perceptual switching paradigms, like the FIST, tend to engage fusiform, inferior temporal, and occipital cortices more so than response- or task-switching paradigms (Kim, Cilles et al. 2012), which may explain their engagement during the FIST. Endogenous switching paradigms, in comparison with explicit tasks, have been shown to specifically activate the frontopolar cortex, implicating its role in the internal generation of cognitive representations (Kim, Cilles et al. 2012). Accordingly, we also observed higher activity in the left frontal pole. The dlPFC has most notably been associated with working memory (Petrides 2000), and has specifically been shown to elicit sustained activation during working memory tasks, suggesting its primary role is in maintenance of information as opposed to updating or manipulation (Cohen, Perlstein et

al. 1997). Consistent with the neuroimaging evidence for the role of the dlPFC in working memory maintenance, behavioral studies demonstrate that there are unique contributions of cognitive flexibility and working memory to FIST performance, and working memory maintenance contributes to item difficulty for the FIST (Dick 2014).

We observed activation of bilateral posterior parietal cortices (PPC) in response to flexibility trials. Evidence from a neuropsychological model of executive function suggests that all executive function tasks require the contribution of "common EF", or ongoing task monitoring and maintenance (Friedman, Miyake et al. 2011). Conjunction analyses across an array of executive function tasks have demonstrated few regions involved in common EF, restricted to the PPC, including the SPL and IPS (Collette, Van der Linden et al. 2005). Therefore, we surmise the role of the PPC in cognitive flexibility is attending to salient stimuli in the service of maintaining task set representations (Kim, Cilles et al. 2012).

Regions that are a part of canonical task-positive networks, including the DAN (Fox, Corbetta et al. 2006) and SN (Seeley, Menon et al. 2007) displayed stronger activation in response to Flexibility trials while controlling for visual and motor processes. The AI and dACC are the major nodes of the SN, which is implicated in detecting behaviorally relevant stimuli. Specifically, the AI has been linked to initiating neural switches between engagement of the ECN and DMN, guiding externally oriented cognition (Uddin 2015). Bilateral frontal eye fields, which are core nodes of the DAN (Fox, Corbetta et al. 2006), were also engaged during flexibility trials, which may have enabled externally directed attention towards the stimulus features necessary to complete the task. FEF activity cannot simply be explained by saccades to the particular pair of

cards chosen by participants, since this was also required for successful completion of control trials, where the "correct" card pairs were provided. Therefore, we suggest that heightened FEF activity during flexibility trials is driven by switches in spatial attention to the specific perceptual features of the stimuli guiding successful trial completion.

We also observed flexibility-specific activity in the left dorsal IFG. In a functional parcellation study of the right IFG, the dorsal portion dissociated into functionally distinct anterior and posterior areas, corresponding to bilateral co-activation patterns that mapped on to executive functions and spatial attention, respectively (Hartwigsen, Neef et al. 2018). Difficulty in switching between card pair choices on the FIST can be attributed to two processes: attentional inertia, the need to shift *away* from the previously attended dimension, and negative priming, the need to shift *towards* the previously ignored dimension. Hartwigsen et al. (2018) suggest the anterior dorsal IFG may be particularly important for overcoming the effects of attentional inertia. The posterior dorsal IFG is functionally connected to regions comprising the DAN, and may be particularly important for top-down attentional orienting (Hartwigsen, Neef et al. 2018).

Although multiple meta-analyses of cognitive flexibility tasks report activation in subcortical structures including the thalamus, basal ganglia, and cerebellum (Wager, Jonides et al. 2004, Kim, Cilles et al. 2012, Niendam, Laird et al. 2012), relatively little attention is paid to the contribution of these subcortical structures to cognitive flexibility. In line with these previous meta-analyses, we identified higher activation in response to flexibility trials in bilateral thalamus, globus pallidus, and cerebellum. The globus pallidus and cerebellum are thought to contribute to cognition through their projections to the dlPFC via distinct thalamic nuclei (Middleton and Strick 2000). In accord, a

neuroimaging study of cognitive flexibility demonstrated that the globus pallidus modulated top-down attentional biasing by the IFG on visual regions in the service of a perceptual switching task (van Schouwenburg, den Ouden et al. 2010). Further, (van Schouwenburg, Onnink et al. 2014) demonstrated that white matter integrity within the globus pallidus and cerebellum related to optimal cognitive flexibility in a mixed sample of healthy adults and individuals with ADHD. These data provide evidence for the role of the globus pallidus in mediating cognitive flexibility via associative basal ganglia loops.

Comparatively, the role of the cerebellum and thalamus in executive function is not well known. Recently, the potential role of the thalamus in executive functions has been acknowledged based on dense connections between its mediodorsal nucleus (MD) and the ACC, dlPFC, and premotor cortex (Halassa and Kastner 2017). Further, lesions to the MD (and internal medullary lamina) are associated with reduced cognitive flexibility (Van der Werf, Witter et al. 2000). While we are unable to distinguish individual thalamic nuclei in this study due to spatial resolution limitations and spatial smoothing, based on past studies and the use of the oxford thalamic connectivity probabilistic atlas, we infer that thalamic regions projecting to PFC were engaged during the FIST. The cerebellum is hypothesized to contribute not only to postural and sensorimotor functions, but also to higher-level cognition such as language, emotion, and executive functions (Schmahmann 1991). In a meta-analysis of task-based neuroimaging studies, (Stoodley and Schmahmann 2009) reported an anterior to posterior gradient underlying sensorimotor and cognitive/emotional functions, respectively. Here, we identified clusters with higher activity in response to flexibility trials distributed bilaterally across both anterior (lobules I-IV) and posterior portions (crus I, crus II, VIIb, and VI) of the

cerebellum. It is likely that activity in the posterior cerebellum, which projects to dlPFC, was related to working memory and/or cognitive flexibility, whereas activity in anterior regions, which project to primary motor cortex, reflects automatization of motor responses necessary to complete the FIST (Kelly and Strick 2003).

Several limitations of this study should be noted. This was the first study to examine the psychometric properties of the fMRI-adapted version of the FIST, and one of the few cognitive neuroscience studies to assess psychometric properties of an fMRI task. In this initial effort, we used a sample size that is considered small for psychometric studies. This small sample size may have led to the failure to detect relationships between the FIST and self-report measures of executive function and ritualistic behavior, as the sample size needed to detect a medium effect (r=.30) with 80% power is 85. Therefore, future studies should follow up on the psychometric properties of this task, especially concerning validity and the task's relationship to objective measure of cognitive flexibility and other executive functions with a larger sample size. We are confident that this task taps cognitive flexibility in adults, but it is likely that other executive functions, including working memory and inhibition, are also engaged in response to the FIST's flexibility trials. To isolate these processes, one potential strategy would be to define contrasts between selections *within* a trial, as an alternative to defining contrasts at the trial-level. In the present study, we sought to minimize working memory demands by making selections occur one after the other, without intervening fixation trials, which would have increased trial length and subsequent working memory demands. Unfortunately, this design precluded the analysis of within-trial contrasts. Future studies may seek to optimize this task design in order to allow within-trial contrasts, which

isolate cognitive flexibility processes from abstraction skills and provide the ability to parametrically model working memory load across selections.

For the first time, we adapted a laboratory-based inductive cognitive flexibility task, the FIST, to the fMRI environment. We provide evidence that the fMRI-adapted task is reliable when collecting four runs per participant, and report preliminary evidence of the task's convergent validity with the laboratory-based version of the task. We also found strong evidence for validity of the task based on robust activation of canonical regions of the executive control, salience and subcortical networks in response to flexibility trials. These results provide support for using the FIST in future cognitive neuroscience studies seeking to understand the neural correlates of cognitive flexibility.

CHAPTER 2. STUDY II: Parsing heterogeneity in autism spectrum disorder and attention-deficit/hyperactivity disorder using neural substrates of cognitive flexibility

Autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) are prevalent neurodevelopmental disorders most commonly diagnosed in the United States according to a symptom-based classification system, the Diagnostic and Statistical Manual of Mental Health Disorders 5 (DSM-5, American Psychiatric Association 2013). Although these disorders are characterized by separate core deficits (in ASD, social communication deficits and restricted and repetitive behaviors [RRBs]; in ADHD, primarily inattentive and/or hyperactive and impulsive symptoms), overlap in behavioral presentation and biological substrates obfuscate the distinctions between these diagnostic categories. In particular, there is significant variability among children with ASD and ADHD concerning deficits in a subdomain of executive function (EF) – cognitive flexibility (Dajani, Llabre et al. 2016). Mixed evidence for distinct diagnostic categories suggests an alternative diagnostic system focusing on the full range of variation in behavior (i.e., Research Domain Criteria [RDoC], Cuthbert and Insel 2013) may be better-suited to identify individuals with a common biological pathway to abnormality. As a first step towards developing an improved nosology for neurodevelopmental disorders, this study aims to use individual connectome mapping to identify children with altered brain network connectivity that may contribute to impaired cognitive flexibility.

Cognitive flexibility undergoes protracted development across childhood to young adulthood (Anderson 2008), supporting a wide range of behaviors that impact life outcomes (Diamond and Lee 2011). In childhood, effective cognitive flexibility predicts

better math and reading abilities (Bull, Espy et al. 2008) and social-emotional development such as false belief understanding (Farrant, Maybery et al. 2012). Deficits in cognitive flexibility can occur in healthy children and adults without accompanying mental illness, but deficits are more prevalent in almost every psychiatric population, such as autism spectrum disorder, attention-deficit/hyperactivity disorder, conduct disorder, depression, obsessive-compulsive disorder, substance abuse and schizophrenia (Diamond and Lee 2011, McTeague, Goodkind et al. 2016). Heightened rigidity in cognition and behavior may predispose individuals to poorer psychiatric morbidity, as evidenced by higher relapse rates in schizophrenia (Chen, Hui et al. 2005). On the other hand, behavioral interventions that improve cognitive flexibility have shown to also ease associated psychiatric symptoms (Tamm, Nakonezny et al. 2014). The ramifications of deficits in cognitive flexibility cannot be ignored. It is therefore imperative that we understand the neural underpinnings of this essential skill across the lifespan.

Traditional diagnostic systems for mental illnesses are limited in that they define diagnostic categories that have high biological and behavioral heterogeneity, allow for high between-category overlap, and cannot accurately predict treatment responsiveness or prognosis. Person-centered studies, which take into account within-population heterogeneity, substantiate evidence that the majority of (but not all) children with ASD and ADHD exhibit severe cognitive flexibility deficits (Gioia, Isquith et al. 2002, Fair, Bathula et al. 2012). Deficits in cognitive flexibility are particularly concerning because they are related to higher levels of core symptomatology in both ASD and ADHD: elevated RRBs in children with ASD and worse hyperactive-impulsive symptoms, oppositional defiant disorder symptoms, and lower intelligence in children with ADHD

(Lopez, Lincoln et al. 2005, D'Cruz, Ragozzino et al. 2013, Roberts, Martel et al. 2013). In addition, there are high rates of comorbidity between diagnostic groups, with rates of comorbid ADHD in children with ASD ranging from 37-85% (Leitner 2014). Numerous studies have demonstrated the detrimental impact of comorbidity between ASD and ADHD diagnoses, citing poorer adaptive functioning, health-related quality of life (Sikora, Vora et al. 2012) and higher rates of clinically impaired cognitive inflexibility than children with either disorder alone (Yerys, Wallace et al. 2009, Dajani, Llabre et al. 2016). Further, there is strong evidence for shared heritability of ASD and ADHD (Ghirardi, Brikell et al. 2018). An alternative nosology based on neurobiologically homogeneous subgroups will aid in the ultimate goal to identify children who stand to benefit from targeted treatments specific to their brain network connectivity alterations (Cuthbert 2014). Due to the controversies surrounding the separability of ASD and ADHD diagnostic categories based on both biological and behavioral characteristics, we propose that children with one disorder or the other should be considered together in an effort to develop an alternative nosology. This conceptualization is in line with the RDoC approach, which advocates for single studies to span multiple diagnostic groups (Cuthbert 2014).

One unresolved issue in psychiatry is whether psychopathology should be conceptualized categorically, similar to traditional systems, or dimensionally, including the full range of behavior from normal to abnormal (Coghill and Sonuga-Barke 2012). While a categorical approach implies that mental disorders are qualitatively different from typical behavior, dimensional approaches assume that mental disorders are an extreme on a continuum of behavior represented across the population. Although many

studies have attempted to find homogeneous, categorical subgroups within ASD and ADHD diagnostic categories using psychological or brain imaging data (e.g., van der Meer, Oerlemans et al. 2012, Gates, Molenaar et al. 2014, Costa Dias, Iyer et al. 2015, Kernbach, Satterthwaite et al. 2018), very few studies explicitly compare categorical and dimensional fits of the data using taxometric or factor mixture modeling. These few available studies have only focused on symptom-based and neuropsychological measures, demonstrating that ASD is a discrete category distinct from typical social communication and repetitive behaviors (Frazier, Youngstrom et al. 2010, Frazier, Youngstrom et al. 2012) and suggest the presence of three subgroups within the ASD category (Georgiades, Szatmari et al. 2013). Conversely, studies have consistently demonstrated a dimensional structure for ADHD symptoms in children (Haslam, Williams et al. 2006, Frazier, Youngstrom et al. 2007, Lubke, Hudziak et al. 2009, Marcus and Barry 2011), though one notable exception found a categorical fit using measures of EF and reward processing tasks (Stevens, Pearlson et al. 2018). Other studies focusing on resting state functional connectivity of large-scale brain networks report that ASD and ADHD symptomatology comprise both categorical and dimensional aspects, but did not explicitly test model fit (Chabernaud, Mennes et al. 2012, Elton, Alcauter et al. 2014, Elton, Di Martino et al. 2016). Therefore, it remains an open question whether the structure of neurodevelopmental disorders is categorical or dimensional at the neurobiological level.

One well-replicated finding in clinical psychology is a hierarchical taxonomy of psychopathology that exists dimensionally across healthy and patient populations (Lahey, Krueger et al. 2017). This ‘p’ factor encapsulates the propensity to develop any form of psychopathology, including anxiety, depression, substance use disorders, and

schizophrenia. In individuals with a diagnosed mental health disorder, the 'p' factor describes the severity, chronicity, and comorbidity of symptoms. In children, the 'p' factor extends to symptoms characteristic of neurodevelopmental disorders, including ADHD and ASD (Neumann, Pappa et al. 2016, Martel, Pan et al. 2017). Further, in children, this general liability to develop psychopathology has been related to poorer EF abilities, poorer effortful control, and is heritable via single-nucleotide polymorphisms (Neumann, Pappa et al. 2016, Martel, Pan et al. 2017). Dimensional structures at the symptom-level do not necessarily imply the same structure at the brain network-level, but initial neuroimaging findings support dimensional models of psychopathology. Dysfunctional activation and functional connectivity of networks important for EF underlie this 'p' factor in children and adolescents (Shanmugan, Wolf et al. 2016, Liu, Liao et al. 2018).

Similar to the 'p' factor models, the RDoC framework opts to model behavior dimensionally, but moves away from focusing on DSM symptoms and towards neurobiologically validated functional constructs, such as EF (Garvey, Avenevoli et al. 2016). Deficits in EF are a transdiagnostic marker of psychopathology, meaning deficits can be detected in children without any traditional diagnoses, and across diagnostic groups such as ASD and ADHD (Dajani, Llabre et al. 2016, McTeague, Goodkind et al. 2016). The current study focuses on a specific aspect of EF, cognitive flexibility, because cognitive flexibility is impacted in various neurodevelopmental disorders (Dajani, Llabre et al. 2016) and these impairments relate to worse symptom severity and lower academic achievement in youth (Lopez, Lincoln et al. 2005, D'Cruz, Ragozzino et al. 2013, Roberts, Martel et al. 2013). Moreover, established flexibility-specific interventions exist

that improve deficits in children with ASD and ADHD (Kenworthy, Anthony et al. 2014, Tamm, Nakonezny et al. 2014). EF deficits that are left untreated may persist and impact social competence and friendship quality in adolescence (Rosenthal, Wallace et al. 2013, Lieb and Bohnert 2017). Thus, it is imperative that researchers develop a reliable method to identify children who exhibit the most severe flexibility deficits.

As a basis for an alternative nosology, we leverage the strength of neurobiological variables due to the many limitations of behavior-based classifications (Waterhouse and Gillberg 2014). Most notably, behavior does not map one-to-one to underlying neurobiology (Pessoa 2014), meaning many disparate brain alterations may manifest as a singular phenotype (referred to as redundancy, see Licinio and Wong 2013). This is especially problematic if treatments only benefit subgroups of children with similar underlying neurobiological deficits (Loth, Spooren et al. 2016). Our group recently demonstrated that subgrouping approaches applied only to behavioral variables do not map well onto neurobiologically defined metrics. We developed an EF-based classification system based on parent-report and behavioral measures. This behavior-based classification system did *not* produce neurobiologically distinct subgroups, as assessed by functional connectivity metrics indexing the integrity of large-scale functional brain networks important for cognition (Dajani, Burrows et al., https://www.biorxiv.org/content/early/2018/08/22/396317). These data underscore the importance of moving beyond behavioral observations and focusing to a greater extent on differences in underlying *neurobiological* mechanisms of these disorders. Therefore, in this study we leverage large-scale neural networks as the foundation for developing an alternative nosology for two prevalent neurodevelopmental disorders.

Here, we focus on functional connectivity approaches estimated with resting-state fMRI data due to the large body of work implicating fronto-parietal functional networks in cognitive flexibility (Kim, Cilles et al. 2012). Specifically, we aim to investigate functional brain networks important for cognitive flexibility using the triple network model of psychopathology (Menon 2011), focusing on the frontoparietal (FPN), salience (SN), and default mode networks (DMN). We use a sample of children with ASD, ADHD and TD children with a wide range of EF abilities to identify subgroups with similar brain network connectivity profiles or "connectomes". This study applies a cutting-edge method to construct individual-level connectomes by combining structural equation modeling (SEM) and unsupervised machine learning to rs-fMRI data called Group Iterative Multiple Model Estimation (GIMME, Gates and Molenaar 2012).

Most studies intending to characterize the neurobiological underpinnings of ASD and ADHD do not aim to understand heterogeneity inherent to each disorder, but instead effectively treat this variability as noise (Lenroot and Yeung 2013). As a result, only two published studies have investigated subgroups of children with ADHD based on functional connectivity metrics, and no study has examined brain network connectivity-based subgroups in ASD. In ADHD, 5 subgroups have been found that differ in connectivity of fronto-parietal regions (Gates, Molenaar et al. 2014), whereas 3 subgroups emerged that differed in connectivity of reward-related networks (i.e., nucleus accumbens-whole brain connectivity Costa Dias, Iyer et al. 2015). We previously showed that three subgroups exist amongst children with ASD, ADHD, and TD children differing in their levels of cognitive flexibility (Dajani, Llabre et al. 2016), therefore, we predict *at*

least three subgroups will emerge that differ on network connectivity profiles important for implementing cognitive flexibility.

Following the results of our previous study (Dajani, Llabre et al. 2016), we hypothesize that one subgroup will include predominantly older TD children with average to above average EFs and low levels of psychopathology. This subgroup may exhibit network connectivity that resembles mature, adult-like networks, characterized by strong, positive connectivity within the FPN, negative functional connectivity between the DMN and FPN (Fair, Nigg et al. 2012, Satterthwaite, Wolf et al. 2013), and integration between SN, FPN and subcortical nodes (Marek, Hwang et al. 2015, Morgan, White et al. 2018). We expect another subgroup to include predominantly younger TD children with less mature network profiles, including weaker within-network connectivity (Bassett, Xia et al. 2018), stronger between-network connectivity, and weaker integration between task-positive networks. In line with the delayed maturation hypothesis (Rubia 2018), we expect this group to also include older children with ADHD who have intact EFs. Children with ASD without elevated comorbid ADHD symptomatology may also comprise this subgroup. Finally, we expect a third subgroup will include children with ADHD and ASD with impaired EFs, elevated general psychopathology (Martel, Pan et al. 2017) and aberrant modular architecture of brain networks (Xia, Ma et al. 2018), which may present as weak connectivity within the FPN and higher FPN-DMN connectivity (Zhong, Rifkin-Graboi et al. 2014, Stevens, Pearlson et al. 2018).

Methods

Participants

Participants ages 8 to 13 years (N=132, Table 5) included a subset of children used in our previous study investigating heterogeneity in EF ability in TD, ADHD and ASD groups (Dajani, Llabre et al. 2016). Written informed consent was obtained from all legal guardians and written assent was obtained from all children. All procedures were approved by the Institutional Review Board at the Johns Hopkins School of Medicine and all methods were carried out in accordance with the approved guidelines.

Diagnostic and neuropsychological measures

Community diagnoses of ASD were confirmed with the Autism Diagnostic Observation Schedule (ADOS-G, Lord, Risi et al. 2000 or ADOS-2, Lord, Rutter et al. 2012, based on study enrollment date) and Autism Diagnostic Interview-Revised (ADI-R, Rutter, Le Couteur et al. 2005). All ASD participants scored ≥7 on the total score on the ADOS-2 or the communication and social interaction score on the ADOS-G. All ASD participants with data for the ADI-R (n=34, 94% of ASD sample) met criteria for ASD based on established cutoffs on the ADI-R (≥10 for social interaction, ≥8 for communication/language and ≥3 for RRBs) except for one ASD participant. This participant was still included in the ASD group because they met criteria based on the ADOS-G (communication and social interaction score: 13).

The Diagnostic Interview for Children and Adolescents IV (Reich, Welner et al. 1997) was used to confirm community ADHD diagnoses, determine whether children with ASD had comorbid ADHD, and for exclusionary purposes. Community diagnoses of ADHD were also confirmed with the Conners' Parent Rating Scales (CPRS-R:L,

Conners 1997 or CPRS-3, Conners 2008, based on study enrollment date) and the ADHD Rating Scale IV, Home version (DuPaul, Power et al. 1998). All participants with ADHD met criteria based on the DICA-IV, except for one child who had missing data. This child met criteria based on both the CPRS-3 and the ADHD Rating Scale IV. TD participants all had T-Scores <65 on both the Hyperactive/Impulsive or Inattentive scales of the Conners' and only met criteria for ≤3 symptoms on either the Hyperactive/Impulsive or Inattention scales of the ADHD Rating Scale IV. In accordance with the RDoC framework, participants with comorbid psychiatric disorders were not excluded. See Table 6 for detailed diagnostic information.

CBCL. The Child Behavior Checklist (Achenbach 1991) is a parent-report of children's emotional and behavioral problems. The social problems subscale was used to index social communication and interaction symptoms that are a hallmark of ASD. This subscale is internally consistent (α = .82), test-retest reliable (r = .90), and validly distinguish between children with and without an ASD diagnosis (Achenbach and Rescorla 2001, Mazefsky, Anderson et al. 2011). T-scores between 67 and 69 represent the borderline clinical range; T-scores≥70 are considered clinically elevated.

RBS-R. The Repetitive Behavior Scale-Revised (RBS-R, Bodfish, Symons et al. 2000) is a parent-report of six domains of RRBs: rituals, sameness, self-injurious behavior, stereotypic behavior, compulsive behavior, and restricted interests. The subscales have poor to good interrater reliability (.55-.78) and test-retest reliability (.52-.96). Total RRBs and a subdomain of RRBs, insistence on sameness, have been shown to be correlated with deficits in cognitive flexibility in ASD (Lopez, Lincoln et al. 2005, Miller,

Ragozzino et al. 2015), therefore the total score and sameness indices were used in this study. Higher scores indicate greater impairment.

Conners' PRS. The Conners' Parent Rating Scales-Revised, Long Version (CPRS-R:L, Conners 1997) is a parent report of children's ADHD symptoms, oppositional defiant disorder and conduct disorder. Here, we used the T-scores from the DSM-IV inattentiveness and hyperactive/impulsive symptom subscales. Higher scores indicate greater impairment.

BRIEF. The Behavior Rating Inventory of Executive Function (BRIEF, Gioia, Isquith et al. 2000) is a parent-report of EF impairment of children 5-18 years of age. All T-score subscales will be used to assess EF impairment: inhibition, shift, emotional control, initiate, working memory, plan/organize, organization of materials and monitor. To avoid redundant analyses, we did not analyze the composite scores, which are combinations of the subscales (Behavioral Regulation Index, Metacognition Index, and Global Executive Composite). The subscales are reliable in normative (r = .76-.85) and clinical samples (r = .72-.84) and can distinguish clinical populations from TD children (Gioia, Isquith et al. 2000). Higher scores indicate greater impairment, with T-scores ≥65 indicating clinical impairment.

Data acquisition

Children completed a mock scanning session prior to fMRI data collection to acclimatize them to the scanning environment. rs-fMRI data were acquired for participants on a Phillips 3T scanner using an 8-channel head coil (TR=2.5s, flip angle=70°, sensitive encoding acceleration factor=2, 3mm slices, voxel size= 2.7x2.7x3 mm, 156 volumes). The first 10 volumes were immediately discarded to account for

magnet stabilization. For the rs-fMRI data acquisition, children were asked to relax with their eyes open and focus on a crosshair while remaining as still as possible. High-resolution T1-weighted scans were also acquired to facilitate registration of the functional image to standard space (TR=8.0ms, TE=3.7ms, 1mm isotropic voxels). Participants were asked to withhold stimulant medication (e.g., Adderall) on the day of MRI scanning, similar to prior neuroimaging studies comparing children with ASD and ADHD (Di Martino, Zuo et al. 2013, Dennis, Jahanshad et al. 2014). Non-stimulant medications were continued as prescribed (e.g., antidepressants, allergy medication). TD children were not taking any psychotropic medications.

Preprocessing

There were systematic differences in the length of resting-state scans by diagnostic group (*n*=16 ASD children were scanned using a 128-volume protocol, while only *n*=1 TD and *n*=1 ADHD child were scanned using this shorter scan length protocol). To maximize power to estimate connectivity maps, which is dependent on the length of the timecourse, only participants with the longer protocol (156 volumes) were included in this study. Participants with maximum absolute motion in any of the six rigid directions >3mm/degrees were excluded. Preprocessing was conducted using a combination of FSL 5.0.9 (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) and SPM12 (http://www.fil.ion.ucl.ac.uk/spm/doc/). First, structural images were brain extracted using FSL's BET tool. Using SPM12 and custom MATLAB scripts, structural images underwent resampling to the EPI image resolution, coregistration to the subject's mean EPI, and segmentation into grey matter, white matter (WM), and CSF components. These WM and CSF masks were used to compute average WM and CSF timecourses to be used

as nuisance regressors at a later preprocessing step. Using FSL's FEAT, raw fMRI data underwent motion correction, 4D intensity normalization, smoothing with a 6mm kernel, and estimation of linear and non-linear warping parameters to normalize to the MNI152 2mm template. Next, independent component analysis-based automatic removal of motion artifacts (ICA-AROMA) was used to remove motion-related artifacts in native space (Pruim, Mennes et al. 2015). ICA-AROMA works by running an individual-level ICA for each subject, classifying motion-related components as noise, and regressing out motion-related components' signals from the individual's 4D time course. The residual time course (with motion-related signal regressed out) was used for subsequent analyses. The denoised data underwent additional nuisance regression (WM, CSF, and linear trends) and band-pass filtering (.01-.10 Hz). Finally, warping parameters generated at an earlier step using FSL were used to normalize the data to the MNI 152 2mm template.

Region of interest selection

The ideal set of regions of interest (ROIs) for this study would be regions that are consistently activated when engaging cognitive flexibility in children, identified by a meta-analysis of neuroimaging studies that used psychometrically validated cognitive flexibility tasks. Unfortunately, no current meta-analyses of cognitive flexibility tasks exist that are specific to middle childhood. Further, the meta-analyses of cognitive flexibility that do exist for adults include a mix of explicit and inductive tasks that were not psychometrically validated (e.g., (Kim, Cilles et al. 2012). Individual neuroimaging studies of cognitive flexibility in children report inconsistent results regarding regions engaged, and are therefore not ideal to guide ROI selection (Dajani and Uddin 2015). Despite the use of the FIST in adults in Study I, due to the numerous advantages of the

task, including its established reliability and validity, we opted to use the neuroimaging results of Study I to inform ROI selection for Study II. Further supporting the FIST's use in guiding ROI selection, the laboratory-based version of the FIST has been validated in children (Dick 2014) and has previously been used in a study of children with ASD (Yerys, Wolff et al. 2012).

Using an fMRI-adapted version of the Flexible Item Selection Task, we identified cortical and subcortical brain regions that were activated over-and-above basic visual and motor processes associated with the task (Flexibility > Control contrast). Additionally, to define regions within the DMN, which generally deactivate when engaging in cognitive flexibility, we identified regions that were more active during the visual-motor control than for the flexibility trials (Control > Flexibility contrast). To facilitate network analyses and interpretations, the Power (2011) parcellation (Power, Cohen et al. 2011) scheme was used to assign ROIs to networks. Nodes within four large-scale networks integral to cognitive flexibility were included: frontoparietal network (FPN), salience network (SN), DMN and the subcortical network (SUB) (Niendam, Laird et al. 2012, Vatansever, Manktelow et al. 2016). Specific nodes were chosen based on their activation or deactivation in response to the Flexible Item Selection Task (Figure 13, Table 7). Within the FPN, we included the left inferior frontal junction (lIFJ), bilateral frontal eye fields (FEF) and bilateral superior parietal lobule (SPL). Within the SN, we included bilateral anterior insula (AI) and the dorsal anterior cingulate cortex (dACC). Within the DMN, we included the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), and bilateral temporal poles. Finally, within SUB, we included bilateral globus pallidus and thalamus. The SEM-based connectivity analysis utilized here

(GIMME) does not perform well with large numbers of ROIs. Accordingly, to reduce the total number of ROIs included, cingulo-opercular, dorsal attention, ventral attention, primary sensory and cerebellar networks were excluded from this analysis. ROIs were centered on the coordinates listed in Table 7 with a 4mm-radius sphere. Timecourses were averaged within each ROI and carried forward for estimation of network connectivity maps using the 'gimme' package (version: 0.4) in R (version: 3.3.1) (Gates, Lane et al. 2017, Lane and Gates 2017).

Group-, subgroup-, and individual-level connectivity map estimation

GIMME (Gates, Lane et al. 2017) capitalizes on the strengths of the unified SEM framework (uSEM, Kim, Zhu et al. 2007, Gates, Molenaar et al. 2011), which is an extension of SEM to timeseries applications. As part of GIMME, uSEM is used to estimate effective connectivity between pre-specified ROIs on the group-, subgroup- and individual-levels. The uSEM framework incorporates both contemporaneous (*t*) and time-lagged (*t*-1) information between brain regions, which reliably recovers both the existence of a connection and its directionality (Kim, Zhu et al. 2007). Directed connections are estimated contemporaneously, controlling for lagged and autoregressive effects (AR). AR effects indicate the relationship between activity in a single region at time *t* and *t*-1. The benefit of including AR effects, in addition to aiding in reliable estimation of contemporaneous effects for a given brain region, is the ability to estimate path *directionality.* Using a Granger causality framework (Granger 1969), a brain region η_1 is said to Granger-cause activity in another region η_2 if η_1 explains variance in η_2 beyond the variance explained in η_2 by its AR term. Lagged directed connections are also estimated to reduce the chance for spurious contemporaneous effects, but these are not

considered to represent underlying neural signal as the temporal resolution of fMRI data is much lower (i.e., seconds) compared to that of neural activity (i.e., milliseconds, Smith, Miller et al. 2011). Instead, when using fMRI data, directional information due to underlying neural signal is best captured in contemporaneous effects (Granger 1969). For this reason, we focus our analysis of network features on contemporaneous effects. The uSEM model is illustrated below, where A is a matrix of contemporaneous effects, Φ is the matrix of lagged effects with AR effects along the diagonal, η is the observed time series for a given region of interest, and ζ is the residual for each point in time *t*.

$$\eta_t = A\eta_t + \Phi\eta_{t-1} + \zeta_t$$

The GIMME algorithm begins by first conducting a group-level search starting with an initial null model, and paths are iteratively added which contribute to better model fit according to multiple modification indices for the majority of participants in the sample, as defined by the user (here, set at 75%). Specifically, GIMME iteratively counts the number of participants whose model would significantly improve if that path were freely estimated, and the path with the highest count is then added to the group-level model. Typically, this process begins by estimating all AR effects first, as this leads to the best performing model search procedure (Gates, Lane et al. 2017). The search procedure continues until there are no paths that would significantly improve the majority of individuals' models. Next, if the subgroup option is enabled, subgroups consisting of individuals with similar network connectivity patterns are identified using the Walktrap community detection algorithm computed on a sparse count similarity matrix that takes into account the presence of a path and its sign (i.e., positive or negative). The sparse count similarity matrix decidedly outperforms correlation-based similarity matrices

according to simulation studies (Gates, Lane et al. 2017). Subgroup-level maps are constructed using an iterative path-adding approach similar to the group-level approach, using the group-level effects as a prior. GIMME adds paths that improve the model for the greatest number of individuals in the subgroup, which must be at least the majority of the sample (here, 51%). Simulation studies demonstrate that given a sample size of at least 75 participants, GIMME accurately recovers up to four subgroups even in cases where subgroup size is unequal (Gates, Lane et al. 2017). The final step estimates individual-level models by adding any additional paths to the group- and subgroup-level paths to best explain that individual's data. The individual-level model search procedure stops after meeting criteria for excellent model fit for two of four fit indices: comparative fit index (CFI≥.95), non-normed fit index (NNFI≥.95), root-mean-square error of approximation (RMSEA≤.05), and the standardized root-mean-square residual (SRMR≤.05) (Brown 2006). These criteria were also used to identify good individual-level model fit in the current study.

Cluster validation

To determine the validity of the cluster solution arrived at using the Walktrap hierarchical clustering algorithm within the GIMME framework, stability and validity of the cluster solution was evaluated using the R package perturbR (Gates, Fisher et al. 2018). This algorithm incrementally introduces noise to network edges while maintaining the original graph's overall properties and compares resulting cluster solutions with the solution for the original network (i.e., using the full sample). A stable solution will not change drastically given small changes to the network. Quantitatively, a cluster solution is said to be stable if the graph had 20% or more of its edges perturbed before the cluster

solution for the rewired graph is as different as when 20% of the nodes are randomly placed into different clusters. This is quantified by two distinct, but complementary, metrics that describe the degree to which two community solutions differ: Hubert-Arabie Adjusted Rand Index (ARI) and Variation of Information (VI). To ensure the subgrouping solution is not simply capitalizing on chance where no true subgroups exist, a relative measure of cluster solution quality (i.e., modularity) was used to compare the original cluster solution's quality with a solution obtained from a random graph that contains no clusters. The cluster solution was considered valid if modularity for the original solution is greater than or equal to the 95th percentile of modularity obtained from random graphs.

Subgroup characterization: network features

Multiple network metrics were calculated for each individual derived from their connectome data produced by GIMME. Prior work implicates the right anterior insula (rAI) as a hub of causal outflow, interacting with the dACC, dorsolateral prefrontal cortex (dlPFC), ventrolateral prefrontal cortex (vlPFC), and posterior parietal cortex (PPC) (Uddin, Supekar et al. 2011, Supekar and Menon 2012). Therefore, we assessed subgroup differences in out-degree of the rAI normalized by the in-degree (i.e., the number of nodes with which the rAI has an 'outward' contemporaneous connection minus the number of contemporaneous connections that terminate on the rAI). Out- and in-degree was calculated using the R package 'igraph' (Csardi and Nepusz 2006). To characterize whether subgroups were "hyperconnected" or "hypoconnected" relative to one another, total number of contemporaneous connections were compared. Within- and between-network connectivity was calculated as the number of contemporaneous ROI-

ROI connections within- and between-networks, respectively. Integration was indexed with a graph metric called the participation coefficient (PC), which describes the relative distribution of between- and within-network connections for a given node (Guimera and Amaral 2005). A PC approaching 1 indicates that the node's connections are evenly distributed among all networks and a PC of 0 indicates that all of a node's connections are within-network. PC was calculated for each node and averaged within network, resulting in a mean PC value for each network, following the methods of a recent network development study (Marek, Hwang et al. 2015). PC was calculated using the R package 'brainGraph' (Watson 2018).

Results

Preliminary analyses

To quantify the relationship between in-scanner head motion and demographic variables of interest, we characterized the relationship between mean framewise displacement FD (Power, Barnes et al. 2012) for data following ICA-AROMA, age, and diagnostic group. A 2x1 repeated-measures ANOVA demonstrated a significant decrease in mean FD following preprocessing, demonstrating the efficacy of this preprocessing pipeline, $F(1, 260)=33.00$, $p<.001$ (raw: *M(SD)*= 0.20 (0.11), AROMA: *M(SD)*= 0.04 (0.02), Figure 14). There was no difference between ASD, ADHD, and TD groups in mean FD for ICA-AROMA-preprocessed data ($F(2,129)=0.57$, $p=.57$). Similarly, there was no association between age and motion following preprocessing ($r(124)=.01$, $p=.89$). Using the full sample ($N=132$), the GIMME-derived subgroups (which did not align with diagnostic subgroups) differed significantly in mean FD for preprocessed data ($F(1,130)=16.23$, $p<.001$). Therefore, GIMME with subgrouping was rerun using a low

motion subsample to reduce potential confounding effects of motion on subgroup formation. This subsample (n=99) included participants who were at or below the sample's 75th percentile of mean FD (≤0.239). This analysis resulted in subgroups who did not differ on motion for preprocessed data (F(2,96)=0.14, p=.87). Below are the results of subgroup-GIMME with the low motion subsample (n=99).

Subgroup-GIMME results using low motion sample

The group-level model included AR effects for all ROIs, one lagged effect from the right pallidum to the right thalamus, and two contemporaneous effects: right pallidum → right thalamus and lSPL → rSPL (Figure 15). These paths were estimated for all 132 participants.

Model fit. According to approximate fit indices, 98 out of 99 participants' models had good model fit (CFI: *M(SD)*= 0.95 (0.004), NNFI: *M(SD)*= 0.92 (0.007), RMSEA: *M(SD)*= 0.08 (0.007), SRMR: *M(SD)*=.03 (0.004)). The one remaining participant's model had excellent model fit according only to SRMR, demonstrating overall acceptable model fit (CFI=.91, NNFI=.86, RMSEA=.11, SRMR=.05).

Cluster validation. Using the Walktrap hierarchical clustering algorithm, three subgroups emerged (n=26, n=41, n=32; Figure 15). According to VI, ARI, and modularity, the cluster solution attained was not stable nor valid (Figure 16). Based on VI and ARI, only 1% of edges had to be perturbed before 20% of participants were placed into different clusters than the original solution, demonstrating that minor perturbations to the data caused large changes in the clustering solution (Figure 16a and b). Modularity attained (0.02) was not better than expected by chance (95th%ile of random graphs=0.063, Figure 16c), suggesting that clusters are not well defined and participants in different clusters

may have more in common than expected if clusters were truly distinct. Based on these results, we concluded that the clustering solution was untrustworthy. Therefore, we do not report subgroup differences in network nor behavioral features.

Discussion

To contribute to an improved nosology for neurodevelopmental disorders, we attempted to characterize subgroups of children across ASD, ADHD, and TD groups differentiated by the topography of directed functional connectomes as an alternative to traditional, symptom-based diagnostic systems. Unexpectedly, we were unable to uncover a reliable or valid categorical scheme based on connectomes important for cognitive flexibility in a heterogeneous group of children who ranged from above average to clinical impairment in EF. These unanticipated results highlight the heterogeneity of the topology and strength of connectivity of connectomes important for cognitive flexibility in children with and without neurodevelopmental disorders. Further, these results may suggest that the neural substrates of cognitive flexibility in children may not differ categorically and individual differences in children's cognitive flexibility may be better represented dimensionally.

The results of the GIMME analysis, which included 16 ROIs and potentially 256 contemporaneous paths to be estimated, identified only two contemporaneous paths common to all 132 participants: from the left to right SPL and from the right pallidum to right thalamus. Unsurprisingly, these two paths are within-network connections (within the FPN and subcortical networks, respectively), which strengthen from middle childhood to early adulthood (Fair, Nigg et al. 2012). Given this result occurred across a heterogeneous sample of children with ASD, ADHD, and TD children, these connections

may be foundational to network topology in middle childhood, regardless of psychopathology present. The bilateral SPL group-level path is an example of functional homotopy, which is evident throughout the healthy brain and across the lifespan (Zuo, Kelly et al. 2010). Surprisingly, functional homotopy was not observed for the majority of children in the other bilateral regions included in this study: FEF, AI, temporal poles, pallidum, and thalamus. This suggests that there is heterogeneity in functional homotopy in childhood that may be moderated by levels of psychopathology.

Strikingly, an additional 475 subgroup- and individual-level paths were estimated. The paucity of group-level paths highlights the extreme heterogeneity in network topology among children with and without a diagnosed mental health disorder. This is in contrast with many group-based studies of the development of the *undirected* functional connectome, which conclude that network topology is stable by about 8 years of age (Fair, Nigg et al. 2012, Marek, Hwang et al. 2015). Of note, this result cannot simply be attributed to the mixed diagnostic status of the sample. For example, had all TD children exhibited a similar network topology, a stable subgroup would likely have formed to reflect that. Using a novel individual directed connectome estimation technique, we were able to identify large differences in both within- and between-network topology in middle childhood. These results echo recent calls to regard heterogeneity among healthy and patient populations as not only ubiquitous, but adaptive, due to evolutionary processes which result in many "optimal" brain network profiles (Holmes and Patrick 2018).

The large amount of heterogeneity apparent in the directed functional topography of networks important for cognitive flexibility across a mixed group of children may have impeded the formation of a stable clustering solution in this study, which may be

overcome in future studies with a much larger sample size. Simulation studies demonstrate that with a sample size of 75, subgroup recovery is good to excellent using GIMME for up to four subgroups (higher subgroup number was not tested, Gates, Lane et al. 2017). Unfortunately, we had to decrease our sample size from 132 to 99 subjects because the initial clustering solution on the full sample primarily led to clustering based on in-scanner motion. This highlights the importance of considering subgroup differences in nuisance variables such as motion to ensure unsupervised algorithms do not produce subgroups driven by artifacts (Bassett, Xia et al. 2018). Nonetheless, based on the simulation studies it is unlikely that a larger sample size was needed to recover subgroups accurately unless the number of subgroups exceeded 4. This may certainly have been the case, given two DSM-defined diagnostic groups were included, each of which may include multiple subgroups (Georgiades, Szatmari et al. 2013, Stevens, Pearlson et al. 2018). Therefore, the present results do not preclude the existence of 5 subgroups or more within this heterogeneous sample.

Another potential source of the large heterogeneity in functional connectomes observed is spatial variability in the precise location of network nodes across children (Dickie, Ameis et al. 2018). Using templates derived from healthy, young adult samples, Dickie et al. (2018) showed that children show marked variation in the precise location of network nodes and that children with ASD deviate even more than children without a psychiatric diagnosis. Thus, applying ROI coordinates using the Power et al. (2012) parcellation may have led to "missing" true connections due to poor ROI specification on an individual-level, leading to fewer than expected group-level paths.

Despite the above explanations for the unexpected results of this study, a more parsimonious account may be that individual differences in the directed functional connectomes important for cognitive flexibility may best be represented dimensionally instead of categorically. Here, we assumed a categorical structure would best parse heterogeneity in cognitive flexibility in children based on previous studies that identified subgroups present within ASD and ADHD diagnostic categories, which differed in functional connectivity and/or behavioral metrics. For example, subgroups have been shown to exist within children who have ADHD based on differences in functional connectivity of fronto-parietal and reward-related networks (Gates, Molenaar et al. 2014, Costa Dias, Iyer et al.). Subgroups within both ADHD and ASD categories have been demonstrated based on disorder-specific symptoms (Georgiades, Szatmari et al.), neuropsychological task performance (Rommelse, van der Meer et al. 2016, Feczko, Balba et al. 2017), and parent-reports of children's executive functions (Dajani, Llabre et al. 2016). Moreover, neurodevelopmental disorders are traditionally characterized as distinct, categorical entities, which is practical for clinical translation, where categorical decisions must be made to diagnose and provide treatment (Coghill and Sonuga-Barke).

On the other hand, the traditional categorical approach has recently been challenged by mounting evidence for a dimensional taxonomy of psychopathology (Lahey, Krueger et al. 2017). Studies focusing on parent-report and neuropsychological measures of ADHD symptoms consistently conclude that inattention and hyperactivity/impulsivity are dimensional by nature (Haslam, Williams et al. 2006, Frazier, Youngstrom et al. 2007, Lubke, Hudziak et al. 2009, Marcus and Barry 2011). Further, neuroimaging studies demonstrate that psychopathology, conceptualized as a

transdiagnostic ‘p’ factor, dimensionally relates to FPN hypoactivation during a working memory task (Shanmugan, Wolf et al. 2016) and a loss of segregation between the DMN and the FPN and SN at rest (Xia, Ma et al. 2018). Xia et al. (2018) also found a specific relationship between externalizing symptoms (i.e., inattention, hyperactivity/impulsivity, and oppositional defiant symptoms) and stronger SN-FPN coupling at rest. Resting-state fMRI studies focusing on individual disorders have found support for a hybrid categorical/dimensional model in ASD and ADHD based on functional connectivity data (Chabernaud, Mennes et al. 2012, Elton, Alcauter et al. 2014, Elton, Di Martino et al. 2016).

In sum, there is support for both categorical and dimensional models of psychopathology at the behavioral and large-scale neurobiological levels. But, there is a major limitation of the majority of these studies presented, which employ factor or cluster analyses. In these cases, a dimensional or categorical structure is *assumed* to fit the data well, without any formal quantitative tests to determine whether one model is superior to the other. Further complicating matters, clustering algorithms are prone to producing false positives, meaning clusters are produced even in cases where none truly exist. These shortcomings limit the validity of past factor and cluster-based studies of psychopathology. Currently, the only methods that formally determine whether dimensional or categorical models best fit data are taxometric analyses and factor mixture models, which have never before been applied to neuroimaging data, limiting our understanding of how psychopathology should be modeled in consideration of the large-scale neural substrates of behavior. With advances in methodology, future studies may begin to tackle whether a dimensional or categorical model is supported at the brain

network level by applying taxometric or factor mixture modeling studies to functional connectome data.

It is important to note that the unreliable and non-modular clustering results presented here are specific to the *a priori* brain regions supplied to the GIMME algorithm. There are many advantages that the GIMME tool has to offer, including individual-level connectome estimation without contamination by group averaging and the estimation of path direction, but this method is not completely data-driven in that only a limited number of ROIs can be used. Thus, researchers are required to use prior literature to guide ROI selection, possibly leading to missed information about functional network topology outside of the networks examined. It is possible that stable subgroups may have been identified using a different set of ROIs, and thus does not preclude the existence of a categorical structure of psychopathology in consideration of other regions and/or functional networks.

Towards an effort to understand whether psychopathology is best represented categorically or dimensionally at the neurobiological level, several future directions are notable. First, in order to construct an alternative nosology that captures the full range of psychopathology, studies should opt to include a wider swath of diagnostic groups to identify relationships beyond ASD and ADHD symptoms, such as frequently comorbid internalizing symptoms (Zald and Lahey 2017). By capturing multiple symptom types, researchers can test whether the neural substrates of psychopathology operate at multiple hierarchical levels from symptoms to the general 'p' factor (Zald and Lahey 2017). Considering that neurodevelopmental disorders unfold across age, it may be more fruitful to investigate developmental trajectories in place of cross-sectional studies (Morgan,

White et al. 2018). Recent advances in neuroimaging allow for different modalities such as structural, functional, and diffusion-weighted images to be combined in multilayer networks, which may be more informative than one modality alone (Morgan, White et al. 2018).

This study was one of the first to use functional connectome data estimated at the individual level for a heterogeneous group of children spanning TD, ASD, and ADHD diagnoses to test whether the neural substrates of EF and psychopathology follow a categorical or dimensional pattern. Results demonstrated high levels of heterogeneity in the topography of directed functional connectomes important for cognitive flexibility in children with a range of EF abilities. Further, our results did not support a categorical scheme. These results may suggest a dimensional model may better describe individual differences in the neural substrates of cognitive flexibility in children.

Chapter 4. Summary and conclusions

My previous work and numerous other studies have characterized the distribution of cognitive flexibility skills across development, from above-average to impaired levels (Dajani and Uddin 2015, Dajani, Llabre et al. 2016, Buttelmann and Karbach 2017). In accordance with the guiding principles of RDoC, this dissertation took the next logical step to characterize the neural substrates of cognitive flexibility in a healthy sample of adults with no reported mental illness. Even within this "neurotypical" sample, there was great variation in performance on a laboratory-based task of cognitive flexibility and in self-reported measures of cognitive flexibility, some even scoring in the clinically impaired range. Group-averaged task-based activation analyses robustly identified the neural substrates of mature cognitive flexibility skills among regions within the frontoparietal, salience, subcortical, and cerebellar networks, consistent with findings from past neuroimaging studies (Niendam, Laird et al. 2012). Findings based on neurotypical samples such as these can be used to draw conclusions of healthy brain function and be extended to studies of children and typical development. Further, these findings can be used as a benchmark for comparison to draw conclusions of deviations from healthy brain and behavioral development. Accordingly, I capitalized on the results of Study I by using the neurotypical correlates of cognitive flexibility to parse heterogeneity in this crucial skill in children who ranged from impaired to above average in their cognitive flexibility in Study II. The results overwhelmingly demonstrated the heterogeneity in the topology of directed functional brain networks supportive of cognitive flexibility in children with a wide range of executive function skills. Further,

we did not find evidence for a valid subgrouping system according to these directed functional connectomes among a heterogenous group of children.

In consideration of past literature demonstrating the viability of a dimensional and transdiagnostic model of psychopathology across the lifespan (Lahey, Krueger et al. 2017), the results of this dissertation may suggest that, at the large-scale functional brain network level, categorical models of psychopathology may not best describe the data (Frazier, Youngstrom et al. 2007, Neumann, Pappa et al. 2016, Friedman and Miyake 2017, Elliott, Romer et al. 2018, Xia, Ma et al. 2018). Future studies should directly test whether dimensional models of ASD symptoms, ADHD symptoms and executive function are supported by directed functional connectome data in children. In the future, this work may guide development of treatments to improve cognitive flexibility in children, regardless of their DSM-based diagnosis.

Tables

Table 1. Study I: Tests of ecological and divergent validity of the FIST. Zero-order correlations between "real-world" measures of executive function, restricted and repetitive behaviors and the average accuracy-RT metric derived from the fMRI-adapted FIST.

Variable	*M*	*SD*	1	2	3	4	5	6
1. Average Acc-RT	6.50	1.58						
2. BRIEF-A Shift	45.19	9.87	-.06 [-.40, .29]					
3. BRIEF-A Inhibition	46.44	10.27	.07 [-.29, .41]	.22 [-.14, .53]				
4. BRIEF-A Working Memory	48.41	9.07	-.11 [-.44, .25]	.56** [.26, .76]	.60** [.32, .78]			
5. RBQ-2A RMB	1.47	0.47	.10 [-.26, .43]	.35* [.00, .62]	.80** [.62, .90]	.61** [.34, .79]		
6. RBQ-2A IS	1.30	0.30	-.04 [-.38, .32]	.67** [.42, .83]	.44* [.11, .68]	.71** [.48, .85]	.55** [.26, .76]	
7. RBQ-2A Total	1.32	0.29	.05 [-.31, .39]	.53** [.22, .74]	.70** [.46, .84]	.73** [.51, .86]	.87** [.74, .93]	.86** [.73, .93]

Note. Values in square brackets indicate the 95% confidence interval for each correlation. * indicates $p < .05$. ** indicates $p < .01$. RBQ-2A: Adult Repetitive Behavior Questionnaire-2. RMB: repetitive motor behaviors. IS: insistence on sameness.

Table 2. Study I: Behavioral data from fMRI-adapted FIST task.

	Min	Max	Mean	SD
Run 1				
Flexibility Accuracy	0	1.00	0.72	0.23
Control Accuracy	0.10	1.00	0.76	0.23
Flexibility Median RT (ms)	4295.00	7428.00	5744.00	837.00
Flexibility Accuracy-RT	0.00	8.27	5.57	1.94
Run 2				
Flexibility Accuracy	0	1.00	0.84	0.22
Control Accuracy	0.80	1.00	0.95	0.07
Flexibility Median RT (ms)	3884.00	7095.00	5317.00	729.00
Flexibility Accuracy-RT	0.00	9.27	6.66	1.92
Run 3				
Flexibility Accuracy	0.20	1.00	0.83	0.18
Control Accuracy	0.70	1.00	0.94	0.09
Flexibility Median RT (ms)	4003.00	7036.00	5441.00	718.00
Flexibility Accuracy-RT	1.47	8.93	6.53	1.59
Run 4				
Flexibility Accuracy	0.40	1.00	0.88	0.16
Control Accuracy	0.60	1.00	0.95	0.09
Flexibility Median RT (ms)	3275.00	6556.00	5068.00	789.00
Flexibility Accuracy-RT	4.19	9.50	7.22	1.36

Table 3. Study I: Brain regions with significant activation for the Flexibility - Control contrast

Brain region	Cluster size	Peak Z	Peak MNI		
			x	y	z
R medial cerebellar crus II	1174	8.07	6	-70	-30
preSMA/dACC	601	7.85	-2	24	38
L inferior parietal lobule	2314	7.67	-36	-48	38
L temporal occipital fusiform	1293	7.62	-34	-46	-22
R cerebellar crus I	1341	7.49	28	-60	-36
L anterior insula	334	7.40	-32	18	-4
L dlPFC	553	7.31	-50	32	26
L FEF	249	7.29	-26	10	56
Medial occipital cortex	912	6.97	4	-72	8
R anterior insula	295	6.88	32	26	-6
R inferior parietal lobule	332	6.81	34	-66	40
L thalamus	44	6.71	-12	-12	8
L IFJ	10	6.51	-46	10	30
R thalamus	201	6.41	10	-8	8
R globus pallidus	8	6.39	16	-4	-2
R thalamus	2	6.34	10	-16	12
R dlPFC	18	6.06	44	36	16
R FEF	19	6.03	26	14	48
R hippocampus	5	5.89	24	-24	-8
L frontal pole	2	5.81	-48	44	8
R cerebellar crus I	2	5.80	14	-78	-20
R dlPFC	1	5.79	50	38	18

Note. Results are voxel-level thresholded at FWE-corrected $p<.05$. Clusters defined using FSL's "cluster" command using a $Z>5.75$ threshold (except for two clusters which required a higher threshold, $Z>6.3$, to break up into anatomically distinct regions: IFJ/dlPFC and R thalamus/L thalamus/basal ganglia). Coordinates in white matter, brainstem, or outside of brain are not included in this table. Results are organized by peak Z value in descending order.

Table 4. Study I: Results for the Control - Flexibility contrast

Brain region	Cluster size	Peak Z	Peak MNI		
			x	y	z
L cerebellar crus II	174	6.99	-22	-78	-36
L STG	2382	6.86	-52	0	-14
mPFC/frontal pole	4002	6.73	4	56	14
L angular gyrus	990	6.58	-56	-58	28
L PCC	585	6.32	-8	-46	26
R cerebellar crus II	47	6.18	26	-78	-36
R hippocampus	164	6.01	24	-8	-24
L hippocampus	67	5.72	-24	-12	-22
L planum temporale	157	5.41	-64	-24	12
R precentral gyrus	123	5.32	2	-30	58
L temporal pole	17	5.26	-34	4	-22
L temporal pole	1	4.79	-28	2	-26
L temporal pole	1	4.69	-40	8	-22
L cingulate gyrus	90	5.25	0	-22	42
R frontal pole	7	5.07	22	62	28
R temporal pole	1	4.76	28	8	-28
L central opercular cortex	1	4.69	-40	8	-22

Note. Results are voxel-level thresholded at FWE-corrected $p<.05$. Clusters defined using FSL's "cluster" command using a $Z>5.75$ threshold. Coordinates in white matter, brainstem, or outside of brain are not included in this table. Results are organized by peak Z value in descending order.

Table 5. Study II: Sample demographics

	Diagnostic groups			
	TD *n*=53	ADHD *n*=43	ASD *n*=36	*P* value
Sex	39 M/14 F	32 M/11 F	28 M/8 F	.90
Age	10.37 (1.04)	9.93 (1.22)	10.52 (1.38)	.09
range	[8.00 - 12.58]	[8.00 - 12.33]	[8.00 - 12.92]	
Race[a]	7, 4, 9, 33	7, 0, 8, 28	2, 0, 3, 29	.10
Ethnicity, *No. Hispanic/Latino*	1	5	4	.14
FSIQ[b]	118.23 (13.25)	109.00 (11.95)	103.17 (12.45)	<.001
range	[90 - 147]	[87 - 136]	[73 - 131]	
Motion[c]	0.17 (0.09)	0.22 (0.12)	0.21 (0.12)	.11
Handedness[d]**,** *No. L,A,R*	4, 1, 47	5, 0, 38	4, 0, 31	.74

a: Numbers for each of the following racial categories presented in the following order: African American, Asian, Biracial, Caucasian, b: FSIQ: WISC-IV full-scale IQ, c: Mean framewise displacement for raw rs-fMRI data calculated in FSL, d: Number of children with left, ambidextrous, right handedness.

Table 6. Study II: Diagnostic information

	Primary diagnosis		
	TD	ADHD	ASD
	n=53	*n*=43	*n*=36
Secondary Dx, No. (%)	**1 (1.9%)[a]**	**20 (46.5%)**	**30 (83.3%)**
ADHD-I	0	8 (18.6%)	9 (25.0%)
ADHD-C	0	34 (79.1%)	12 (33.3%)
ODD	0	18 (41.9%)	9 (25.0%)
Simple Phobia	1 (1.9%)	5 (11.6%)	8 (22.2%)
GAD	0	0	5 (13.9%)
OCD	0	0	4 (11.1%)
Dysthymia	0	0	1 (2.8%)
CD,MDD,Mania,Panic,Som,Sep Ax[b]	0	0	0
ADHD Measures[c], *M* (*SD*) *[range]*			
Conners Hyper/Impulsive	47.79 (5.51)	70.55 (12.54)	66.22 (10.85)
Conners Inattention	45.35 (4.71)	73.02 (8.2)	66.78 (10.7)
Conners 3 Hyper/Impulsive	44.93 (7.15)	75.50 (11.74)	80.00 (9.55)
Conners 3 Inattention	43.00 (8.82)	77.20 (9.57)	83.50 (4.85)
ADHD Hyperactivity	0.22 (0.59) [0-3]	4.00 (2.89) [0-9]	3.76 (2.23) [0-8]
ADHD Inattention	0.17 (0.53) [0-3]	7.00 (1.91) [2-9]	5.76 (2.75) [0-9]
ASD Measures, *M (SD)*			
ADI A	--	--	21.03 (5.91)
ADI B	--	--	15.71 (4.99)
ADI C	--	--	6.21 (2.09)
ADOS-2 Social Affect	--	--	7.06 (3.25)
ADOS-2 RRB	--	--	4.41 (1.23)
ADOS-G CS	--	--	11.9 (2.90)
ADOS-G RRB	--	--	3.42 (1.68)

[a]One TD participant had a past simple phobia of dogs, [b]No child had a diagnosis of conduct disorder, major depressive disorder, mania or hypomania, panic disorder, somatization disorder, or separation anxiety disorder. [c]Reporting n=104 for Conners (1997) and n=26 with Conners-3 (2008); missing data on Conners n=2; Missing data on ADHD Home Rating Scale IV n=3; missing data on ADI for n=2 ASD participants; n=17 ASD participants have ADOS-2 data and n=19 ASD participants have ADOS-G data.

Table 7. Study II: Regions of interest used in GIMME analysis

	Brain region	MNI		
		x	y	z
ROI#	*Frontoparietal network*			
1	L SPL	-28	-58	48
2	R SPL	33	-53	44
3	L IFJ	-47	11	23
4	L FEF	-23	11	64
5	R FEF	32	14	56
	Salience network			
6	L anterior insula	-35	20	0
7	R anterior insula	36	22	3
8	dorsal ACC	5	23	37
	Default mode network			
9	vmPFC	6	54	16
10	R temporal pole	46	16	-30
11	L temporal pole	-53	3	-27
12	PCC	8	-48	31
	Subcortical network			
13	L pallidum	-15	4	8
14	R pallidum	15	5	7
15	L thalamus	-10	-18	7
16	R thalamus	9	-4	6

Note. All regions in the FPN, SN, and subcortical network were active in response to flexibility trials during the Flexible Item Selection Task (Study I). Regions within the DMN were more active for control trials compared with flexibility trials. Specific coordinates were chosen from the Power et al. (2011) parcellation.

Figures

Figure 1. Study I: fMRI-adapted 4-Match FIST. Participants were asked to choose 3 successive pairs of cards that "go together in one way" ("Now you choose"). During control trials, the correct card pairs were highlighted by a thick black border, eliminating the need to enable cognitive flexibility, while still controlling for lower-level visual and motor processes ("Follow along"). Fixation trials had jittered presentation times optimized for a fast event-related design. The computer-based task included only flexibility trials.

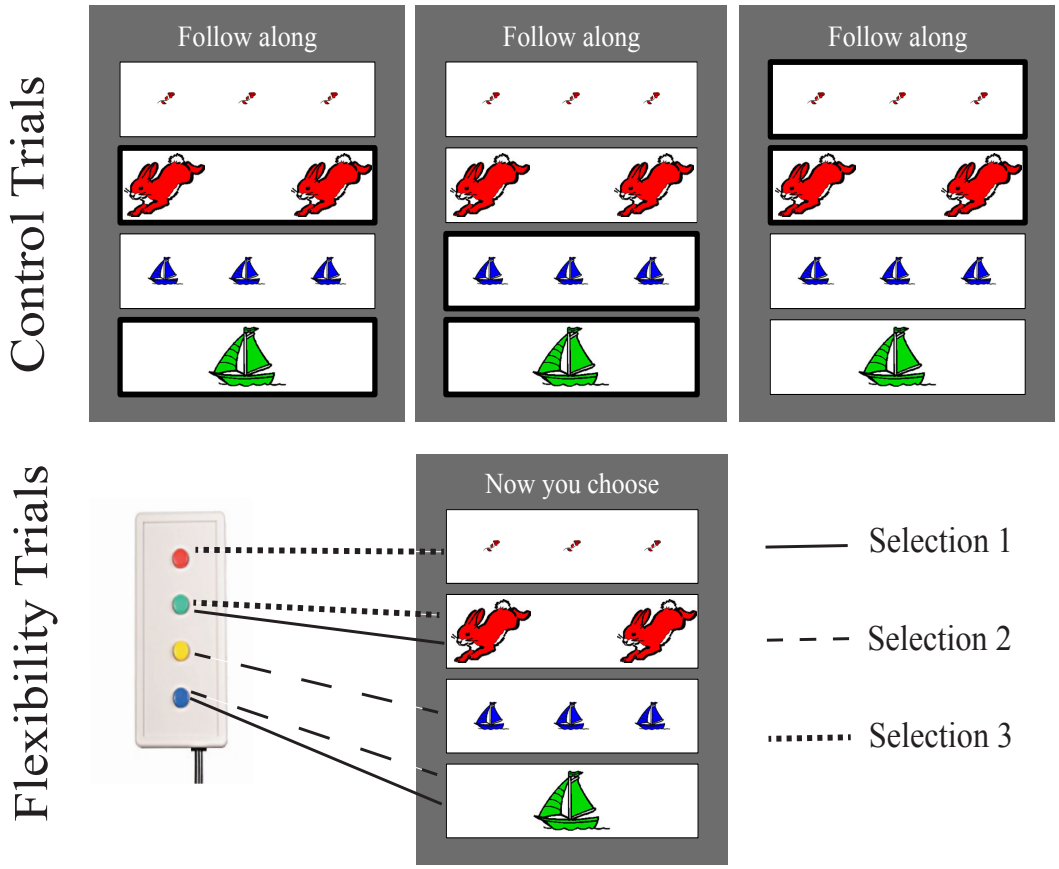


Figure 2. Study I: A schematic of the fMRI-adapted 4-Match FIST.

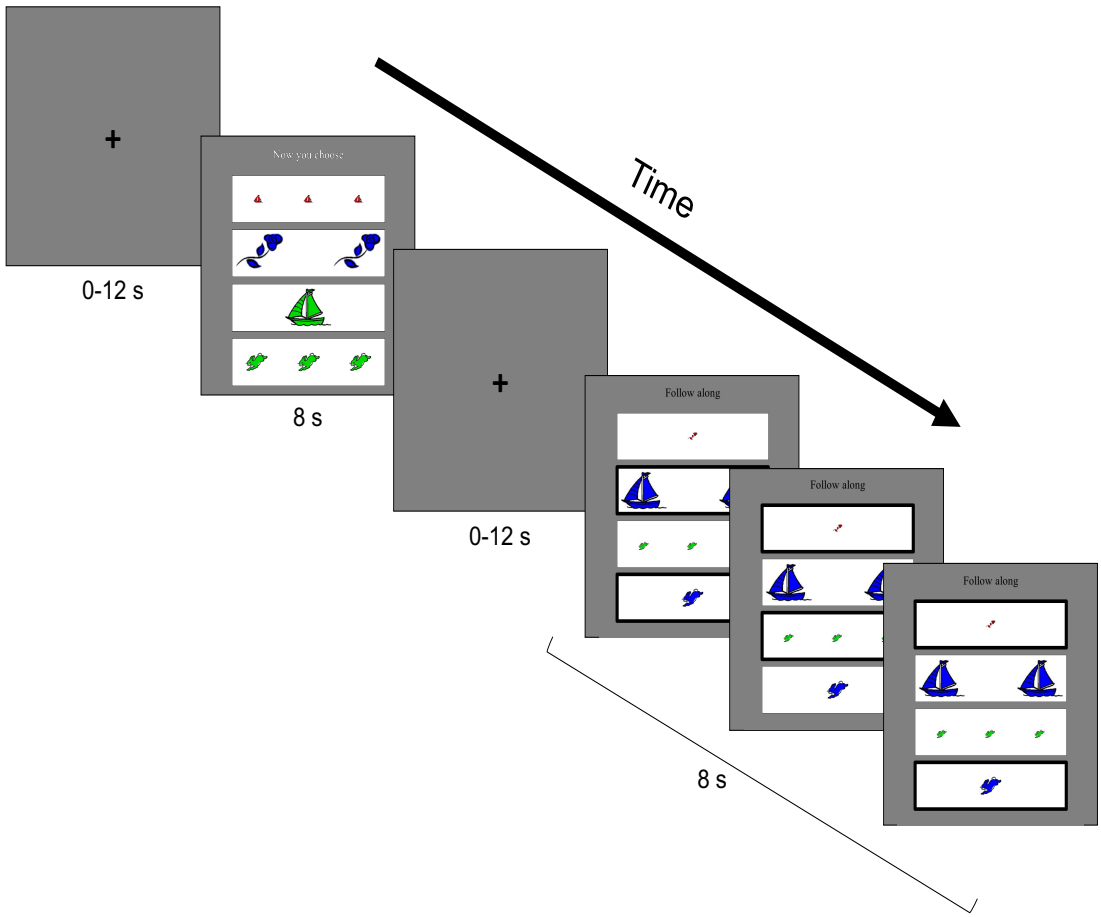


Figure 3. Study I: Convergent validity between fMRI-adapted and computer-based 4-Match FIST. $r(30)=.33$, $p = .07$.

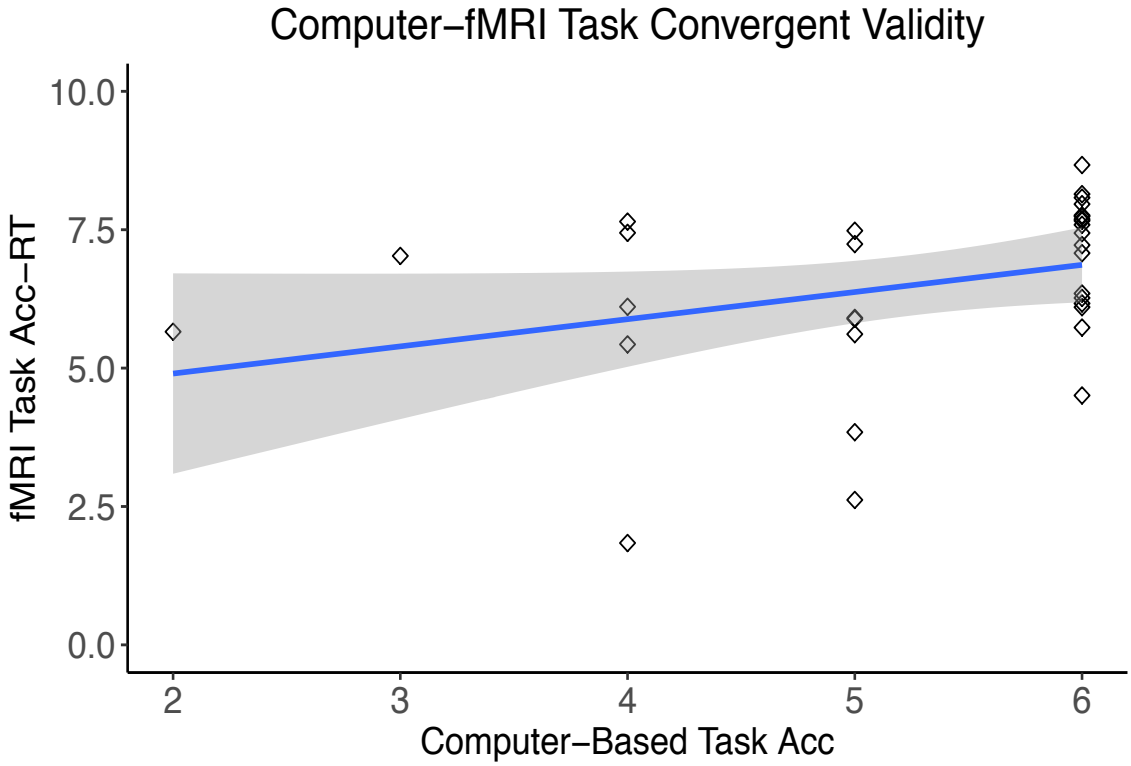


Figure 4. Study I: Ecological validity of the fMRI-adapted 4-Match FIST. The Acc-RT metric from the fMRI-adapted FIST is plotted against self-reported cognitive flexibility (blue), working memory (black), and inhibition scores (red; measured with the BRIEF T scores). All correlations are non-significant ($p > .05$).

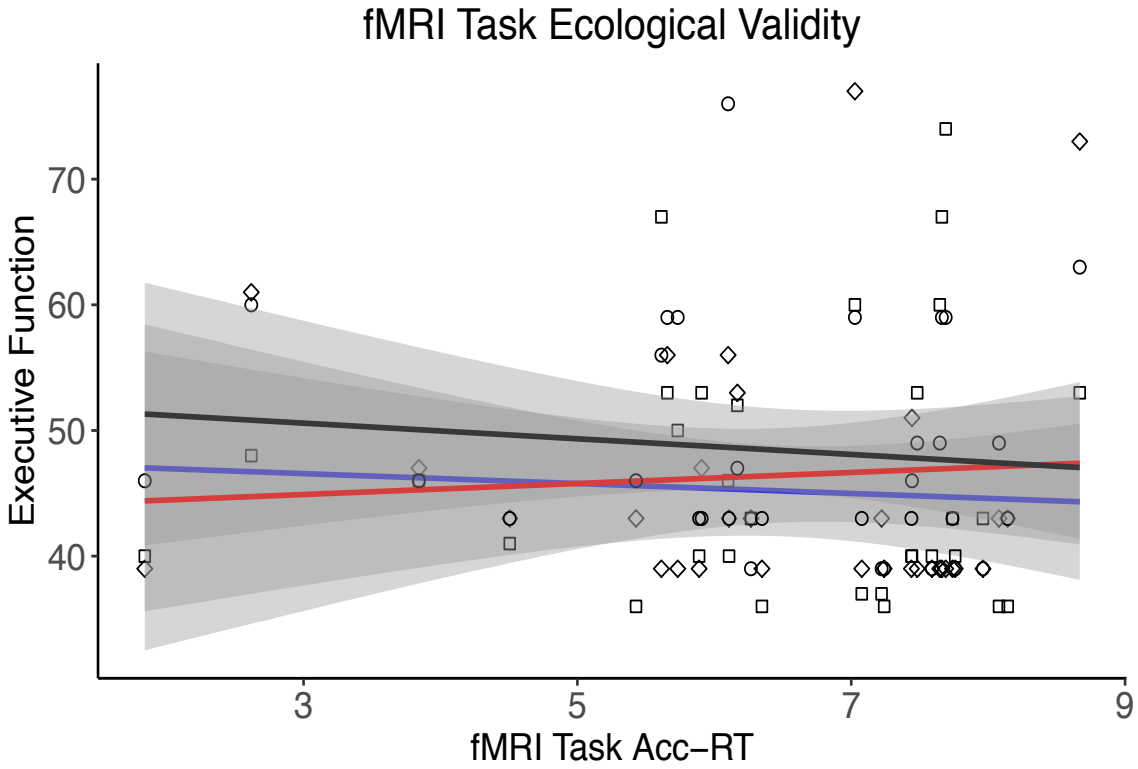


Figure 5. Study I: Selection-level computer-based task accuracy. A box plot of trial-level accuracy for each selection. Horizontal bars indicate the median and dots represent outliers, whose values lie outside of the interquartile range of the sample. An accuracy score of 6 represents a perfect score (100%).

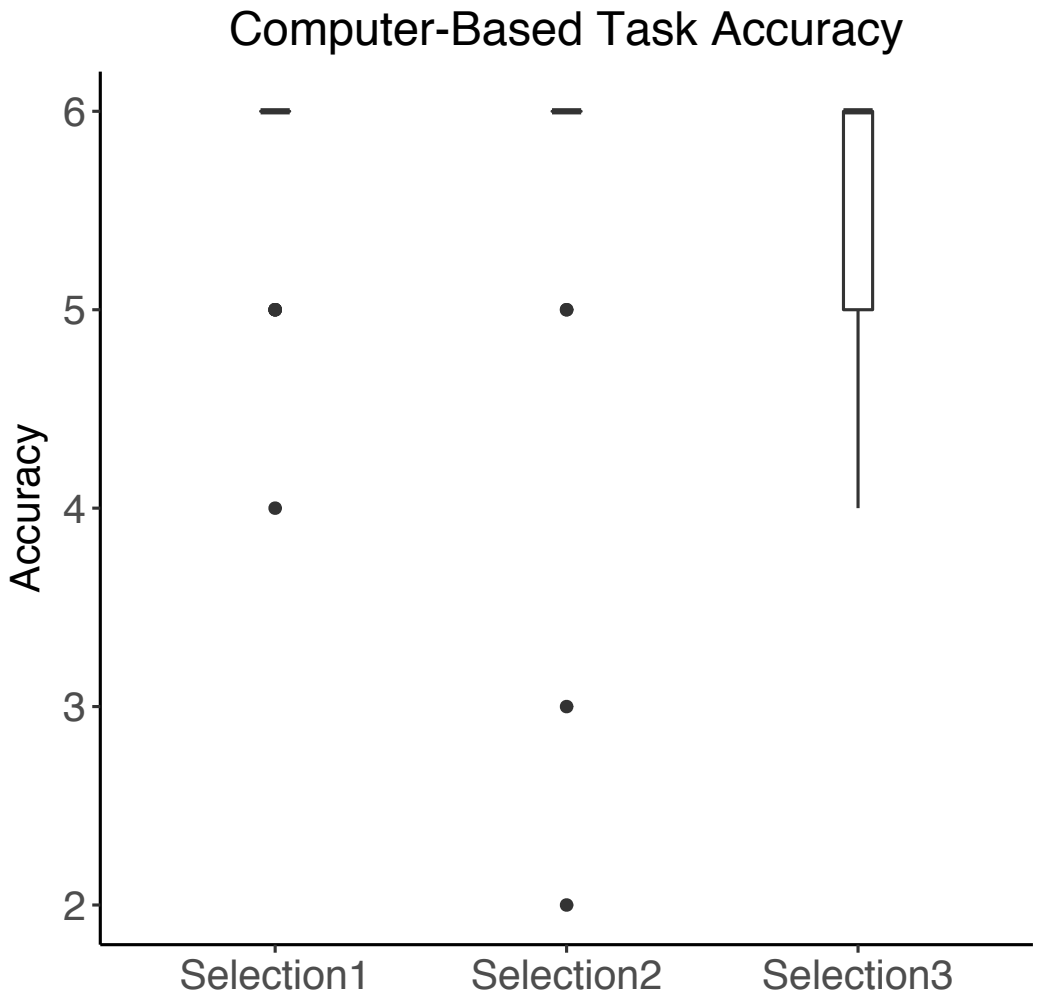


Figure 6. Study I: Trial-level computer-based task accuracy. Histogram of trial-level accuracy where the blue dashed line represents the sample mean. An accuracy score of 18 represents a 100% accuracy across all selections.

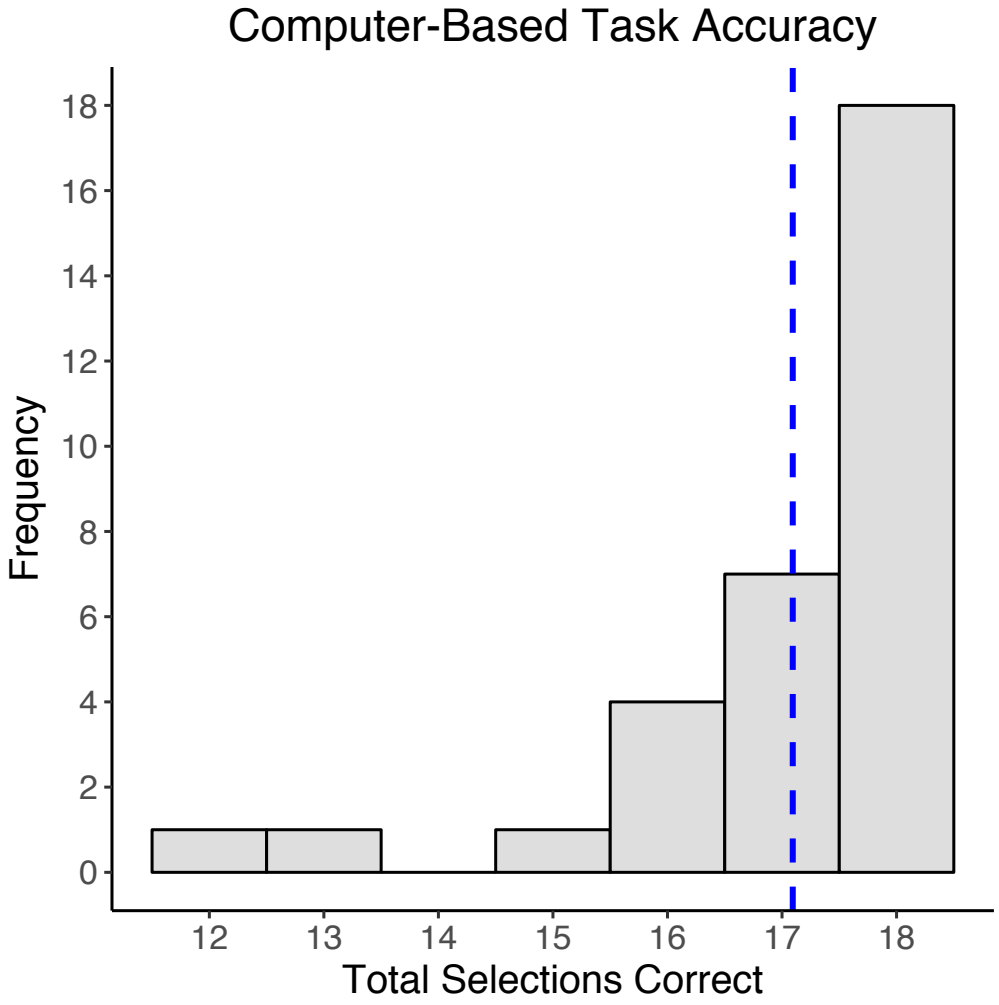


Figure 7. Study I: Frequency of dimension choice. A boxplot based on data from the computer-based FIST is displayed: the number of trials that included a selection of a particular dimension type (in grey, outliers black points) and number of trials where a particular dimension was chosen for each selection (in color, only interquartile range displayed). Color was most frequently chosen on the first selection. Number was least frequently chosen for any given trial, and if chosen, was most frequently chosen on the third selection.

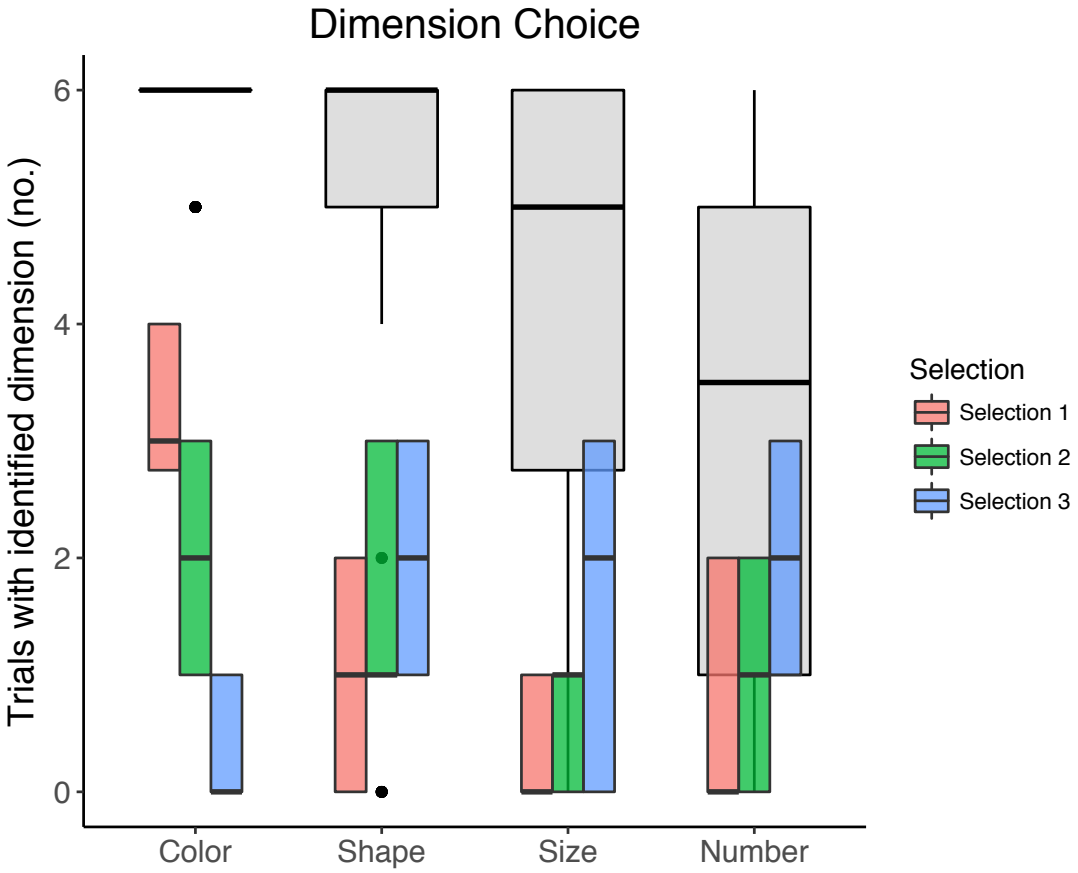


Figure 8. Study I: Frequency of dimension choice per selection. Box plot of the number of trials that participants chose a particular dimension for selections 1, 2 and 3.

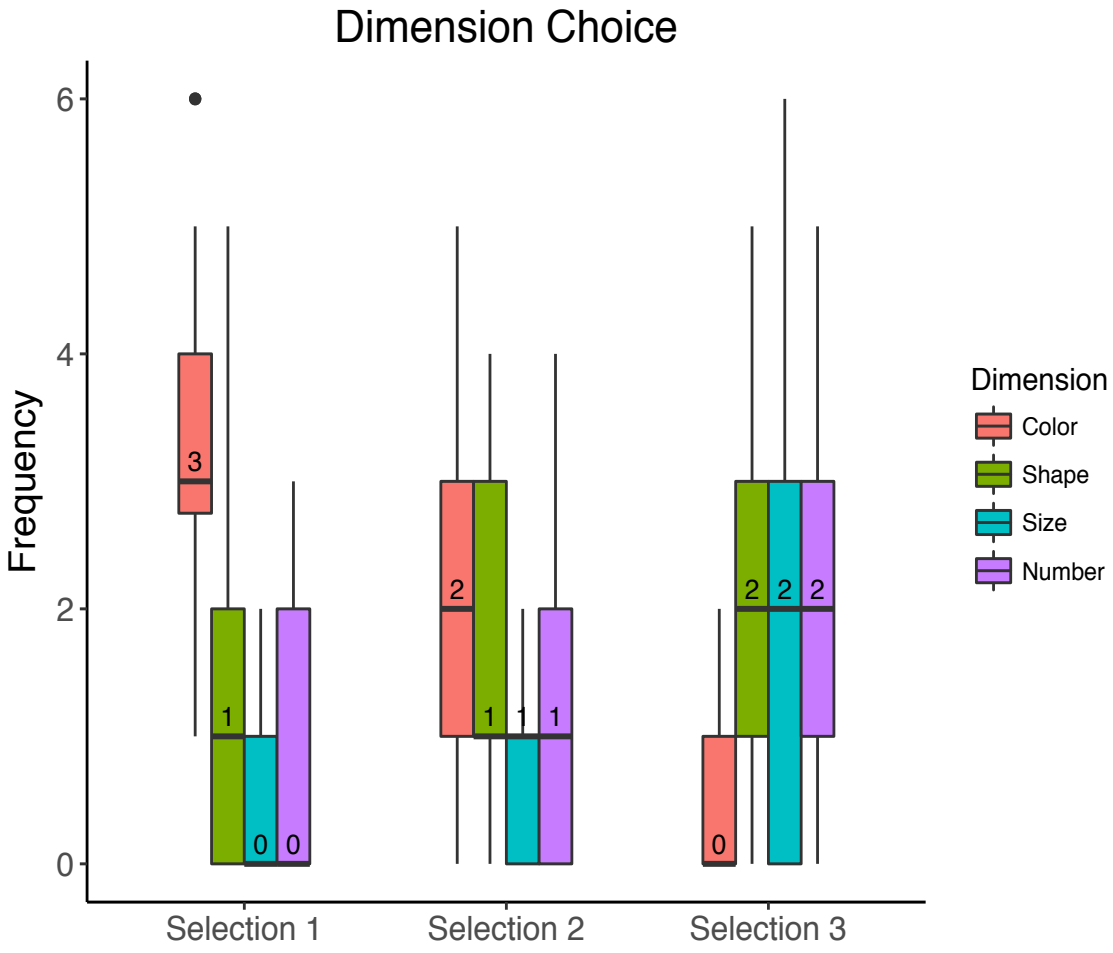


Figure 9. Study I: Accuracy for flexibility and control trials for fMRI-adapted 4-Match FIST. Violin plot (in color) with overlay of boxplot (in white) displayed.

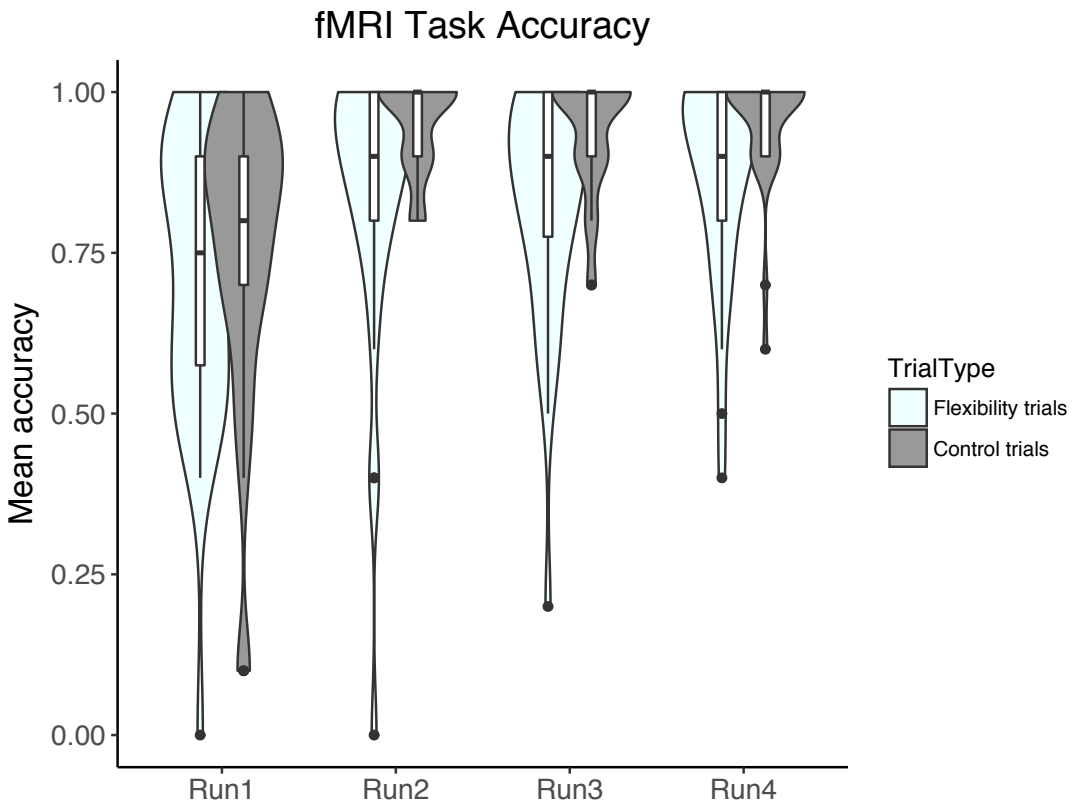


Figure 10. Study I: Accuracy-RT scores for each run of the fMRI-adapted FIST. Scores changed across runs in a cubic fashion, demonstrating an improvement in performance over time. Red diamonds indicate mean scores for each run.

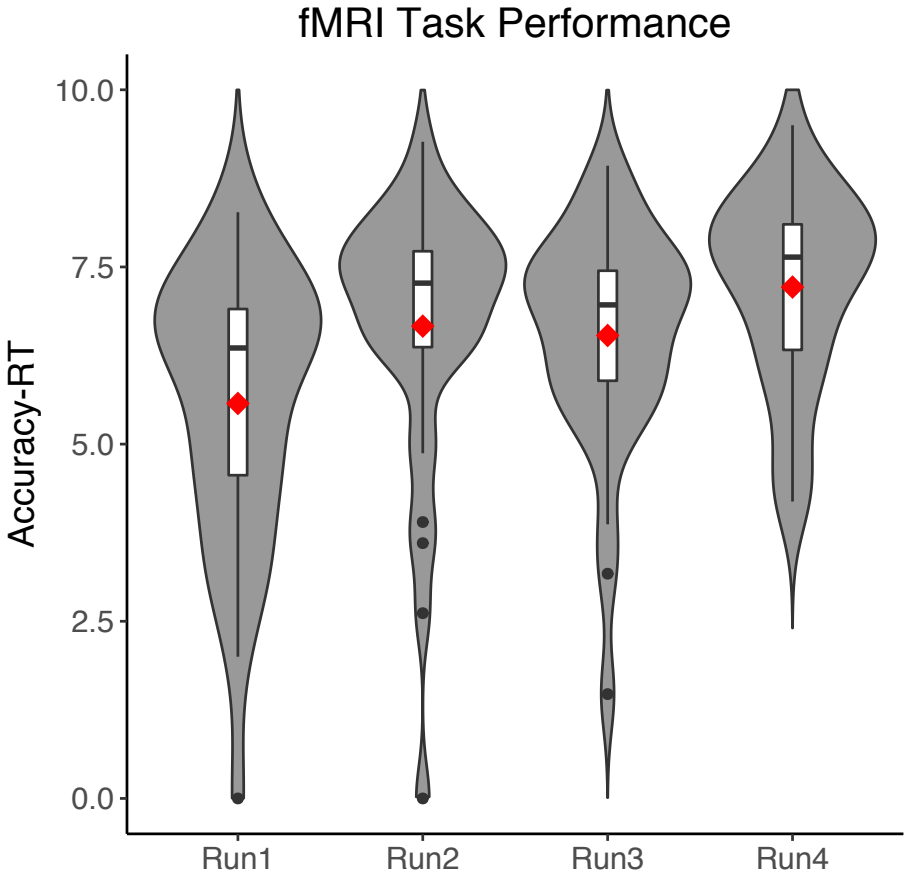


Figure 11. Group activation maps for the main effect of flexibility and control trials**.** Z-maps are voxel-level thresholded at an FWE-corrected $p<.05$. The "overlap" panel displays thresholded maps that include both activations and deactivations.

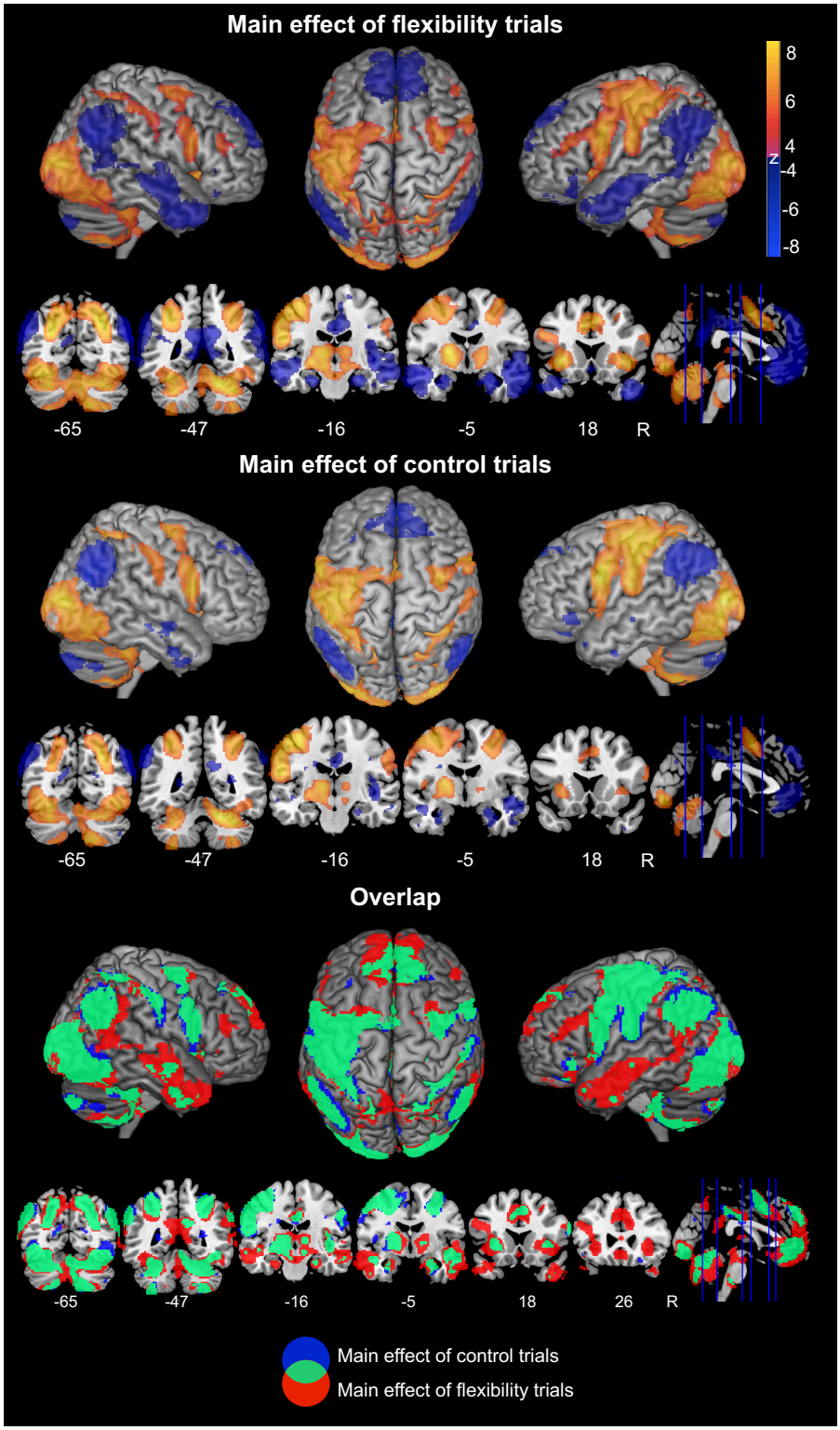


Figure 12. Study I: Group activation maps for the Flexibility – Control and Control – Flexibility contrasts**.** Z-maps are voxel-level thresholded at an FWE-corrected $p<.05$.

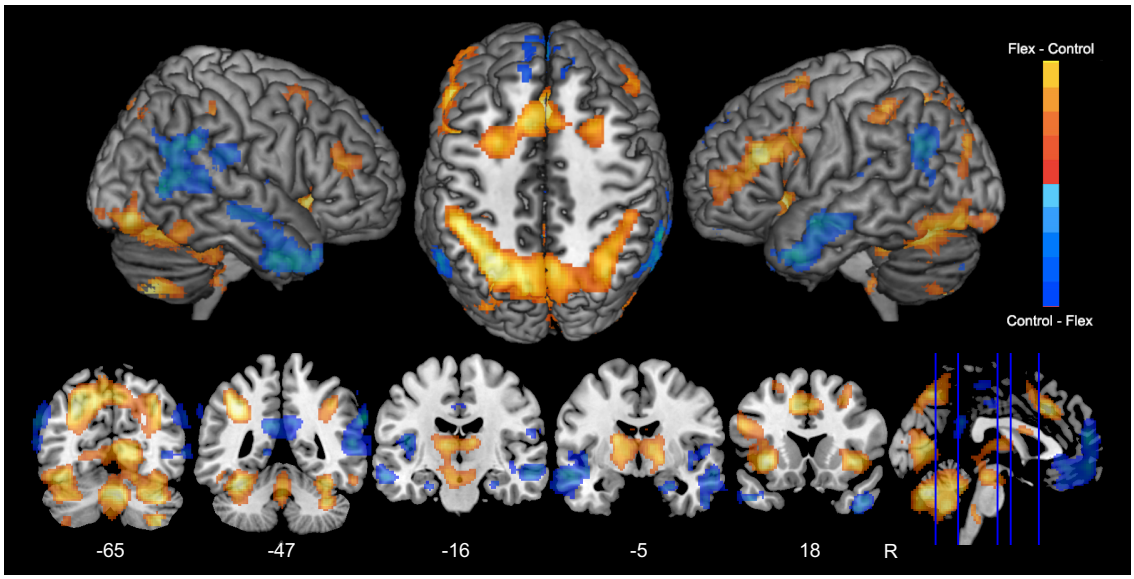
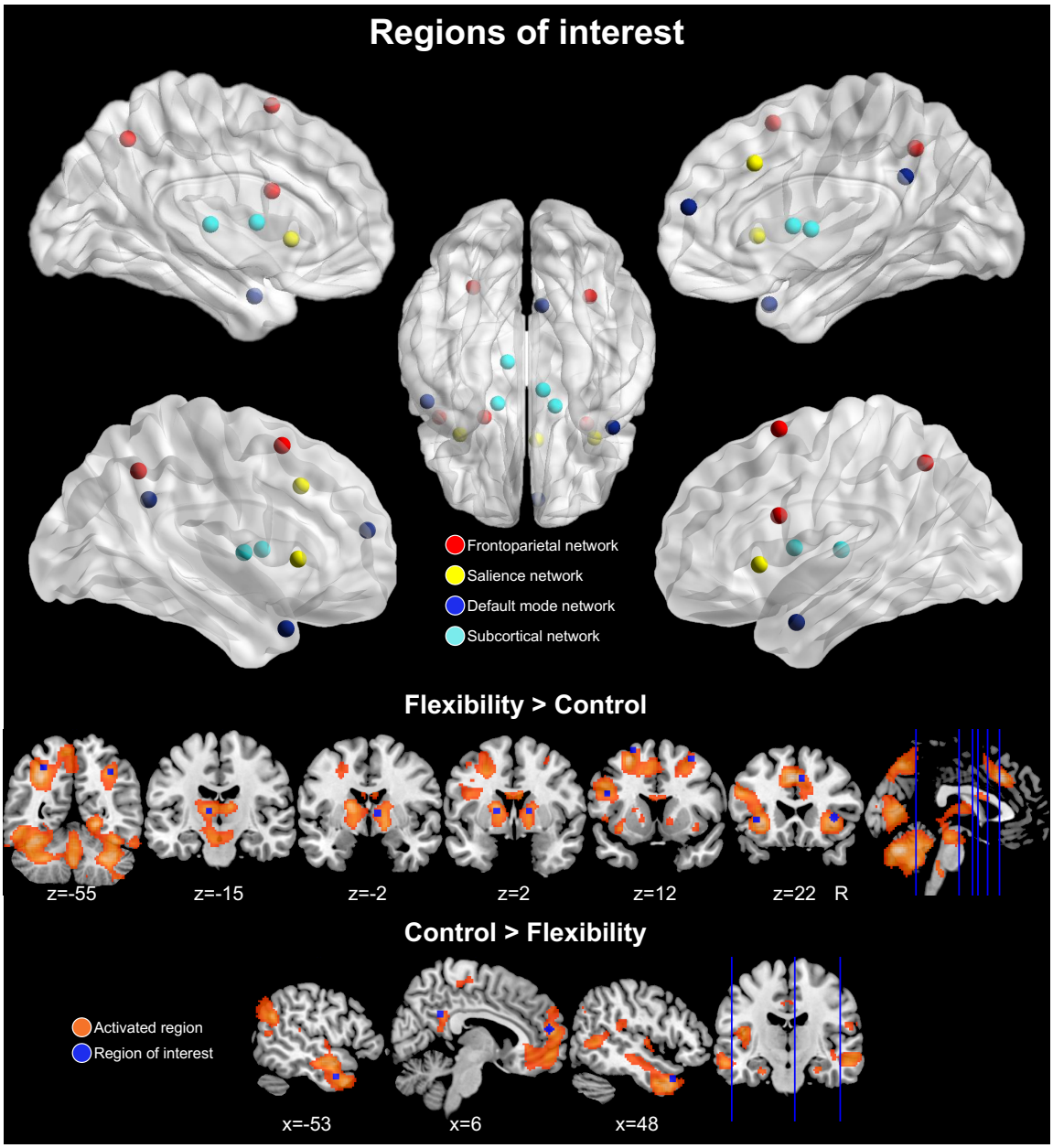


Figure 13. Study II: Regions of interest.



Sixteen regions of interest used in the GIMME network connectivity analysis. Regions were chosen based on their activation (Flexibility > Control) or deactivation (Control > Flexibility) in response to the Flexible Item Selection Task (Study I). Specific coordinates were based on the Power et al. (2011) parcellation to facilitate categorization within large-scale brain networks (frontoparietal, salience, default mode, and subcortical). Brain nodes in upper panel were visualized with BrainNet Viewer (http://www.nitrc.org/projects/bnv/, Cao, Wang et al. 2014).

Figure 14. Study II: Differences in motion estimates in raw and pre-processed data. AROMA denotes data that has undergone motion artifact removal with ICA-AROMA. Data shown for the entire sample (*N*=132).

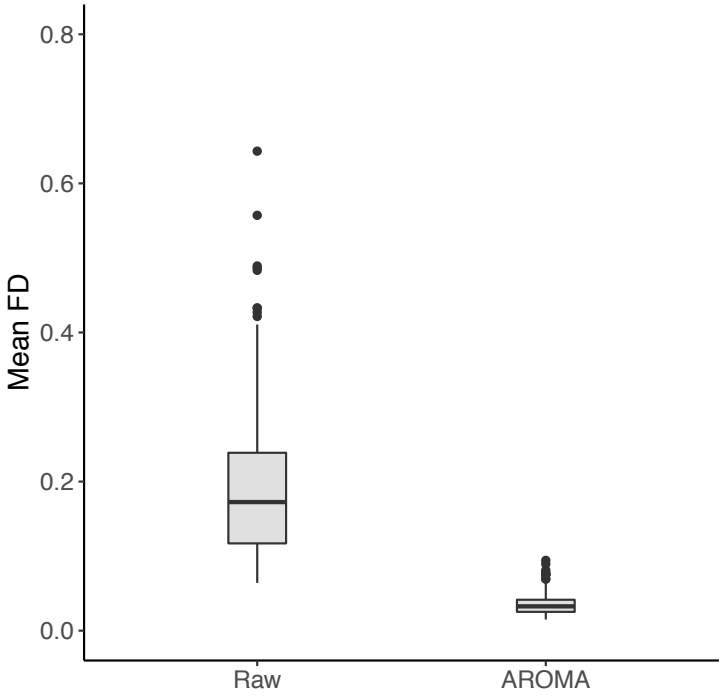


Figure 15. Study II: Subgroup-GIMME network models for each subgroup.

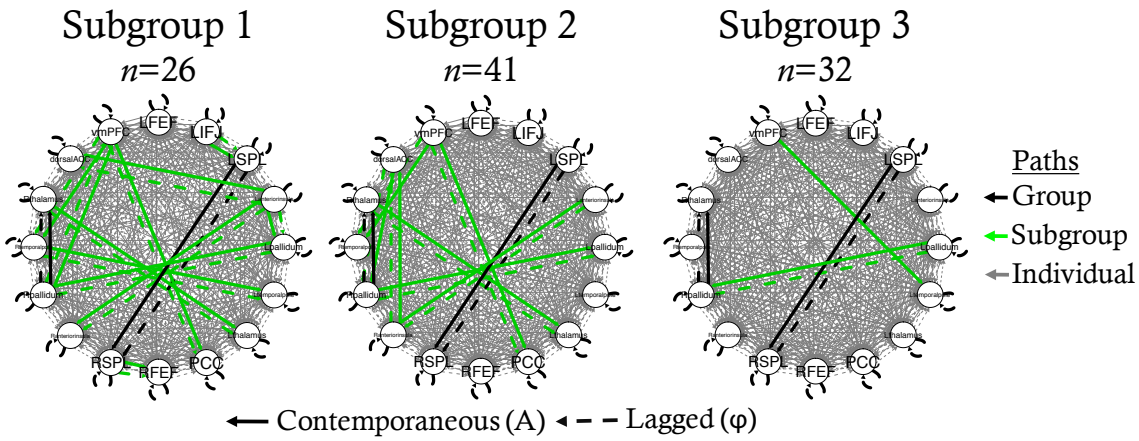
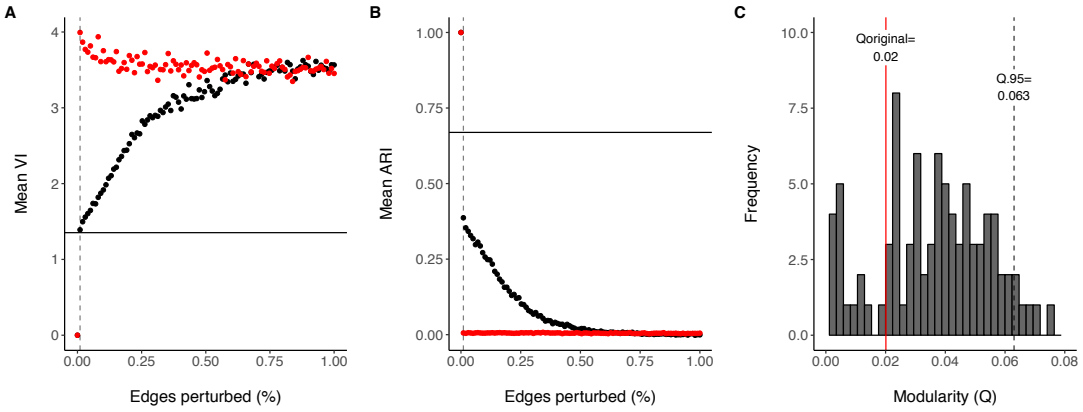


Figure 16. Study II: Cluster validation. Results from both VI and ARI demonstrate that the clustering solution is not stable. For panels A and B, the black horizontal line represents the point at which 20% of participants were placed into different clusters than the original solution (20% of nodes perturbed). The dashed vertical line identifies the point at which the perturbed graph reached 20% of nodes perturbed. Black dots represent the perturbed graph based on the original clustering solution while the red dots represent a perturbed random graph. Panel C demonstrates that modularity for the original clustering solution (0.02) was not better than expected by chance (>.06), suggesting this clustering solution is not valid.



References

Achenbach, T. (1991). Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles, Department of Psychiatry, University of Vermont.

Achenbach, T. and L. Rescorla (2001). "Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment Burlington, VT: University of Vermont." Research Center for Children, Youth, & Families.

American Psychiatric Association (2013). DSM 5, American Psychiatric Association.

Anderson, P. (2002). "Assessment and development of executive function (EF) during childhood." Child Neuropsychology **8**(2): 71-82.

Anderson, V., Jacobs, R., Anderson, P.J. (2008). Executive functions and the frontal lobes: A lifespan perspective, Taylor & Francis.

Anticevic, A., M. W. Cole, J. D. Murray, P. R. Corlett, X. J. Wang and J. H. Krystal (2012). "The role of default network deactivation in cognition and disease." Trends in Cognitive Sciences **16**(12): 584-592.

Armbruster, D. J. N., K. Ueltzhoffer, U. Basten and C. J. Fieback (2012). "Prefrontal cortical mechanisms underlying individual differences in cognitive flexibility and stability." Journal of Cognitive Neuroscience **24**(12): 2385-2399.

Badre, D. and A. D. Wagner (2006). "Computational and neurobiological mechanisms underlying cognitive flexibility." Proceedings of the National Academy of Sciences **103**(18): 7186-7191.

Bailey, C. E. (2007). "Cognitive accuracy and intelligent executive function in the brain and in business." Annals of the New York Academy of Sciences **1118**(1): 122-141.

Barrett, S. L., M. Uljarević, E. K. Baker, A. L. Richdale, C. R. Jones and S. R. Leekam (2015). "The Adult Repetitive Behaviours Questionnaire-2 (RBQ-2A): a self-report measure of restricted and repetitive behaviours." Journal of Autism and Developmental Disorders **45**(11): 3680-3692.

Bassett, D. S., C. H. Xia and T. D. Satterthwaite (2018). "Understanding the emergence of neuropsychiatric disorders with network neuroscience." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **3**(9): 742-753.

Blijd-Hoogewys, E. M., M. L. Bezemer and P. L. van Geert (2014). "Executive functioning in children with ASD: an analysis of the BRIEF." Journal of Autism and Developmental Disorders **44**(12): 3089-3100.

Bodfish, J. W., F. J. Symons, D. E. Parker and M. H. Lewis (2000). "Varieties of repetitive behavior in autism: Comparisons to mental retardation." Journal of Autism and Developmental Disorders **30**(3): 237-243.

Brown, T. A. (2006). Confirmatory factor analysis for applied research, New York, Guilford Press.

Brunswik, E. (1956). Perception and the representative design of psychological experiments, Univ of California Press.

Bull, R., K. A. Espy and S. A. Wiebe (2008). "Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years." Developmental Neuropsychology **33**(3): 205-228.

Burgess, P. W., N. Alderman, C. Forbes, A. Costello, L. MA Coates, D. R. Dawson, N. D. Anderson, S. J. Gilbert, I. Dumontheil and S. Channon (2006). "The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology." Journal of the International Neuropsychological Society **12**(02): 194-209.

Burt, K. B. and A. A. Paysnick (2012). "Resilience in the transition to adulthood." Development and Psychopathology **24**(2): 493-505.

Buttelmann, F. and J. Karbach (2017). "Development and plasticity of cognitive flexibility in early and middle childhood." Frontiers in Psychology **8**: 1040.

Campbell, D. T. and D. W. Fiske (1959). "Convergent and discriminant validation by the multitrait-multimethod matrix." Psychological Bulletin **56**(2): 81.

Cao, M., J. H. Wang, Z. J. Dai, X. Y. Cao, L. L. Jiang, F. M. Fan, X. W. Song, M. R. Xia, N. Shu, Q. Dong, M. P. Milham, F. X. Castellanos, X. N. Zuo and Y. He (2014). "Topological organization of the human brain functional connectome across the lifespan." Developmental Cognitive Neuroscience **7**: 76-93.

Chabernaud, C., M. Mennes, C. Kelly, K. Nooner, A. Di Martino, F. X. Castellanos and M. P. Milham (2012). "Dimensional brain-behavior relationships in children with attention-deficit/hyperactivity disorder." Biological Psychiatry **71**(5): 434-442.

Chan, R. C., D. Shum, T. Toulopoulou and E. Y. Chen (2008). "Assessment of executive functions: review of instruments and identification of critical issues." Archives of Clinical Neuropsychology **23**(2): 201-216.

Chen, E. Y., C. L. Hui, E. L. Dunn, M. Y. Miao, W. S. Yeung, C. K. Wong, W. F. Chan and W. N. Tang (2005). "A prospective 3-year longitudinal study of cognitive predictors of relapse in first-episode schizophrenic patients." Schizophrenia Research **77**(1): 99-104.

Chen, Q., W. Yang, W. Li, D. Wei, H. Li, Q. Lei, Q. Zhang and J. Qiu (2014). "Association of creative achievement with cognitive flexibility by a combined voxel-based morphometry and resting-state functional connectivity study." NeuroImage **102**: 474-483.

Cicchetti, D. V. (1994). "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology." Psychological Assessment **6**(4): 284.

Cicchetti, D. V. and S. S. Sparrow (1990). Developmental assessment in clinical child psychology: A handbook, Elmsford, NY, US, Pergamon Press**.**

Coghill, D. and E. J. Sonuga-Barke (2012). "Annual research review: categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders--implications of recent empirical study." Journal of Child Psychology and Psychiatry **53**(5): 469-489.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn, Erlbaum Associates, Hillsdale.

Cohen, J. D., W. M. Perlstein, T. S. Braver, L. E. Nystrom, D. C. Noll, J. Jonides and E. E. Smith (1997). "Temporal dynamics of brain activation during a working memory task." Nature **386**(6625): 604-608.

Collette, F., M. Van der Linden, S. Laureys, G. Delfiore, C. Degueldre, A. Luxen and E. Salmon (2005). "Exploring the unity and diversity of the neural substrates of executive functioning." Human Brain Mapping **25**(4): 409-423.

Conners, C. K. (1997). Conners' Rating Scales--revised: User's Manual, Multi-Health Systems, Incorporated.

Conners, C. K. (2008). Conners 3rd edition: Manual, Multi-Health Systems.

Costa Dias, T. G., S. P. Iyer, S. D. Carpenter, R. P. Cary, V. B. Wilson, S. H. Mitchell, J. T. Nigg and D. A. Fair (2015). "Characterizing heterogeneity in children with and without ADHD based on reward system connectivity." Developmental Cognitive Neuroscience **11**: 155-174.

Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research.

Cubillo, A., R. Halari, C. Ecker, V. Giampietro, E. Taylor and K. Rubia (2010). "Reduced activation and inter-regional functional connectivity of fronto-striatal networks in adults with childhood Attention-Deficit Hyperactivity Disorder (ADHD) and persisting symptoms during tasks of motor inhibition and cognitive switching." Journal of Psychiatric Research **44**(10): 629-639.

Cuthbert, B. N. (2014). "The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology." World Psychiatry **13**(1): 28-35.

Cuthbert, B. N. and T. R. Insel (2013). "Toward the future of psychiatric diagnosis: the seven pillars of RDoC." BMC medicine **11**(1): 126.

D'Cruz, A. M., M. E. Ragozzino, M. W. Mosconi, S. Shrestha, E. H. Cook and J. A. Sweeney (2013). "Reduced behavioral flexibility in autism spectrum disorders." Neuropsychology **27**(2): 152-160.

Dajani, D. R., M. M. Llabre, M. B. Nebel, S. H. Mostofsky and L. Q. Uddin (2016). "Heterogeneity of executive functions among comorbid neurodevelopmental disorders." Scientific Reports **6**: 36566.

Dajani, D. R. and L. Q. Uddin (2015). "Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience." Trends in Neurosciences **38**(9): 571-578.

Dennis, E. L., N. Jahanshad, K. L. McMahon, G. I. de Zubicaray, N. G. Martin, I. B. Hickie, A. W. Toga, M. J. Wright and P. M. Thompson (2014). "Development of insula connectivity between ages 12 and 30 revealed by high angular resolution diffusion imaging." Human Brain Mapping **35**(4): 1790-1800.

Derrfuss, J., M. Brass, J. Neumann and D. Y. von Cramon (2005). "Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies." Human Brain Mapping **25**(1): 22-34.

Di Martino, A., X. N. Zuo, C. Kelly, R. Grzadzinski, M. Mennes, A. Schvarcz, J. Rodman, C. Lord, F. X. Castellanos and M. P. Milham (2013). "Shared and distinct intrinsic functional network centrality in autism and attention-deficit/hyperactivity disorder." Biological Psychiatry **74**(8): 623-632.

Diamond, A., W. S. Barnett, J. Thomas and S. Munro (2007). "Preschool program improves cognitive control." Science **318**(5855): 1387.

Diamond, A. and K. Lee (2011). "Interventions shown to aid executive function development in children 4 to 12 years old." Science **333**(6045): 959-964.

Dick, A. S. (2014). "The development of cognitive flexibility beyond the preschool period: an investigation using a modified Flexible Item Selection Task." Journal of Experimental Child Psychology **125**: 13-34.

Dickie, E. W., S. H. Ameis, S. Shahab, N. Calarco, D. E. Smith, D. Miranda, J. D. Viviano and A. N. Voineskos (2018). "Personalized intrinsic network topography mapping and functional connectivity deficits in autism spectrum disorder." Biological Psychiatry.

Di Martino, A., Damien A. Fair, C. Kelly, Theodore D. Satterthwaite, F. X. Castellanos, Moriah E. Thomason, R. C. Craddock, B. Luna, Bennett L. Leventhal, X.-N. Zuo and Michael P. Milham (2014). "Unraveling the miswired connectome: A developmental perspective." Neuron **83**(6): 1335-1353.

Dougherty, C. C., D. W. Evans, S. M. Myers, G. J. Moore and A. M. Michael (2016). "A comparison of structural brain imaging findings in autism spectrum disorder and attention-deficit hyperactivity disorder." Neuropsychology Review **26**(1): 25-43.

DuPaul, G. J., T. J. Power, A. D. Anastopoulos and R. Reid (1998). ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation, Guilford Press New York.

Eaton, W. W., C. Smith, M. Ybarra, C. Muntaner and A. Tien (2004). "Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R)."

Ehring, T., U. Zetsche, K. Weidacker, K. Wahl, S. Schönfeld and A. Ehlers (2011). "The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking." Journal of Behavior Therapy and Experimental Psychiatry **42**(2): 225-232.

Eklund, A., T. E. Nichols and H. Knutsson (2016). "Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates." Proceedings of the National Academy of Sciences **113**(28): 7900-7905.

Elliott, M. L., A. Romer, A. R. Knodt and A. R. Hariri (2018). "A connectome-wide functional signature of transdiagnostic risk for mental illness." Biological Psychiatry **84**(6): 452-459.

Elton, A., S. Alcauter and W. Gao (2014). "Network connectivity abnormality profile supports a categorical-dimensional hybrid model of ADHD." Human Brain Mapping **35**(9): 4531-4543.

Elton, A., A. Di Martino, H. C. Hazlett and W. Gao (2016). "Neural connectivity evidence for a categorical-dimensional hybrid model of autism spectrum disorder." Biological Psychiatry **80**(2): 120-128.

Engel de Abreu, P. M., N. Abreu, C. C. Nikaedo, M. L. Puglisi, C. J. Tourinho, M. C. Miranda, D. M. Befi-Lopes, O. F. Bueno and R. Martin (2014). "Executive functioning and reading achievement in school: a study of Brazilian children assessed by their teachers as “poor readers”." Frontiers in Psychology **5**: 550.

Fair, D. A., D. Bathula, M. A. Nikolas and J. T. Nigg (2012). "Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD." Proceedings of the National Academy of Sciences **109**(17): 6769-6774.

Fair, D. A., J. T. Nigg, S. Iyer, D. Bathula, K. L. Mills, N. U. Dosenbach, B. L. Schlaggar, M. Mennes, D. Gutman, S. Bangaru, J. K. Buitelaar, D. P. Dickstein, A. Di Martino, D. N. Kennedy, C. Kelly, B. Luna, J. B. Schweitzer, K. Velanova, Y. F. Wang, S. Mostofsky, F. X. Castellanos and M. P. Milham (2012). "Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data." Frontiers in Systems Neuroscience **6**: 80.

Farrant, B. M., M. T. Maybery and J. Fletcher (2012). "Language, cognitive flexibility, and explicit false belief understanding: longitudinal analysis in typical development and specific language impairment." Child Development **83**(1): 223-235.

Feczko, E., N. Balba, O. Miranda-Dominguez, M. Cordova, S. L. Karalunas, L. Irwin, D. V. Demeter, A. P. Hill, B. H. Langhorst, J. Grieser Painter, J. Van Santen, E. J. Fombonne, J. L. Nigg and D. A. Fair (2017). "Subtyping cognitive profiles in Autism Spectrum Disorder using a random forest algorithm." NeuroImage.

Fox, M. D., M. Corbetta, A. Z. Snyder, J. L. Vincent and M. E. Raichle (2006). "Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems." Proceedings of the National Academy of Sciences **103**(26): 10046-10051.

Frazier, T. W., E. A. Youngstrom and R. I. Naugle (2007). "The latent structure of attention-deficit/hyperactivity disorder in a clinic-referred sample." Neuropsychology **21**(1): 45-64.

Frazier, T. W., E. A. Youngstrom, L. Sinclair, C. S. Kubu, P. Law, A. Rezai, J. N. Constantino and C. Eng (2010). "Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample." Assessment **17**(3): 308-320.

Frazier, T. W., E. A. Youngstrom, L. Speer, R. Embacher, P. Law, J. Constantino, R. L. Findling, A. Y. Hardan and C. Eng (2012). "Validation of proposed DSM-5 criteria for autism spectrum disorder." Journal of the American Academy of Child & Adolescent Psychiatry **51**(1): 28-40. e23.

Friedman, N. P. and A. Miyake (2017). "Unity and diversity of executive functions: Individual differences as a window on cognitive structure." Cortex **86**: 186-204.

Friedman, N. P., A. Miyake, J. L. Robinson and J. K. Hewitt (2011). "Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis." Developmental Psychology **47**(5): 1410-1430.

Garvey, M., S. Avenevoli and K. Anderson (2016). "The National Institute of Mental Health Research Domain Criteria and Clinical Research in Child and Adolescent Psychiatry." Journal of the American Academy of Child & Adolescent Psychiatry **55**(2): 93-98.

Gates, K. M., Z. Fisher and C. Arizmendi (2018). perturbR: Random Perturbation of Count Matrices. R package version 0.1.2.

Gates, K. M., S. T. Lane, E. Varangis, K. Giovanello and K. Guiskewicz (2017). "Unsupervised classification during time-series model building." Multivariate Behav Res **52**(2): 129-148.

Gates, K. M. and P. C. Molenaar (2012). "Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples." NeuroImage **63**(1): 310-319.

Gates, K. M., P. C. Molenaar, F. G. Hillary and S. Slobounov (2011). "Extended unified SEM approach for modeling event-related fMRI data." NeuroImage **54**(2): 1151-1158.

Gates, K. M., P. C. Molenaar, S. P. Iyer, J. T. Nigg and D. A. Fair (2014). "Organizing heterogeneous samples using community detection of GIMME-derived resting state functional networks." PloS one **9**(3): e91322.

Genet, J. J. and M. Siemer (2011). "Flexible control in processing affective and non-affective material predicts individual differences in trait resilience." Cognition and Emotion **25**(2): 380-388.

Georgiades, S., P. Szatmari, M. Boyle, S. Hanna, E. Duku, L. Zwaigenbaum, S. Bryson, E. Fombonne, J. Volden, P. Mirenda, I. Smith, W. Roberts, T. Vaillancourt, C. Waddell, T. Bennett, A. Thompson and A. S. D. S. T. Pathways in (2013). "Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach." Journal of Child Psychology and Psychiatry **54**(2): 206-215.

Ghirardi, L., I. Brikell, R. Kuja-Halkola, C. M. Freitag, B. Franke, P. Asherson, P. Lichtenstein and H. Larsson (2018). "The familial co-aggregation of ASD and ADHD: a register-based cohort study." Mol Psychiatry **23**(2): 257-262.

Gioia, G. A., P. K. Isquith, S. C. Guy and L. Kenworthy (2000). Behavior Rating Inventory of Executive Function: BRIEF, Psychological Assessment Resources Odessa, FL.

Gioia, G. A., P. K. Isquith, S. C. Guy and L. Kenworthy (2000). "Test review Behavior Rating Inventory of Executive Function." Child Neuropsychology **6**(3): 235-238.

Gioia, G. A., P. K. Isquith, P. D. Retzlaff and K. A. Espy (2002). "Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample." Child Neuropsychology **8**(4): 249-257.

Granger, C. W. (1969). "Investigating causal relations by econometric models and cross-spectral methods." Econometrica: Journal of the Econometric Society: 424-438.

Guimera, R. and L. A. N. Amaral (2005). "Cartography of complex networks: modules and universal roles." Journal of Statistical Mechanics: Theory and Experiment **2005**(02): P02001.

Halassa, M. M. and S. Kastner (2017). "Thalamic functions in distributed cognitive control." Nature Neuroscience **20**(12): 1669-1679.

Hartwigsen, G., N. E. Neef, J. A. Camilleri, D. S. Margulies and S. B. Eickhoff (2018). "Functional segregation of the right inferior frontal gyrus: evidence from coactivation-based parcellation." Cerebral Cortex.

Haslam, N., B. Williams, M. Prior, R. Haslam, B. Graetz and M. Sawyer (2006). "The latent structure of attention-deficit/hyperactivity disorder: A taxometric analysis." Australian and New Zealand Journal of Psychiatry **40**(8): 639-647.

Holmes, A. J. and L. M. Patrick (2018). "The myth of optimality in clinical neuroscience." Trends in Cognitive Sciences **22**(3): 241-257.

Hunter, S. J. and E. P. Sparrow (2012). Executive function and dysfunction: Identification, assessment and treatment, Cambridge University Press.

Insel, T., B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow and P. Wang (2010). "Research domain criteria (RDoC): toward a new classification framework for research on mental disorders." American Journal of Psychiatry **167**(7): 748-751.

Jacques, S. and P. D. Zelazo (2001). "The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers." Developmental Neuropsychology **20**(3): 573-591.

Jeste, S. S. and D. H. Geschwind (2014). "Disentangling the heterogeneity of autism spectrum disorder through genetic findings." Nature Reviews Neurology **10**(2): 74-81.

Kapp, S. K., A. Gantman and E. A. Laugeson (2011). Transition to adulthood for high-functioning individuals with autism spectrum disorders. A comprehensive book on autism spectrum disorders, InTech.

Kelly, R. M. and P. L. Strick (2003). "Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate." Journal of Neuroscience **23**(23): 8432-8444.

Kenworthy, L., L. G. Anthony, D. Q. Naiman, L. Cannon, M. C. Wills, C. Luong-Tran, M. A. Werner, K. C. Alexander, J. Strang, E. Bal, J. L. Sokoloff and G. L. Wallace (2014). "Randomized controlled effectiveness trial of executive function intervention for children on the autism spectrum." Journal of Child Psychology and Psychiatry **55**(4): 374-383.

Kernbach, J. M., T. D. Satterthwaite, D. S. Bassett, J. Smallwood, D. Margulies, S. Krall, P. Shaw, G. Varoquaux, B. Thirion, K. Konrad and D. Bzdok (2018). "Shared endo-phenotypes of default mode dsfunction in attention deficit/hyperactivity disorder and autism spectrum disorder." Transl Psychiatry **8**(1): 133.

Kim, C., S. E. Cilles, N. F. Johnson and B. T. Gold (2012). "Domain general and domain preferential brain regions associated with different types of task switching: a meta-analysis." Human Brain Mapping **33**(1): 130-142.

Kim, J., W. Zhu, L. Chang, P. M. Bentler and T. Ernst (2007). "Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data." Human Brain Mapping **28**(2): 85-93.

Konishi, S., T. Hayashi, I. Uchida, H. Kikyo, E. Takahashi and Y. Miyashita (2002). "Hemispheric asymmetry in human lateral prefrontal cortex during cognitive set shifting." Proceedings of the National Academy of Sciences **99**(11): 7803-7808.

Konishi, S., K. Nakajima, I. Uchida, M. Kameyama, K. Nakahara, K. Sekihara and Y. Miyashita (1998). "Transient activation of inferior prefrontal cortex during cognitive set shifting." Nature Neuroscience **1**(1): 80-84.

Kuder, G. F. and M. W. Richardson (1937). "The theory of the estimation of test reliability." Psychometrika **2**(3): 151-160.

Kvavilashvili, L. and J. Ellis (2004). "Ecological validity and twenty years of real-life/laboratory controversy in memory research: A critical (and historical) review." History and Philosophy of Psychology.

Lahey, B. B., R. F. Krueger, P. J. Rathouz, I. D. Waldman and D. H. Zald (2017). "A hierarchical causal taxonomy of psychopathology across the life span." Psychological Bulletin **143**(2): 142-186.

Lane, S. T. and K. M. Gates (2017). "Automated selection of robust individual-level structural equation models for time series data." Structural Equation Modeling: A Multidisciplinary Journal **24**(5): 768-782.

Leber, A. B., N. B. Turk-Browne and M. M. Chun (2008). "Neural predictors of moment-to-moment fluctuations in cognitive flexibility." Proceedings of the National Academy of Sciences **105**(36): 13592-13597.

Leitner, Y. (2014). "The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know?" Frontiers in Human Neuroscience **8**: 268.

Lenroot, R. K. and P. K. Yeung (2013). "Heterogeneity within Autism Spectrum Disorders: what have we learned from neuroimaging studies?" Frontiers in Human Neuroscience **7**: 733.

Licinio, J. and M. L. Wong (2013). "A novel conceptual framework for psychiatry: vertically and horizontally integrated approaches to redundancy and pleiotropism that co-exist with a classification of symptom clusters based on DSM-5." Mol Psychiatry **18**(8): 846-848.

Lieb, R. W. and A. M. Bohnert (2017). "Relations between executive functions, social impairment, and friendship quality on adjustment among high functioning youth with Autism Spectrum Disorder." Journal of Autism and Developmental Disorders **47**(9): 2861-2872.

Liu, J., X. Liao, M. Xia and Y. He (2018). "Chronnectome fingerprinting: Identifying individuals and predicting higher cognitive functions using dynamic brain connectivity patterns." Human Brain Mapping **39**(2): 902-915.

Lopez, B. R., A. J. Lincoln, S. Ozonoff and Z. Lai (2005). "Examining the relationship between executive functions and restricted, repetitive symptoms of Autistic Disorder." Journal of Autism and Developmental Disorders **35**(4): 445-460.

Lord, C., S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles and M. Rutter (2000). "The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism." Journal of Autism and Developmental Disorders **30**(3): 205-223.

Lord, C., M. Rutter, P. C. DiLavore, S. Risi, K. Gotham and S. L. Bishop (2012). Autism diagnostic observation schedule: ADOS-2, Western Psychological Services Los Angeles, CA.

Loth, E., W. Spooren, L. M. Ham, M. B. Isaac, C. Auriche-Benichou, T. Banaschewski, S. Baron-Cohen, K. Broich, S. Bölte and T. Bourgeron (2016). "Identification and validation of biomarkers for autism spectrum disorders." Nature Reviews Drug Discovery **15**(1): 70.

Lubke, G. H., J. J. Hudziak, E. M. Derks, T. C. van Bijsterveldt and D. I. Boomsma (2009). "Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum." Journal of the American Academy of Child & Adolescent Psychiatry **48**(11): 1085-1093.

Marcus, D. K. and T. D. Barry (2011). "Does attention-deficit/hyperactivity disorder have a dimensional latent structure? A taxometric analysis." J Abnorm Psychol **120**(2): 427-442.

Marek, S., K. Hwang, W. Foran, M. N. Hallquist and B. Luna (2015). "The contribution of network organization and integration to the development of cognitive control." PLoS Biology **13**(12): e1002328.

Martel, M. M., P. M. Pan, M. S. Hoffmann, A. Gadelha, M. C. do Rosario, J. J. Mari, G. G. Manfro, E. C. Miguel, T. Paus, R. A. Bressan, L. A. Rohde and G. A. Salum (2017). "A general psychopathology factor (P factor) in children: Structural model analysis and external validation through familial risk and child global executive function." J Abnorm Psychol **126**(1): 137-148.

Mazefsky, C. A., R. Anderson, C. M. Conner and N. Minshew (2011). "Child Behavior Checklist Scores for school-aged children with autism: Preliminary evidence of patterns suggesting the need for referral." J Psychopathol Behav Assess **33**(1): 31-37.

McTeague, L. M., M. S. Goodkind and A. Etkin (2016). "Transdiagnostic impairment of cognitive control in mental illness." J Psychiatr Res **83**: 37-46.

Menon, V. (2011). "Large-scale brain networks and psychopathology: a unifying triple network model." Trends in Cognitive Sciences **15**(10): 483-506.

Middleton, F. A. and P. L. Strick (2000). "Basal ganglia and cerebellar loops: motor and cognitive circuits." Brain Research Reviews **31**(2-3): 236-250.

Miller, H. L., M. E. Ragozzino, E. H. Cook, J. A. Sweeney and M. W. Mosconi (2015). "Cognitive set shifting deficits and their relationship to repetitive behaviors in autism spectrum disorder." Journal of Autism and Developmental Disorders **45**(3): 805-815.

Morgan, S. E., S. R. White, E. T. Bullmore and P. E. Vertes (2018). "A network neuroscience approach to typical and atypical brain development." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.

Mostert, J. C., M. Hoogman, A. M. Onnink, D. van Rooij, D. von Rhein, K. J. van Hulzen, J. Dammers, C. C. Kan, J. K. Buitelaar, D. G. Norris and B. Franke (2015). "Similar subgroups based on cognitive performance parse heterogeneity in adults with ADHD and healthy controls." Journal of Attention Disorders.

Neumann, A., I. Pappa, B. B. Lahey, F. C. Verhulst, C. Medina-Gomez, V. W. Jaddoe, M. J. Bakermans-Kranenburg, T. E. Moffitt, I. M. H. van and H. Tiemeier (2016). "Single Nucleotide Polymorphism heritability of a general psychopathology factor in children." Journal of the American Academy of Child & Adolescent Psychiatry **55**(12): 1038-1045 e1034.

Niendam, T. A., A. R. Laird, K. L. Ray, Y. M. Dean, D. C. Glahn and C. S. Carter (2012). "Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions." Cogn Affect Behav Neurosci **12**(2): 241-268.

Pelphrey, K. (2017). "Charting a Course for Autism Biomarkers." Biological Psychiatry **82**(3): 155-156.

Pessoa, L. (2014). "Understanding brain networks and brain organization." Physics of Life Reviews **11**(3): 400-435.

Petrides, M. (2000). "The role of the mid-dorsolateral prefrontal cortex in working memory." Exp Brain Res **133**(1): 44-54.

Pine, D. S. (2017). "Clinical Advances From a Computational Approach to Anxiety." Biological Psychiatry **82**(6): 385-387.

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

Power, J. D., A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar and S. E. Petersen (2011). "Functional network organization of the human brain." Neuron **72**(4): 665-678.

Pruim, R. H., M. Mennes, D. van Rooij, A. Llera, J. K. Buitelaar and C. F. Beckmann (2015). "ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data." NeuroImage **112**: 267-277.

Reich, W., Z. Welner and B. Herjanic (1997). Diagnostic interview for children and adolescents-IV (DICA-IV) Multi-Health Systems, Toronto, Canada.

Revelle, W. (2017). psych: Procedures for psychological, psychometric, and personality research. Evanston, Illinois, Northwestern University.

Roberts, B. A., M. M. Martel and J. T. Nigg (2013). "Are there executive dysfunction subtypes within ADHD?" Journal of Attention Disorders.

Rommelse, N. N. J., J. M. J. van der Meer, C. A. Hartman and J. K. Buitelaar (2016). "Cognitive profiling useful for unraveling cross-disorder mechanisms." Clinical Psychological Science **4**(6): 957-970.

Rosenthal, M., G. L. Wallace, R. Lawson, M. C. Wills, E. Dixon, B. E. Yerys and L. Kenworthy (2013). "Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders." Neuropsychology **27**(1): 13.

Roth, R. M., P. K. Isquith and G. A. Gioia (2005). BRIEF-A: Behavior Rating Inventory of Executive Function-Adult version: Professional manual, Psychological Assessment Resources.

Rubia, K. (2018). "Cognitive neuroscience of Attention Deficit Hyperactivity Disorder (ADHD) and its clinical translation." Frontiers in Human Neuroscience **12**: 100.

Rutter, M., A. Le Couteur, C. Lord and R. Faggioli (2005). ADI-R: Autism diagnostic interview--revised: Manual, OS, Organizzazioni speciali.

Sainani, K. L. (2017). "Reliability Statistics." PM R **9**(6): 622-628.

Satterthwaite, T. D., D. H. Wolf, K. Ruparel, G. Erus, M. A. Elliott, S. B. Eickhoff, E. D. Gennatas, C. Jackson, K. Prabhakaran, A. Smith, H. Hakonarson, R. Verma, C. Davatzikos, R. E. Gur and R. C. Gur (2013). "Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth." NeuroImage **83**: 45-57.

Schimmack, U. (2010). "What multi-method data tell us about construct validity." European Journal of Personality **24**(3): 241-257.

Schmahmann, J. D. (1991). "An emerging concept: the cerebellar contribution to higher function." Archives of Neurology **48**(11): 1178-1187.

Scott, W. A. (1962). "Cognitive complexity and cognitive flexibility." Sociometry: 405-414.

Seeley, W. W., V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss and M. D. Greicius (2007). "Dissociable intrinsic connectivity networks for salience processing and executive control." J Neurosci **27**(9): 2349-2356.

Shanmugan, S., D. H. Wolf, M. E. Calkins, T. M. Moore, K. Ruparel, R. D. Hopson, S. N. Vandekar, D. R. Roalf, M. A. Elliott, C. Jackson, E. D. Gennatas, E. Leibenluft, D. S. Pine, R. T. Shinohara, H. Hakonarson, R. C. Gur, R. E. Gur and T. D. Satterthwaite (2016). "Common and dissociable mechanisms of executive system dysfunction across psychiatric disorders in youth." American Journal of Psychiatry **173**(5): 517-526.

Shrout, P. E. and J. L. Fleiss (1979). "Intraclass correlations: uses in assessing rater reliability." Psychological Bulletin **86**(2): 420.

Signorell, A. (2017). DescTools: Tools for descriptive statistics. R package version 0.99.23.

Sikora, D. M., P. Vora, D. L. Coury and D. Rosenberg (2012). "Attention-deficit/hyperactivity disorder symptoms, adaptive functioning, and quality of life in children with autism spectrum disorder." Pediatrics **130 Suppl 2**: S91-97.

Silver, C. H. (2014). "Sources of data about children's executive functioning: review and commentary." Child Neuropsychol **20**(1): 1-13.

Smith, S. M., K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey and M. W. Woolrich (2011). "Network modelling methods for FMRI." NeuroImage **54**(2): 875-891.

Stevens, M. C., G. D. Pearlson, V. D. Calhoun and K. L. Bessette (2018). "Functional neuroimaging evidence for distinct neurobiological pathways in Attention-Deficit/Hyperactivity Disorder." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **3**(8): 675-685.

Stoodley, C. J. and J. D. Schmahmann (2009). "Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies." NeuroImage **44**(2): 489-501.

Supekar, K. and V. Menon (2012). "Developmental maturation of dynamic causal control signals in higher-order cognition: a neurocognitive network model." PLoS Comput Biol **8**(2): e1002374.

Tamm, L., P. A. Nakonezny and C. W. Hughes (2014). "An open trial of a metacognitive executive function training for young children with ADHD." Journal of Attention Disorders **18**(6): 551-559.

Team, R. C. (2017). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Toplak, M. E., R. F. West and K. E. Stanovich (2013). "Practitioner review: do performance-based measures and ratings of executive function assess the same construct?" Journal of Child Psychology and Psychiatry **54**(2): 131-143.

Uddin, L. Q. (2015). "Salience processing and insular cortical function and dysfunction." Nature Reviews Neuroscience **16**(1): 55-61.

Uddin, L. Q., K. S. Supekar, S. Ryali and V. Menon (2011). "Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development." The Journal of Neuroscience **31**(50): 18578-18589.

van der Meer, J. M., A. M. Oerlemans, D. J. van Steijn, M. G. Lappenschaar, L. M. de Sonneville, J. K. Buitelaar and N. N. Rommelse (2012). "Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample." Journal of the American Academy of Child and Adolescent Psychiatry **51**(11): 1160-1172 e1163.

Van der Werf, Y. D., M. P. Witter, H. B. Uylings and J. Jolles (2000). "Neuropsychology of infarctions in the thalamus: a review." Neuropsychologia **38**(5): 613-627.

van Schouwenburg, M. R., H. E. den Ouden and R. Cools (2010). "The human basal ganglia modulate frontal-posterior connectivity during attention shifting." Journal of Neuroscience **30**(29): 9910-9918.

van Schouwenburg, M. R., A. M. Onnink, N. ter Huurne, C. C. Kan, M. P. Zwiers, M. Hoogman, B. Franke, J. K. Buitelaar and R. Cools (2014). "Cognitive flexibility depends on white matter microstructure of the basal ganglia." Neuropsychologia **53**: 171-177.

Vatansever, D., A. E. Manktelow, B. J. Sahakian, D. K. Menon and E. A. Stamatakis (2016). "Cognitive flexibility: A default network and basal ganglia connectivity perspective." Brain Connect **6**(3): 201-207.

Wager, T. D., J. Jonides and S. Reading (2004). "Neuroimaging studies of shifting attention: a meta-analysis." NeuroImage **22**(4): 1679-1693.

Waterhouse, L. and C. Gillberg (2014). "Why autism must be taken apart." Journal of Autism and Developmental Disorders **44**(7): 1788-1792.

Watson, C. G. (2018). brainGraph: Graph Theory Analysis of Brain MRI Data.

Xia, C. H., Z. Ma, R. Ciric, S. Gu, R. F. Betzel, A. N. Kaczkurkin, M. E. Calkins, P. A. Cook, A. Garcia de la Garza, S. N. Vandekar, Z. Cui, T. M. Moore, D. R. Roalf, K. Ruparel, D. H. Wolf, C. Davatzikos, R. C. Gur, R. E. Gur, R. T. Shinohara, D. S. Bassett and T. D. Satterthwaite (2018). "Linked dimensions of psychopathology and connectivity in functional brain networks." Nat Commun **9**(1): 3003.

Yerys, B. E., G. L. Wallace, J. L. Sokoloff, D. A. Shook, J. D. James and L. Kenworthy (2009). "Attention deficit/hyperactivity disorder symptoms moderate cognition and behavior in children with autism spectrum disorders." Autism Res **2**(6): 322-333.

Yerys, B. E., B. C. Wolff, E. Moody, B. F. Pennington and S. L. Hepburn (2012). "Brief Report: impaired Flexible Item Selection Task (FIST) in school-age children with autism spectrum disorders." Journal of Autism and Developmental Disorders **42**(9): 2013-2020.

Zald, D. H. and B. B. Lahey (2017). "Implications of the hierarchical structure of psychopathology for psychiatric neuroimaging." Biol Psychiatry Cogn Neurosci Neuroimaging **2**(4): 310-317.

Zelazo, P. D. (2006). "The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children." Nature Protocols **1**(1): 297.

Zelazo, P. D., J. E. Anderson, J. Richler, K. Wallner-Allen, J. L. Beaumont and S. Weintraub (2013). "II. NIH toolbox cognition battery (CB): measuring executive function and attention." Monographs of the Society for Research in Child Development **78**(4): 16-33.

Zhong, J., A. Rifkin-Graboi, A. T. Ta, K. L. Yap, K. H. Chuang, M. J. Meaney and A. Qiu (2014). "Functional networks in parallel with cortical development associate with executive functions in children." Cerebral Cortex **24**(7): 1937-1947.

Zuo, X. N., C. Kelly, A. Di Martino, M. Mennes, D. S. Margulies, S. Bangaru, R. Grzadzinski, A. C. Evans, Y. F. Zang, F. X. Castellanos and M. P. Milham (2010). "Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy." Journal of Neuroscience **30**(45): 15034-15043.