INDIVIDUAL DIFFERENCES IN EXECUTIVE (DYS)FUNCTION IN RELATION TO SLEEP DURATION AND PSYCHOPATHOLOGY

By CHELSIE E. BENCA

B.A., University of Michigan, 2009

M.A., University of Colorado Boulder, 2016

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement of the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2017

This thesis entitled:

Individual Differences in Executive (Dys)Function in Relation to Sleep Duration and Psychopathology

Written by Chelsie E. Benca
has been approved for the Department of Psychology

_____
(Naomi Friedman – Committee Chair)

_____
(John Hewitt– Committee Member)

_____
(Marie Banich – Committee Member)

_____
(Matthew Keller – Committee Member)

_____
(Kenneth Wright– Committee Member)

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable standards of scholarly work in the above mentioned discipline

IRB protocol # 11-0614, 14-0421, 17-005

Benca, Chelsie E. (Ph. D., Department of Psychology and Neuroscience)

Individual Differences in Executive (Dys)Function in Relation to Sleep Duration and Psychopathology

Thesis directed by Professor Naomi Friedman

This dissertation presents four studies that examined how executive functions (EFs) relate to problematic behaviors, such as psychopathology and atypical sleep, as well as how EFs are characterized in healthy individuals. The first study examined whether genetic risk for five different forms of psychopathology predicted EFs at the latent variable level in a population sample. The second study characterized the neural activation patterns in health individuals in response to 3 different types of EF tasks. This study (a) assessed the overlapping activation elicited across the different EF task types, and (b) used neural activation to predict high or low EF ability. The third study examined whether individual differences in sleep duration influenced either depression or EFs (a) within a single time of assessment, (b) across time, and (c) how sleep duration, depression, and EFs influence each other in the same model. Study 4 decomposed the relationships of sleep duration with depression and then EF into their genetic and environmental factors in order to better understand the underlying architecture for these relationships.

In the first study, publically available datasets from the Psychiatric Genomics Consortia were used to generate polygenic risk scores for 5 different psychiatric disorders: Autism, Attention Deficit Hyperactivity Disorder (ADHD), Bipolar Disorder, Major Depressive Disorder (MDD), and Schizophrenia. I then used a deeply phenotyped (and genotyped) subset of 354 twins in the Colorado Longitudinal Twin Study (LTS) from the University of Colorado Boulder to test whether or not genetic risk in these individuals predicted EF abilities. I also examined whether the appropriate risk scores were associated with ADHD and MDD symptoms or lifetime diagnoses to the same relative extent as the EF scores. Results indicated polygenic risk for psychopathology did not significantly predict EFs after controlling for multiple testing. Results also suggested that effect sizes for EFs were comparable to those for ADHD and MDD symptoms and lifetime diagnoses.

The second study was a pilot study that included 30 subjects from the Colorado Twin Study from the University of Colorado Boulder at approximately age 28. These subjects were chosen because they were either high or low in Common EF ability as measured in a previous wave of data collection 7 years prior, until we had 15 of each. Each subject completed 3 EF tasks in a functional magnetic resonance imaging scanner. Results indicated that common brain activation in response to these tasks both overlapped with a frontoparietal network typically associated with cognitive tasks, and extended beyond this network. The common areas associated with individual differences in EF ability fell outside of the frontoparietal network.

The third and fourth studies utilized data from the same group of 857 twins from the LTS sample. These studies examined sleep, depression, and EF when available at approximately ages 12, 17, 21, and 23. Study three looked at the phenotypic relationships

between these variables and found linear and nonlinear relationships between sleep duration and EFs and depression across age. When put together in the same model, depression seems to suppress the relationship between EF and sleep duration in adolescence. Study four results showed that sleep duration is moderately heritable, and that the phenotypic relationships between these variables is typically attributable to non-shared environmental influences.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health T32

DA017637, MH063207, MH001865, MH016880, MH079485, DA011015, DA035804,

AG046938, DA035484 & HD010333

CONTENTS

TABLES

Table

FIGURES

Figure

CHAPTER 1

Brief Introduction

Individual differences in Executive Functions (EFs), higher-level cognitive mechanisms that allow individuals to set and reach goals, predict many important life and health outcomes, ranging from academic and occupational performance, to psychopathology and criminality (Miller, Nevado-Montenegro, & Hinshaw, 2012; Morgan & Lilienfield, 2000). In fact, executive dysfunction is so pervasive in many forms of psychopathology, that it has been proposed as a potential endophenotype for a number of disorders, such as major depression and schizophrenia (Hasler, Drevets, Manji, & Charney, 2004; Snitz, MacDonald, & Carter, 2006). This body of research works to expand upon previous literature by further characterizing a deeply phenotyped model of EFs, and how individual differences in this model relate to psychopathology and sleep duration.

One strength in this body of research is the use of a well-established, deeply phenotyped model of EFs. The Unity and Diversity model of EF uses a battery of 9 EF tasks, selected from three different domains (response inhibition, updating working memory, and task-switching), to extract 3 latent factors: Common EF, Updating-specific (UPD), and Shifting-specific (SHI). The Common EF latent factor explains variance in all 9 of the EF tasks and is thought to reflect goal maintenance. The UPD factor explains additional variance in the 3 updating tasks after the common variation is portioned out, and strongly relates to general fluid intelligence (Friedman et al., 2006). The SHI factor explains remaining variance in the 3 shifting tasks and often shows an opposite pattern with psychopathology compared to Updating-specific abilities and Common EF. These

factors allow for a purer measurement of the underlying EF constructs that can then be related to problematic behaviors.

Individual differences in these EFs relate to many types of problematic behaviors such as depression (Friedman et al., in prep), substance use (Gustavson et al., 2017), and behavioral disinhibition (Young et al., 2009). Longitudinally, these EFs relate to sleep problems (Friedman, Corley, Hewitt, & Writght, 2009), self-restraint (Friedman, Miyake, Robinson & Hewitt, 2011), and attention problems (Friedman et al., 2007). Given previous research suggesting EFs in general as endophenotypes for psychoaphtology and the fact that these latent factors have been shown to be highly heritable (Friedman et al., 2008), and stable (Friedman, 2016), make the Unity and Diversity model of EF a good candidate model for EFs as endophenotypes.

First, I used genetic risk for 5 forms of psychopathology to predict individual differences in EFs. All associations were small and did not survive correction for multiple testing. This study highlighted 1) the need for large discovery and testing samples when using polygenic risk scores, even with deeply phenotyped testing samples, and 2) that purported endophenotypes might not produce larger effect sizes than the more distal behavior of interest, at least at the level of cognitive lab-based measures. Rose & Donohue (2013) conclude that neuroimaging studies of cognition produce more robust associations with Schizophrenia than their lab-based counterparts. This finding and the results of Study 1 helped to motivate the next study.

In Study 2, I characterized the neural networks common to EFs, using functional magnetic resonance imaging of response inhibition, updating working memory, and set-shifting tasks. This study highlighted that in general EF tasks activate a fronto-parietal

network (FPN), similar to ones previously identified as being related to multiple cognitive tasks. However, the overlapping regions from the 3 EF tasks did show some differences from this previously identified FPN, suggesting that they cannot be used interchangeably. Importantly, the regions that predicted individual differences in EFs usually fell outside of the overlapping group activation, suggesting that individuals with higher Common EF might activate areas other than those necessary to complete the task in order to boost their performance when the task is difficult. Before neural networks of EF can be tested as an endophenotypes, many more participants need to be collected. Data collection is ongoing.

In the interim, I further characterized the relationship between sleep duration and EFs, as well as sleep duration and depression in both phenotypic and genetically informed models. Adolescence into young adulthood is a critical period for the development and refinement of EFs (Casey, Getz & Galvin, 2008; Diamond, 2002). This time-period is also marked by changes in sleep pressure and circadian rhythms (Barclay & Gregory, 2013). Therefore, I first phenotypically assessed the relationships between sleep duration and EF, sleep duration and depression, as well as how EF, sleep duration, and depression influence each other. Second, I addressed whether the phenotypic relationships were driven by shared genetic or environmental influences. Overall, shorter sleep durations and sleeping more or less than the average person are associated with increased depression, while sleep duration and EFs have a more nuanced relationship. Shared genetic variation contributed to both sleep duration and depression in early adulthood, but overall most associations between sleep duration and depression or EFs

were explained by non-shared environmental influences that make twins more different from each other.

Overall this body of research sets the groundwork for further assessing the Unity and Diversity model of EF as a potential endophenotype for various forms of psychopathology and maladaptive behaviors.

CHAPTER 2

Predicting Cognitive Executive Functioning with Polygenic Risk Scores for Psychiatric Disorders

Executive functions (EFs) — higher-order cognitive processes that regulate thoughts and actions during goal-directed behavior — are implicated in many types of psychopathology. Individuals with attention-deficit/hyperactivity disorder (ADHD), autism (AUT), schizophrenia (SCZ), major depressive disorder (MDD) and bipolar disorder (BP), as well as other psychiatric disorders, show EF deficits, and many of the symptoms for these disorders reflect EF dysfunction (Amann et al., 2012; Rosenthal et al., 2013; Synder, 2013; Snyder, Miyake, & Hankin, 2015). In fact, researchers have hypothesized that EFs are endophenotypes — intermediate phenotypes on the pathway between genes and diagnosis — for these disorders (Nyden, Hagberg, Gousse, & Rastam, 2011; Glahn, Bearden, Niendam, & Escamilla, 2004; Hasler, Drevets, Manji, & Charney, 2004; Snitz, MacDonald, & Carter, 2006; Willcutt et al., 2005). If so, then the genes that influence EFs also influence vulnerability to psychiatric disorders. In this study, we examine the hypothesis that genetic risk for psychiatric disorders predicts individual differences in EFs. We use large, publically available samples for ADHD, AUT, BP, MDD, and SCZ to find genetic risk variants and construct polygenic risk scores (PRSs) for each disorder, then test whether these risk scores predict EFs in an independent population-based sample that is smaller but more deeply phenotyped.

The EF framework we use is the unity/diversity model (Miyake & Friedman, 2012), which was recently discussed by Snyder et al. (2015) as a particularly promising framework for gaining new insights into the relationship between EFs and

psychopathology. This model examines nine tasks tapping three separable but correlated latent variable EFs (response inhibition, updating working memory, and shifting sets). The covariances among the nine tasks are partitioned into three orthogonal factors: Common EF, which explains variance in all nine tasks, including the response inhibition tasks; Updating-Specific, which explains residual covariance among the updating working memory tasks (once the common factor is accounted for); and Shifting-Specific, which similarly explains residual covariance among the tasks designed to examine task shifting ability.

The Common EF latent factor is thought to reflect active goal maintenance and top-down biasing of lower-level cognitive processing (Miyake & Friedman, 2012), which may be particularly important to avoid dominant or automatic responses. In fact, Common EF is isomorphic with response inhibition; in other words after accounting for Common EF, there is no Inhibition-Specific factor. The Shifting-Specific factor is thought to capture individual differences in the speed with which no-longer-relevant goals are cleared from working memory, and the Updating-Specific factor is thought to capture individual differences in gating information into working memory, as well as possibly memory-specific factors like retrieval (Miyake & Friedman, 2012). Our prior work with this model (see Miyake & Friedman, 2012), as well as existing meta-analyses and reviews (e.g., Snyder, 2013; Snyder et al., 2015) suggests that the Common EF factor is the most closely related to multiple forms of psychopathology. There is less work examining specific variances in updating and shifting (i.e., after removing Common EF variance), but some prior research suggests that they show different relationships with psychopathology-relevant behavior (see summary in Herd et al., 2014). Given this body

of research, we use this model as a candidate endophenotype.

A mediational endophenotype, also referred to as an *intermediate* phenotype, is assumed to be closer to the genetic risk factors for the disorder and the behavioral symptoms (Kendler & Neale, 2010). Therefore, relevant genes should be more strongly associated with the endophenotype than the psychiatric disorder itself (Flint & Munafò, 2007; Walters & Owen, 2007). Proposed criteria for endophenotypes include the following: They are associated with the disorder, heritable, and found in unaffected family members at higher rates than in the general population (Gottesman & Gould, 2003). Endophenotypes should also co-segregate in families, and be state-independent, or exist in probands even when they are not currently exhibiting the disorder (Gottesman & Gould, 2003). Thus, one should be able to find an association between genetic risk for psychopathology and purported endophenotypes even in individuals who do not meet criteria for a disorder at the time they are measured on the endophenotypes.

Twin and family studies have shown that most complex psychiatric disorders are heritable (Shih, Belmonte, & Zandi, 2004), with heritability estimates of 76% for ADHD, 85-92% for AUT, 59-87% for BP, 37% for MDD, and 81% for SCZ (Faraone et al., 2005; Miles, 2011; Smoller & Finn, 2003; Sullivan, Neale, & Kendler, 2000; Sullivan, Kendler, & Neale, 2003). As relatively few to no single nucleotide polymorphisms (SNPs) have been identified at a genome-wide significance level for most of these psychiatric disorders (with the exception of SCZ; Ripke et al., 2014), and at best only a handful of SNPs have been identified for constructs related to EFs (Davis et al., 2010; Ibrahim-Verbaas et al., 2016; Plomin et al., 2013; Rietveld et al., 2014), it is difficult to assess whether the same genetic variants that predict EFs also predict these psychiatric

disorders or vice versa. Even in cases for which a relatively large number of genome-wide significant variants have been identified, such as the 128 independent associations with SCZ identified by Ripke and colleagues (2014), the variants collectively explain very little of the phenotypic variance on a liability scale (3.4%), with the individual SNPs explaining much less (by one estimate for genetic studies more generally, each SNP is typically associated with a 1.1 odds ratio; Dick et al., 2015).

One approach to increasing effect sizes is to use PRSs. PRSs aggregate the signals from multiple SNPs related to the disorder of interest, instead of testing the association of variants one by one (Dudbridge, 2013; Morrison et al., 2007). To calculate a PRS, first a GWAS in a discovery sample is used to quantify the relations between all SNPs and a disorder. Then SNPs that meet a certain *p*-value threshold (for example, $p < .0005$) in the discovery sample are binned together. However, it is unclear what significance threshold is optimal, because adding SNPs can increase noise as well as signal; the threshold that results in the optimal signal to noise ratio likely varies depending on phenotype and sample size. Thus, studies commonly look at PRSs for SNPs at different *p*-value bins (e.g., all SNPs with $p < .10, .05, .005$, etc.). Then, for a given bin or collection of apriori chosen SNPs, in an independent testing sample, the PRS is computed as a summed count of whether or not each individual has 0, 1, or 2 copies of the risk variants. If the discovery sample is sufficiently large, then one might expect the estimates of the regression betas for each SNP to be stable and accurate, and each SNP can be weighted by its beta from the discovery sample (Dudbridge, 2013). Finally, the PRSs can be used to predict the disorder or another phenotype in the testing sample.

A benefit of using a PRS is the ability to use a large discovery set for one

phenotype (i.e., a psychiatric disorder) to estimate genetic risk and then test for association in an independent, more deeply phenotyped sample. Both samples do not need to have both phenotypes, and a larger sample size is more important in the discovery sample for determining the risk variants and estimating the SNP effect sizes (Dudbridge, 2013). So, a smaller testing sample, which in this case has been assessed with great rigor, can be used to test the genetic association between the two phenotypes.

Genome-wide association studies (GWAS) for constructs related to EFs (such as intelligence test subscales, matrix reasoning, the Stroop test, the trail-making test, and educational attainment) have had varying success in identifying significant genetic variants (Davis et al., 2010; Ibrahim-Verbaas et al., 2016; Plomin et al., 2013; Rietveld et al., 2014). However, due to sample-size constraints, reverse-phenotyping is frequently employed. For example, the observed heritability of educational attainment is due in part to cognitive ability, but also reflects much more, such as work ethic, motivation, and behavioral problems (Krapohl et al., 2014). A recent GWAS by Rietveld et al. (2014) found that a PRS for educational attainment predicted general cognitive ability better than it did educational attainment in an independent sample. The authors suggested that the higher relation to general cognitive ability than to the originally investigated trait (educational attainment) arose because general cognitive ability is an endophenotype for educational attainment. The authors describe this phenomenon of using risk variants for the disorder of interest to try to predict a purported endophenotype as "reverse endophenotyping." We utilized this approach because we have an extensive EF battery on a relatively small sample that would be inappropriate for risk score discovery. That is, even though it may be more logical to calculate a PRS for the endophenotype and test it

with a psychiatric phenotype, we do the opposite because there are currently larger sample sizes for psychiatric disorders than for these EFs.

One recent study found associations between psychopathology and single cognitive measures (verbal-numerical reasoning, educational attainment, reaction time, and memory) in sample sizes of 36,035 to 112,067 individuals from the UK Biobank (Hagenaars et al., 2016). Associations were examined in 2 ways: genetic correlations from LD score regression, and PRSs. Schizophrenia was the only disorder consistently related to each measure, with genetic correlations ranging from 0.13 to -0.34 and betas from regressions with PRSs ranging from -0.062 to 0.025. This study shows the best-case scenario for effect sizes in large samples with single measures related to cognition. However, this study focused on individual cognitive tests that did not target particular EFs. In the current study, we use a similar approach to examine relations to multiple EFs, measured at the level of latent variables.

Current Study

We used publicly available genome-wide summary data from five case-control samples (AUT, ADHD, BP, MDD, and SCZ) from the Psychiatric Genomics Consortium (PGC) (Sullivan et al., 2010). We calculated PRSs for each disorder at multiple *p*-value bins, and then used them to predict three separable EFs (Common EF, Updating-Specific, and Shifting-Specific latent variables) in an independent sample composed of unrelated individuals drawn from two Colorado twin studies (n = 386 with both genetic and EF data).

For our EF measures, we employed a latent variable model, which has two major advantages over individual tasks. First, because they only reflect variance that correlates

across tasks, latent variables are free from measurement error due to unreliability (Bollen, 1989). Second, particularly for EF constructs, latent variables are more valid measures, because they remove task impurity (Miyake et al., 2000). EFs are higher-level processes that act on lower-level processes; so individual EF tasks typically include a good deal of variance that is not related to the EF of interest (such as verbal or spatial ability). The EF model that we use includes measures that were selected to tap the same EFs but differ in these non-EF requirements so that this non-EF variance would be removed from the latent variables. The result is a purer measure of the EF, but the consequence is that standard errors for estimates of relations with these latent variables may be larger than those for individual tasks to the extent that the latent variable loadings are low (which they typically are for EF models). High reliability and validity is particularly important in evaluating endophenotypes, because poor measurement can outweigh the benefits gained by an endophenotype's more proximal connections between genes and behavior. Prior research with a subset of the data used here demonstrates that these EF latent variables are highly heritable and show high stability across a 6-year time window (Friedman et al., 2016).

While many previous studies of general cognitive ability have larger samples, deep phenotyping by selecting highly heritable EF constructs based on a well-characterized model of EF should increase our ability to detect an association between psychopathology PRSs and EF, particularly in a smaller sample. Prior work with the data from the Colorado Longitudinal Twin Study sample (Friedman et al., 2016) indicates that these EF latent variables have heritabilities at age 17 of 98% for Common EF, 100% for Updating-Specific latent factor, and 76% for Shifting-Specific. In the same sample,

heritability for a general intelligence factor was estimated at 76% (Friedman et al., 2008). These same latent variables are stable from ages 17 to 23 years, with correlations between the two ages of .86, 1.0, and .91 for Common EF, Updating-Specific, and Shifting-Specific abilities, respectively (Friedman et al., 2016). Moreover the Common EF factor is more strongly related than general cognitive ability to behavior that is relevant to psychopathology, such as attention problems and self-restraint (e.g., Friedman et al., 2007; 2011). Thus, Common EF is a strong candidate for examination as an endophenotype for psychopathology.

For comparison purposes, we also included measures that were more similar to the psychiatric disorders on which the risk scores were based. Specifically, we used the ADHD and MDD PRSs to predict attention problems, depression symptoms, ADHD and MDD lifetime diagnosis, and a joint general anxiety disorder (GAD) and/or MDD lifetime diagnosis (given that depression and anxiety have a high genetic correlation; Kendler et al., 1992). These analyses enabled us to examine whether our effect sizes for EFs are larger or smaller than those for phenotypes that more closely match the original psychiatric phenotypes used to generate the PRSs.

Intelligence quotient (IQ), another proposed endophenotype for psychopathology and a construct related to EF (Friedman et al., 2006), has been previously linked to genetic risk for SCZ. As such, we tested whether the PRSs were correlated with IQ in our sample to see if we replicate this association and to better interpret the observed relationships between PRSs and EFs.

While we are interested in the relationship between all three latent factors in the EF model and psychopathology, we hypothesize that the PRSs will be negatively related

to the Common EF factor, based on prior work suggesting that multiple forms of psychopathology are associated with broad EF deficits (e.g., Snyder et al., 2015). In addition, prior phenotypic and genetic models with one of these samples suggest a possible positive relationship between the Shifting-Specific factor and PRSs for psychopathology, reflecting a stability/flexibility tradeoff with the Common EF factor (Miyake & Friedman, 2012); therefore, we hypothesize that PRSs will be positively related to the Shifting-Specific factor.

Method

Participants

Target sample. For genetic analyses participants were 452 individual twins (178 female; mean age at time of EF testing 19.6 [SD = 2.3]), a subset from 2,935 twins recruited from the Colorado Longitudinal Twin Study (LTS) and the Colorado Community Twin Study (CTS) at the University of Colorado (Rhea et al., 2013). For all models we included all individuals who had phenotypic data in order to get a more robust estimation of phenotypic traits (distributions or thresholds); however, only a subset of 452 to 386 individuals who had both genotypic and phenotypic information, depending on the analysis, contributed to the correlation between PRS and phenotype. For example, in the estimation of the EF models, we used all twins who had EF data (n = 1,543) in order to get better, more stable estimates of the latent factor loadings; however, only one twin from a subset of those twin pairs was genotyped, and of those, only Caucasian samples were imputed to the 1000 genomes reference panel. Out of the 452 individuals with imputed genotype data, 386 also had EF data, 387 had IQ data, 452 had diagnostic information for ADHD, MDD, general anxiety disorder and/or major depression, or

depression symptoms from the Center for Epidemiologic Studies-Depression scale (CES-D), and 257 also had Child Behavioral Checklist (CBCL; Achenbach et al., 1991) data (see supplemental Table S1 for *n*s).

Discovery samples. We used publicly available summary statistics from GWAS to obtain the sets of SNPs (and associated beta weights) to be included in the PRSs for each disorder. The discovery data came from the PGC (Sullivan et al., 2010) and included an AUT sample from the Autism Disorder Working Group (March 2015 Release; URLs:PGC) with 10,610 individuals (5,305 ASD cases and 5,305 pseudocontrols), an ADHD sample (Neale et al., 2010) with 9,543 individuals (896 cases, 2,455 controls, 2,064 trios), a BP sample (Sklar et al., 2011) with 16,731 individuals (7,481 cases, 9,250 controls), a MDD sample (Ripke et al., 2013) with 76,237 individuals (16,023 cases, 60,214 controls), and a SCZ sample (Ripke et al., 2014) with 150,064 individuals (36,989 cases, 113,075 controls). For more details on the discovery samples' characteristics, preprocessing procedures, and analysis methods used by the PGC, see the papers associated with each dataset.

Materials

Attention problem symptoms. Attention problems were assessed by the attention problems subscale of the Child Behavior Checklist (CBCL; Achenbach et al., 1991). This subscale had 11 symptoms that could be endorsed as not true (0), somewhat true (1), or very true (2), for a maximum score of 22 points. For the LTS, we used multiple waves of parent (either mother or father) ratings from age 7 until age 16 years. After taking the square root of the raw score to help normalize the distribution, we regressed out age separately within each sex at each time point, then averaged the

standardized residuals across time. We followed the same procedure for the CTS sample, however we only had parent ratings (mother, father, or both) at one time point. Across both the LTS and CTS samples, mothers' ratings were more common than fathers'; only mothers answered approximately 77% of the time, only fathers answered 9% of the time, and both parents answered approximately 13% of the time. When both were available, we averaged the parents' ratings at that time point, and then averaged the combined rating with the other time points. Descriptive statistics for raw scores are provided in Supplemental Table S1.

Depression symptoms. Participants completed the Center for Epidemiologic Studies-Depression scale (CES-D; Radloff, 1977) at three waves: wave 1 (ages 11.33 to 15.99 years), wave 2 (ages 15.75 to 27.45 years), and wave 3 (ages 21.10 to 34.37 years). This 20-question scale assesses how often a person experiences depressive symptoms on a scale of 0 (rarely or none of the time) to 4 (most or all of the time). At each wave, after reverse-scoring appropriate questions, if an individual answered at least 16 questions, we took the mean of those questions and multiplied it by 20 in order to get a sum score.[1] We used a square root transformation to help normalize the distribution and regressed out age, sex, and their interaction, then averaged the standardized residuals across waves to get a single score for each participant.

Lifetime diagnoses. We examined three lifetime diagnoses: ADHD, MDD, and GAD and/or MDD. Adult case-control status was assessed by the DSM-IV diagnostic criteria, or the DSM-IIIR adjusted to be equivalent with the DSM-IV diagnostic criteria if data were collected before 2002. We used the Diagnostic Interview Schedule (DIS;

[1] Across all waves, only 4 scores were not computed because the participant did not answer at least 16 questions.

Robins et al., 2000) for participants 18 or older, and the Diagnostic Interview Schedule for Children (DISC; Shaffer, et al, 2000) for participants younger than 18. We had three waves of data available (see *Depression symptoms* section) and used all of the data to create our measures. Age at time of psychopathology assessment ranged from 12-34 with a mean age of 24.4 (SD = 3.7). When there were multiple assessments, the age in supplemental Table S1 is from the most recent wave of available data. Our final variables were dichotomous variables for each disorder, where if the participant had ever met criteria for diagnosis at any wave, he or she was considered a case. Out of the 452 participants who had genetic data and information on lifetime diagnosis, 43 (9.5%) had a lifetime diagnosis of ADHD, 107 (23.6%) had a lifetime diagnosis of MDD, 45 (10%) had a lifetime diagnosis of GAD, and 120 (26.5%) had a lifetime diagnosis of MDD and/or GAD.

Full-scale intelligence. IQ was measured using the Wechsler Adult Intelligence Scale, third edition (WAIS-III; Wechsler, 1997) in the LTS sample, and the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) in the CTS sample. The WAIS-III was collected at a mean age 16.58 (SD = 0.79), with a mean score of 102.2 (range 70 to 142). The WASI was collected at a mean age of 21.09 (SD = 1.72), with a mean score of 106 (range 53 to 135). Scores were regressed on age, sex, and their interaction within sample, and the standardized residuals were then concatenated.

EF tasks. Nine EF tasks were used to construct EF latent variables. The inhibition tasks (antisaccade, stop-signal, and Stroop) required stopping a prepotent behavioral response (eye movements, categorization, or word reading, respectively). The dependent measures were antisaccade accuracy, estimated stop-signal reaction time in the stop-

signal task, and Stroop response time interference (for incongruent minus asterisks stimuli). The updating working-memory tasks (keep track, letter memory, and spatial 2-back) required monitoring incoming stimuli (words, letters, or spatial locations, respectively) updating working-memory with new relevant information (deleting no longer relevant information) when appropriate. The dependent measures were accuracy. The set-shifting tasks (number–letter, color–shape, and category-switch) required participants to switch between two subtasks (categorizing numbers or letters, colors or shapes, or animacy or size, respectively) on the basis of cues that appeared before each trial. The dependent measures were local switch costs, or the difference in reaction time on switch trials minus repeat trials. Additional information is provided in Table 2.1; see Friedman et al. (2008) for full details. Tasks were administered in the LTS sample at mean age 17.25 years (SD = 0.65) and in the CTS sample at mean age 21.01 years (SD = 1.68). The CTS and LTS samples were combined and then age, sex, and their interaction were regressed out of each EF task score. Standardized factor loadings for the three orthogonal EF latent variables are provided in Table 2.1 for the combined sample. See supplemental Table S2.1 for task descriptive statistics for the sample with genetic data.

Table 2.1 Descriptions and Factor Loadings of the Executive Function Tasks

		Standardized Factor Loading		
Measure	**Description**	**Common EF**	**Updating-Specific**	**Shifting-Specific**
Inhibiting				
Antisaccade	Avoid the prepotent response to saccade to a cue and instead look in	.54	-	-

	the opposite direction to view a briefly displayed target			
Stop-signal	Stop a dominant categorization response on infrequent trials in which an auditory signal sounds	.50	-	-
Stroop	Avoid the prepotent tendency to read a word and instead name the color of the font in which the word is printed	.41	-	-
Updating				
Keep-track	From a series of 15 words, remember the most recently presented exemplar of 2-4 specified categories	.38	.63	-
Letter-memory	During a series of letters, continuously rehearse the last three letters and recall them at the end	.38	.47	-
Spatial 2-back	Respond whether an indicated location is the same as that two trials back	.40	.17	-
Shifting				
Number-letter	Categorize whether the number in a letter-number pair is odd or even, or	.42	-	.45

	whether the letter is a consonant or vowel, depending on the location of stimuli (top or bottom of screen)			
Color-shape	Categorize whether a colored shape is a circle or triangle, or red or green, depending on a cue letter (C or S) appearing above the stimulus	.39	-	.43
Category-switch	Categorize a word as living or nonliving, or small or big, depending on a cue symbol appearing above the word	.45	-	.59

Note. Standardized factor loadings (all $p < .05$) from a model with no genetic risk score or principal components included. Models included the full sample (n = 1,549) although only a subset of 389 individuals contributed to the correlation with the genetic risk scores. The model showed an acceptable fit, $\chi^2(21) = 97.22$, $p < .001$; CFI = .959; RMSEA = .048. EF = executive function.

Procedures

Genotyping: discovery sample. The AUT2 and SCZ2 sample were part of a second phase and were imputed to the 1000 Genome reference panel (The 1000 Genomes Project Consortium, 2010). The BP sample was imputed to HapMap phase 2; the ADHD and MDD samples were imputed to HapMap phase 3 (Thorisson, Smith, Krishnan, & Stein, 2005; The International HapMap Project, 2003). After quality control through PGC (see individual references for more information) all results files were downloaded to our servers. All discovery samples went through a clumping procedure in PLINK (Purcell, et. al., 2007) to account for linkage disequilibrium (LD). Clumping accounts for LD by taking the most significant SNPs in a GWAS, then grouping SNPs that meet an LD

threshold with this most significant index SNP, resulting in only one signal per LD block. We used an LD threshold of $R^2 < 0.2$, with no SNPs excluded based on *p*-values for association with the disorder. The resulting SNPs were then put into R (R Core Team, 2003) and the list of SNP names were matched to the imputed SNPs in the testing sample for PRS generation in the testing sample.

Genotyping: testing sample. Individuals were genotyped on the Affymetrix 6.0 platform (Affymetrix, Inc., Santa Clara CA) and called by BEAGLECALL 1.0.1 (Browning & Yu, 2009). See the description for the "Center on Antisocial Drug Dependence (CADD)" sample in Derringer et al. (2015) for full details of the cleaning and quality control procedures before imputation.

Caucasians were identified by visual inspection of the first 10 components from a principal components analysis calculated in PLINK using the full, unrelated CADD sample (described in Derringer et al., 2015). Cut-offs for the first 3 PCs were applied, and then the remaining subjects were imputed to the 1000 Genome reference panel using IMPUTE2 (Howie, Donnelly, & Marchini, 2009)[2]. The 10 ancestry components were also used as covariates in the analyses.

SHAPEIT was used for the prephasing process (Delaneau, Marchini, & Zagury, 2012). A cut-off info score of >= .4 was used to ensure good quality imputed SNPs, resulting in approximately 14.9 million SNPs. After restricting imputed SNPs to those also identified in the discovery sample (see Table S3 in supplemental materials for number of SNPs in each PRS), the beta weights for those SNPs were used to calculate

[2] Visual inspection involved comparing the self-reported ancestry to the places in the distribution that showed breakpoints (or drop-offs) between the sample's ancestry groups. This resulted in identification of European ancestry participants by component 1>0.014, 0< component 2 <0.013, and component 3>-0.006.

weighted risk scores in the testing sample by multiplying 0, 1, or 2 (for copies of the risk allele), or dosages for imputed SNPs, by the beta weight for those SNPs, and summing across SNPs in each *p*-value bin.

Analyses

Analyses were run in Mplus 7.3 (Muthen & Muthen, 2012) to allow for estimation of the EF latent variables. Models used all available phenotypic data when possible; however, only individuals who also had genetic information contributed to the correlation between PRSs and phenotype (i.e., individuals with EF data, but without genetic data were included in the models to obtain the best estimates of the factor loadings, but the covariance with the PRS was only based on the subset with both genetic and phenotypic data). For models with categorical diagnoses, mean and variance adjusted weighted least squares (WLSMV) estimation (delta parameterization) was used, which models the underlying liability as a normal distribution using a probit model; for models with only continuous data, robust maximum likelihood (MLR) was used. Non-independence (due to including both twins) was corrected for with the type= COMPLEX option, which clusters by family. In all analyses, all individual indicators (e.g., all nine EF tasks) as well as the PRS were regressed on 10 ethnicity PCs.[3]

As described earlier, all continuous phenotypic variables were age, sex, and age by sex[4] regressed before analysis. Age (of last diagnostic assessment) and sex were included as covariates for models including diagnoses. PRSs were not regressed on age

[3] To include individuals without genetic data in the estimation of the EF latent variables (Mplus will exclude individuals missing on covariates) and other phenotypic measures, we imputed missing PCs as the average for that self-identified ethnicity in our genetic sample. The number of individuals who contributed to each phenotype was as follows: EFs=1543; CES-D=2875; CBCL=1684; IQ=1571; DIS diagnoses=2875.

[4] This interaction term was included even though it was not significant in any models.

and sex.

PRSs and EFs. We used structural equation modeling to estimate the three EF latent variables: a Common EF latent variable, representing what is shared between all of the tasks (with loadings from all nine tasks), an Updating-Specific latent variable capturing additional variance specific to Updating tasks (with loadings from three updating tasks), and a Shifting-Specific latent variable capturing additional variance unique to the shifting tasks (with loadings from the three shifting tasks). The latent factors in the EF model are orthogonal, where Common EF explains covariance across all nine tasks, and the Updating- and Shifting-Specific factors explain additional covariance among the updating and shifting tasks, respectively, that is not explained by the Common EF factor.

To examine the relations of these EF latent variables to each PRS, we correlated them with the residual of the PRS (after removing the PCs from the PRS). Thus, the correlations we present are actually partial correlations controlling for ethnicity, because the 10 PCs were regressed out of both the PRS and the individual EF tasks (and the EF tasks were also residualized on age and sex).

PRSs and psychopathological symptoms, diagnoses, and IQ. We used the five PRSs to predict IQ, ADHD symptom scores, ADHD lifetime diagnosis, depression symptom scores, depression lifetime diagnosis, and MDD and GAD lifetime diagnoses. IQ was correlated with all five PRSs, however ADHD symptom scores and diagnosis, and depression symptom scores and diagnosis, were correlated only with the ADHD PRSs and the MDD PRSs respectively. As with the EF model, the correlations we present are actually partial correlations controlling for ethnicity, because the 10 PCs were

regressed out of both the PRS and the phenotype (and the phenotype was also regressed on age and sex).

Permutation. PRSs for higher *p*-value bins include the same SNPs as lower threshold bins for PRSs based on the same disorder. Due to high correlations between *p*-value bins within each risk score (see supplemental Table S2), correlations with the same phenotype across bins of the same risk score are not independent. Therefore, we used permutation to correct for multiple testing. For each permutation, we retained the relatedness of the *p*-value bins within PRSs for each disorder. The association between the independent and dependent variables was broken by randomly shuffling scores for the dependent variables 1000 times and constructing a distribution of statistical coefficients under this null. For example, for the EF model, we shuffled the rows of the nine EF task scores (residualized on age and sex), so that the correlations among the nine EF tasks were retained, but their associations with the PRSs were broken. Because the PRSs were not shuffled, the associations among *p*-value bins remained intact. For each shuffle, we then ran the same model (including ethnicity PCs, which were not shuffled), and obtained the newly estimated correlations between PRSs and EFs. We constructed the empirical distribution of correlation coefficients for each disorder in this way, and used it to calculate empirical *p*-values for the correlations we obtained in our unpermuted models (i.e., a correlation would be significant if it was more extreme than 95% of the empirical correlation values in the distribution of permuted correlations). This is ultimately less stringent than a Bonferroni correction (Camargo et al., 2008) for multiple testing, but does not correct for the multiple testing due to examining multiple phenotypes, for which we divided our alpha of .05 by the number of phenotypes tests (nine) examined, for a new

alpha of .006.

Results

PRSs With Cognitive Measures

EFs. To examine the relationship between EFs and genetic risk for psychopathology, we correlated the PRSs (residualized on PCs) with the EF latent variables (individual tasks regressed on PCs). Correlations are shown in Figure 2.1. Common EF was positively correlated with the MDD $p < .05$ bin PRS, but did not significantly relate to the other PRSs at any *p*-value bin, and this correlation did not survive multiple testing correction.

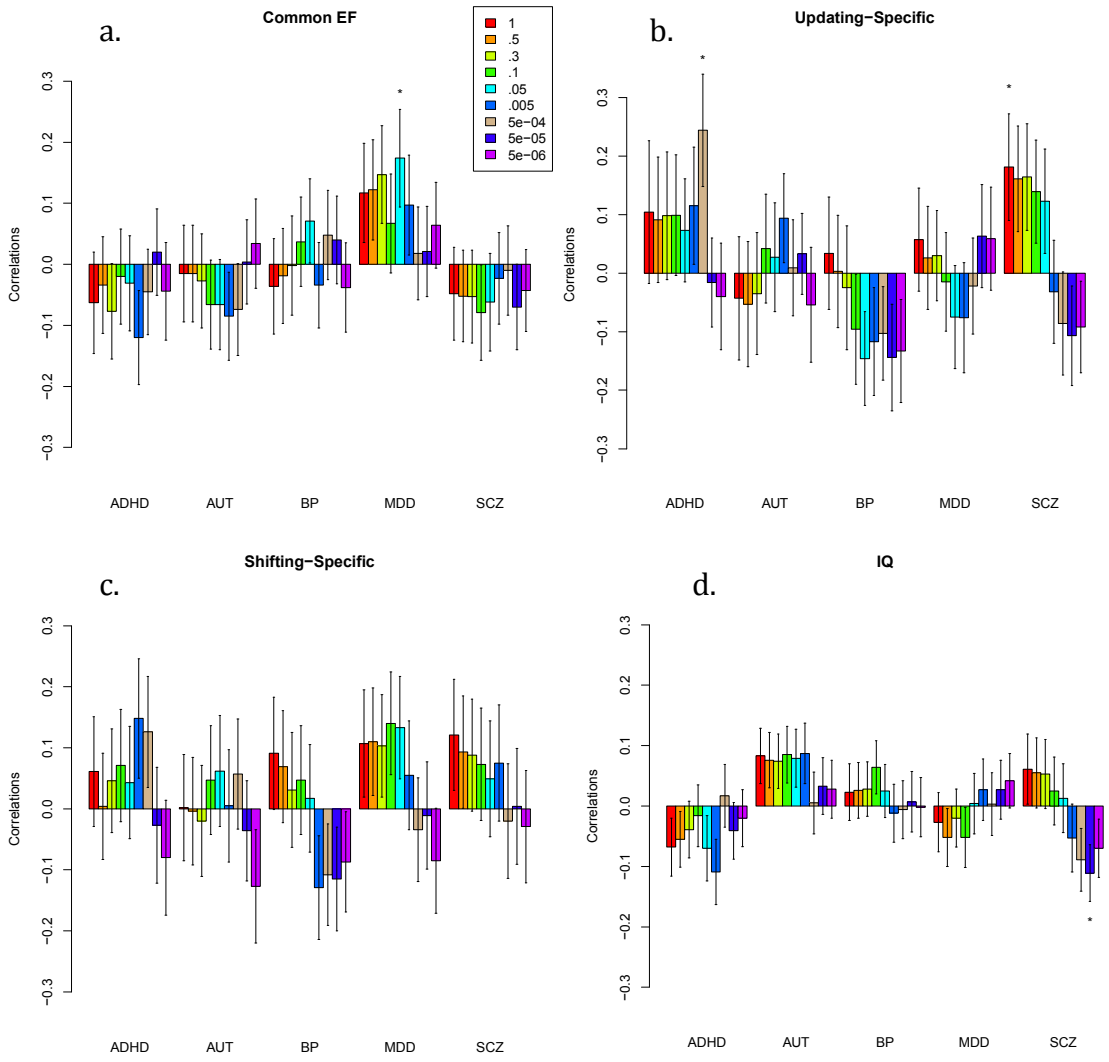

Figure 2.1. Correlations between cognitive measures and psychopathological polygenic risk scores (PRSs). Bars represent standard errors. Legend shows colors corresponding to *p*-value threshold bins for each disorder. (a) Correlations between PRSs and the Common EF latent factor. (b) Correlations between PRSs and the Updating-Specific latent factor. (c) Correlations between PRSs and the Shifting-Specific latent factor. (d) Correlations between PRSs and IQ. EF = executive function; IQ = intelligence quotient; ADHD = Attention Deficit Hyperactive Disorder; AUT =Autism; BP = Bipolar Disorder; MDD = Major Depressive Disorder; SCZ = Schizophrenia. **p*< .05 uncorrected.

The Updating-Specific latent variable significantly positively correlated with the ADHD $p < .0005$ and SCZ $p < 1$ bins. While the former survived permutation, neither of these results were significant after correcting for multiple-testing of the nine phenotypes.

Updating-Specific abilities did not appear to be related to any of the other three disorders. Likewise, Shifting-Specific abilities were not related to genetic risk for any of the five disorders.

IQ. We also examined the relationship between IQ and the PRSs, because IQ is phenotypically associated with EFs (Friedman et al., 2006; 2008) and has been related to PRSs for SCZ (Lencz et al., 2014; McIntosh et al., 2013). As shown in Figure 2.1D, IQ was negatively correlated with the SCZ $p < 5x10^{-5}$ bin, but this result did not survive correction for multiple testing.

PRSs With Measures of Psychopathology

Given the relatively small effects we observed with the proposed endophenotypes (EFs and IQ), we wondered if we would get similarly small effects with phenotypes that were arguably more closely related to the phenotypes used to construct the PRSs. So, we examined how the ADHD and MDD PRSs related to attention and depression symptoms and lifetime diagnoses. The magnitude of effects found for relevant phenotypes within our sample allows for a better understanding of the magnitude of relationship observed with EF and IQ.

The relationships between ADHD and MDD symptom scores and their respective PRSs were assessed with correlational analyses of the residuals of PRS and phenotype after each was regressed on the PCs for ethnicity. As shown in Figure 2.2, PRSs for MDD were not significantly related to any psychopathological phenotypes in our sample. While genetic risk for ADHD was not related to ADHD symptom scores, it was correlated with lifetime diagnosis for ADHD at one bin ($p < .05$), but this result did not survive multiple testing correction. Because we did not find significant results with either

EFs or psychopathology measures after correcting for multiple testing, we did not test whether the magnitudes of effects were significantly larger for EFs.

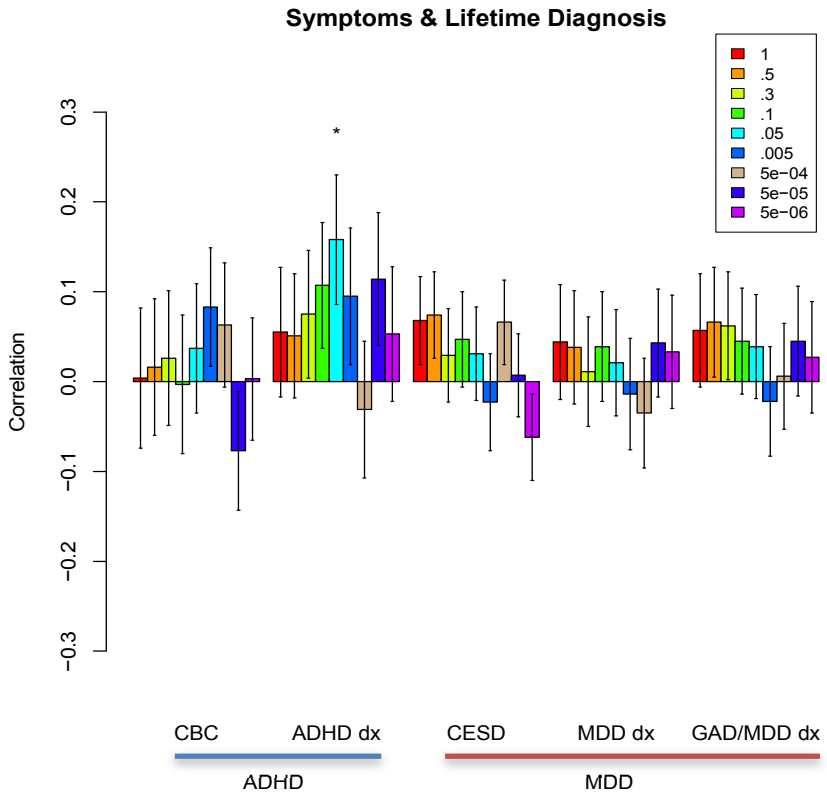

Figure 2.2. Correlations between ADHD, MDD, and GAD/MDD symptoms and lifetime diagnosis, and ADHD and MDD polygenic risk scores (PRSs). Correlations are partial correlations after the 10 principal components for ethnicity have been regressed out of the PRSs and the phenotypic measures and age and sex have been regressed out of the phenotypic measures. Bars represent standard errors. Legend shows colors corresponding to *p*-value threshold bins for each disorder. ADHD = Attention Deficit Hyperactive Disorder; MDD = Major Depressive Disorder; CBC = Child Behavioral Checklist, ADHD dx = lifetime diagnosis of ADHD; CESD = Center for Epidemiologic Studies-Depression Scale; MDD dx = MDD lifetime diagnosis; GAD/MDD dx = General Anxiety Disorder or MDD lifetime diagnosis. *$p < .05$ uncorrected.

Power

To better interpret our results, we conducted power analyses for the EF measures in our sample. As can be seen in Table 2.2, we would have enough power with our sample size used in this study ($N = 386$) if there were a moderate to large correlation ($r =$ .20 to .50) between Common EF and a PRS. However, the observed effect sizes were

smaller than this, and therefore we were underpowered with our sample size. We examined what sample size would be necessary for a power of .80 with a smaller correlation estimate (r = .10). Alpha levels were varied because we examined nine different phenotypes that are not fully independent of one another; for example, ADHD symptoms are correlated with ADHD lifetime diagnosis (r = .25) and MDD symptoms at a lower level (r = .15), so our adjusted alpha should be somewhere between .05 and .006. As shown in Table 2.2, for 80% power to detect an effect with a correlation of .10 or smaller, larger sample sizes, on the order of 1,510 to 2,500 or more, are necessary. In summary, if latent EFs were strong endophenotypes for psychopathology and we observed stronger relations between EFs and these PRSs than previously seen with other cognitive measures, we would have been adequately powered. However with a correlation of .10 or smaller, we would need many more subjects to have adequate power.

Table 2.2 Power for Executive Function Analyses

Simulated Correlations	**1-β; α = .05 and *N*=386**	**Required *N*; α = .05 and 1-β = .8**	**Required *N*; α = .01 and 1-β = .8**	**Required *N*; α = .006 and 1-β = .8**
.50	1	57	84	94
.40	.999	91	135	151
.30	.990	165	245	274
.20	.811	375	558	625
.10	.294	1510	2247	2519

Note. Power analysis for the executive functions (EFs) latent-variable model where simulated correlations represent a theoretical correlation between common EF and the polygenic risk score. 1-β = power; α = alpha; *N* = number of participants included in parameter estimates.

Discussion

To understand the potential of EFs as endophenotypes for psychiatric disorders, we used large discovery datasets to generate PRSs for five disorders (AUT, ADHD, MDD, BP, and SCZ) and related those PRSs to EF latent variables in an independent dataset. We found little evidence for stronger effect sizes for the EFs than measures more similar to these psychopathologies. The general pattern of results indicated that EFs might be related to psychopathology, but they may not lead us to find more genetic variants than symptom or diagnosis measures unless we have significantly larger sample sizes.

At a nominally significant level, a Common EF latent variable was positively related to genetic risk for depression; however, this effect was in the opposite direction than expected, with higher genetic risk for depression indicating better Common EF in a general population sample. Higher genetic risk for ADHD was nominally related to better Updating-Specific abilities; this association was also not in the expected direction. The amount of variance explained by the PRS for each latent factor was $R^2 = 0.03$ for Common EF and $R^2 = 0.06$ for Updating-Specific. However, these results did not survive correction for multiple testing, so they would need to be replicated to determine if they are real effects that are simply underpowered.

Likewise, we examined a measure of general cognitive ability, IQ, which has also been proposed as an endophenotype for psychopathology (Burdick et al., 2009). Although our results did not survive correction for multiple testing, the directionality and variance explained was comparable to what has been observed in previous studies. Lencz et al. (2014) linked a PRS for general cognitive ability to case-control status of SCZ, and

McIntosh et al. (2013) linked a PRS for SCZ to increased cognitive decline between the ages of 11 and 70. We also found that increased genetic risk for SCZ predicted lower IQ, with the amount of variance explained ($R^2 = 0.01$) comparable to that found by Lencz et al. (2014; $R^2 = 0.000$ to 0.019) and McIntosh et al. (2013; $R^2 = 0.006$ to 0.009). The replication of this association between increased genetic risk for SCZ and cognitive ability suggests that we may be seeing real, but underpowered, effects.

A mediation model of an endophenotype (Kendler & Neale, 2010) assumes that the endophenotype is more proximal to genes that influence the psychiatric disorder. If the mediation assumption is incorrect and phenotypes related to disorders of interest, such as depression symptoms, are equally or more strongly related to the PRSs, then EFs as endophenotypes might not be as useful for PRS research. To address this assumption, we also used the PRSs to predict relevant phenotypes more similar to the psychopathologies used to generate the PRSs. Again, we found few associations. A relationship between increased risk for ADHD and lifetime diagnosis for ADHD emerged, where greater genetic risk was related to higher rates of lifetime diagnosis, but it did not survive correction for multiple testing. However, the amount of variance explained ($R^2 = 0.02$) is similar to what we observed for EF and IQ, suggesting that in a small testing sample, psychopathology phenotypes do not have a weaker relationship with PRSs than candidate endophenotypes.

Recently, a few studies have also addressed the assumption that endophenotypes will elicit larger effect sizes with respect to genetic variants. A meta-analysis by Flint and Munafò (2007) concluded that endophenotypes were not necessarily showing larger effect sizes than the disorders of interest. These results could have occurred because the

studies were not using appropriate endophenotypes or because the assumption that endophenotypes have larger effect sizes is incorrect. In particular, if the endophenotypes were not mediators between the genes and phenotypes as often assumed, but instead indices of liability, where the same genes influence both the endophenotypes and the phenotypes of interest (Kendler & Neale, 2010), then one might not expect larger effect sizes for the endophenotypes.

The largest GWAS study to date found no significant hits for EF tasks (Stroop, trail-making, and fluency tests; Ibrahim-Verbaas et al., 2016), despite discovery sample sizes ranging from 5,429 to 32,070. Thus, EF tasks, like other measures, seem to have relatively small effect sizes for individual variants. However, another meta-analysis by Rose and Donohoe (2013) found different effect sizes for two different classes of endophenotypes for SCZ, with larger effect sizes for cognitive neuroimaging endophenotypes than lab-based cognitive measures. More research is needed to establish good estimates of expected effect sizes for different types of endophenotypes.

Another emerging debate focuses on issues of sample size and phenotype specificity when testing for genotype-phenotype associations. Many studies have shown that with the small effect sizes for individual SNPs, large samples will be necessary to detect significant associations with the phenotypes of interest. However, when combining data sets or using large publicly available datasets, often only rudimentary phenotypic assessment is available (e.g., case-control status, without information on which symptoms were endorsed or degree of severity of illness). This thin phenotyping allows for the inclusion of more subjects, but potentially dilutes statistical power and the strength of association (Tracy, 2008). While this trade-off holds in this study with regard to PRS

generation, the deep phenotyping of a candidate endophenotype could possibly help in the testing sample. We had hoped that our deep phenotyping with the latent variable model of EF, which reduces measurement error and extracts highly heritable latent factors that are more stable across time than single measures (Friedman et al., 2016), would enable us to detect a larger effect. We were well powered to detect effects that explained 4% or more of the variance, but the effects we obtained were smaller than that.

Despite being underpowered, there is still useful information to gain pertaining to the effect sizes we can reasonably expect from endophenotypes compared to more direct measures of psychopathology with a small testing sample size. Lab-based measures of EF, even at a highly heritable latent variable level, do not seem to generate substantially larger effect sizes for genes related to risk for psychopathology than measures of symptoms, at least in a population-based sample.

Limitations

In addition to the previously discussed power issues, a limitation of this study is that our sample was population-based with low levels of psychopathology; hence, the genetic variance related to psychopathology was likely restricted compared to a clinical sample. Although endophenotypes are present in individuals without the disorder of interest, particularly in family members of a proband, the use of a population sample might have limited the variance in the endophenotype as well. Thus, a stronger effect would perhaps be seen in a clinical sample.

Although we chose to calculate PRSs from psychiatric disorders and test them with EFs because larger sample sizes are available for the former than the latter, and because its utility has been previously demonstrated in other studies (Lencz et al., 2014;

Rietveld et al., 2014), the reverse endophenotype approach could also be considered a limitation. The relationship between purported endophenotypes and genetic risk for psychopathology is likely a complicated matter (Cannon & Keller, 2006). If an endophenotype is only related to a portion of the genes influencing a given disorder, the strength of the relationship between all genes that affect the disorder and the endophenotype is unclear. Conversely, if the endophenotype is a complex trait itself, such as EF, there are likely unique genetic contributions to EF that do not overlap with the more distal phenotype of interest, such as psychopathology. Due to the unclear genetic relationship between endophenotypes and the more distal phenotype, it is difficult to estimate an expected effect size. However, the genetic architecture of both psychopathology and EF are important for the interpretation of our results.

Multiple testing could also be considered a complication of this study. Associations between five disorders and four phenotypes (Common EF, Updating-Specific, Shifting-Specific and IQ) were tested, as well as one disorder (ADHD) with two phenotypes (ADHD diagnosis and ADHD symptoms), and one disorder (MDD) with three phenotypes (MDD diagnosis, combined GAD / MDD diagnosis, and MDD symptoms), all of these at nine bins. In total, we conducted 225 tests. While this number is not remarkable for those working with GWAS data, it is greater than is typically done in PRS studies. How to adequately correct for multiple testing is complicated by the fact that the nine bins are not independent from each other, the phenotypes are not independent (e.g., ADHD symptom count is correlated with ADHD diagnosis), and the different disorders are also not independent of one another due to comorbidity. We chose to use permutation testing and then use a Bonferonni correction for the number of bins;

however, there is no clear best way to correct for multiple testing in this scenario. One suggestion for future studies would be to reduce the number of bins tested, particularly if testing several phenotypes. However, our initial thorough approach in the exploratory analyses presented here will guide future investigations of relationships between common and specific EFs and a range of psychopathology outcomes.

Conclusion

In this study, we examined the relationship between PRSs for psychopathology and EFs with highly heritable EF latent variables. Despite large sample sizes for deriving PRSs for psychopathology and deeply phenotyped candidate endophenotypes, we did not see substantial effects. The highest observed relations between PRSs for psychopathology and EFs ranged from an R^2 of .03 to .06, which are smaller than we needed for adequate power with our sample size. The highest R^2 for non-EF phenotypes with PRSs was .03, in a similar range as our EF measures. Overall, our results are similar to what was found by Flint and Munafò (2007) and provide little evidence for EFs as endophenotypes that will give significantly larger estimates than psychiatric phenotypes such as lifetime diagnosis. However, even if EFs do not necessarily show larger genetic effect sizes than psychiatric measures, their transdiagnostic associations with psychopathology (Snyder et al., 2015) suggests that increasing understanding of their genetic influences can provide a window into disease mechanisms and pathways.

CHAPTER 3

Is Common Executive Function Related to a Shared Frontopareital Network?

Executive functions (EFs) are higher order cognitive processes that regulate lower-level processes in the service of goal-directed behavior. Studies of normal individual differences indicate that different EFs—such as response inhibition, working memory updating, and set shifting—are correlated but separable, suggesting that they utilize both a common subset of cognitive processes and unique processes specific to each type of EF (Miyake & Friedman, 2012). A large body of neuroimaging research is dedicated to understanding which brain regions are important for particular types of EF at the group-level, usually within the context of individual tasks. When multiple tasks have been examined, the focus has been on which areas show overlapping activation in key contrasts across tasks, rather than which areas predict individual differences in performance. However, it is not necessarily the case that areas important for individual differences are significantly active at the group-level (Yarkoni & Braver, 2010). This study aims to examine whether regions important for EF across a variety of complex cognitive tasks at the group-level, the multiple demand network (Fedorenko, Duncan, & Kanwisher, 2013), predict individual differences in an underlying Common EF latent factor. Specifically, we examine whether individuals who differ in a Common EF factor also differ in their recruitment of a frontoparietal network during three diverse EF tasks loading on this factor. We also examine if individual differences in activation of regions within this frontoparietal network correlate across these three EF tasks.

While there are many subtypes of EFs (Banich, 2009; Diamond, 2013; Jurado & Rosselli, 2007), here we focus on the processes that are common across EFs. Prior work

(Miyake & Friedman, 2012; Friedman & Miyake, in press) used a latent variable model framework and tasks tapping response inhibition, working memory updating, and set shifting to examine the unity and diversity of EFs. They found evidence for a Common EF factor that explained variance in all task types (the unity aspect). Updating-Specific and Shifting-Specific factors explained additional variance, which was specific to the updating and shifting tasks, respectively. The Common EF factor was isomorphic with response inhibition, so there was no inhibiting-specific factor. The Common EF factor is thought to reflect active goal maintenance and top-down bias (Herd et al., 2014; Miyake & Friedman, 2012; Friedman & Miyake, in press).

The neural substrates of specific EFs have been well documented; however regions important for executive functions in general are less well documented. One notable exception is work by Duncan and colleagues, which describes a frontoparietal Multiple Demand (MD) network (Duncan & Owen, 2000; Duncan, 2010; Crittenden & Duncan, 2014) that is recruited in a diverse array of cognitive tasks. The MD network includes many regions previously associated with EFs in other studies, and combinations of these regions have also been called a Frontoparietal (FP) network (Zanto, 2013), Cognitive Control network (Cole & Schneider, 2007), frontoparietal control network (Spreng et al., 2013) as well as other variations. This particular network includes the inferior frontal sulcus (IFS), anterior insula, frontal operculum (AI/FO), pre-supplementary motor area, dorsal anterior cingulate (SMA/ACC), intraparietal sulcus, and sometimes the rostrolateral prefrontal cortex (rlPFC). This MD network has been proposed as related to "goal neglect" and has been linked to general cognitive processes including fluid intelligence(g_f; Bishop et al., 2008; Duncan, 2010; Woolgar et al., 2010).

At the level of behavior, intelligence (both fluid and crystalized) has been linked to a Common EF factor and an Updating Specific factor (Friedman et al., 2006), though the Common EF factor is separable from general intelligence (Friedman et al., 2008). The link to goal maintenance and fluid intelligence makes the MD network a good candidate for areas related to a Common EF factor.

Research by Collette and colleagues (2008) supports the idea that the MD network is an appropriate starting point to look for general group activation related to EF tasks and a Common EF factor. They conducted a positron emission tomography (PET) study looking at the overlap between eight EF tasks tapping inhibition, updating, or set-shifting. A conjunction map of the average group activation for each task resulted in two common areas across tasks where activation was higher in the difficult condition relative to the control condition: the left superior parietal cortex (Brodmann's Area 7) and the right intraparietal sulcus. Other regions were identified at lower significance thresholds across all three tasks, and still other regions were identified across task subtypes.

While areas common across EF tasks are indisputably important for successful performance of these tasks, they might not be the regions important for normal individual differences in performance of the task. One way to better understand the role a region plays in a cognitive process is to look at activation patterns across different tasks or contrasts to see if activity is correlated. For example, if a region shows correlated activation across different attention-shifting tasks, then it is plausible that a common shifting-related cognitive construct underlies activation in response to shifting tasks in that region (Purkayastha, Wager, & Nichols, 2008).

A study by Wager and colleagues (2005) identified a number of regions that were commonly activated across different response inhibition tasks. They then examined correlations of activity patterns across tasks and with behavioral performance to help them understand if these common regions reflected a unified component important for individual differences. A subset of these regions, most consistently the insula, was correlated with poorer performance on these tasks. However, the activation patterns in these regions did not correlate with each other across tasks, and neither did the performance across tasks. This indicates that regions commonly activated during response inhibition tasks might not be the crucial regions for predicting individual differences.

In fact, relatively few individual difference studies find significant effects in areas consistently activated across subjects in within-subject designs (Yarkoni & Braver, 2010). The authors partially attributed this phenomenon to the fact that within-subject variance is considered error in between-subject analyses, and vice versa. In fact, it is not necessary that regions important for individual differences are significantly active at a group-level. An alternative explanation is that on average, people activate areas such as the MD network in order to perform EF tasks, but those who perform better or worse on these tasks recruit regions outside of this frontoparietal network to provide additional support.

Indeed, a recent resting state functional connectivity study suggested that healthy young adults high in Common EF ability might have an expanded frontoparietal network (Reineberg et al., 2015). They found that better Common EF is associated with the connectivity of the frontal pole to an attentional resting state network, as well as

increased connection of Crus I and II of the cerebellum to a frontoparietal resting state network. This study indicates that perhaps those with better EF recruit additional regions outside of the functional areas associated with EF at a group-level.

The Current Study

Based on the unity/diversity EF framework (Miyake & Friedman, 2012), participants completed three functional magnetic resonance imaging (fMRI) tasks: Antisaccade (inhibition), Keep track (working-memory), and Number-letter (set-shifting). We selected 30 subjects, approximately age 28, based on their prior performance on a battery of EF tasks 7 years beforehand: 15 participants high (> 1 SD) and 15 low (< 1 SD) on a Common EF factor score. All participants performed all three tasks in the scanner.

First we assessed whether common areas activated across our three EF tasks, at a within-subject level, were consistent with the previously identified MD network. We broke the MD network into regions of interest (ROIs) to examine activation in our three EF tasks (Federanko, Duncan & Kanwisher, 2013). Particularly we expect that any regions that overlap across our three tasks would fall within this network's ROIs. To examine this we tested whether or not the MD ROIs were significantly active across all three tasks. To accomplish this, we extracted the beta values each ROI for each task, tested if it meaningfully differed from zero, and then looked for a pattern of significance across the 3 tasks for a given ROI.

Second, we examined the relationship between the MD ROIs and Common EF with an individual differences approach, using two different methods. If an underlying Common EF factor predicts performance on three EF tasks, it implies correlations

between the tasks, and to the extent that performance is correlated, neural regions should also be correlated. If MD ROIs are important for individual differences, a person who strongly activates a particular ROI in the inhibition task should also strongly activate that ROI during updating and task switching. Therefore, we examined the pattern of correlations in all pairs of EF tasks across MD ROIs to look for evidence of an underlying Common EF factor.

Alternatively, if MD ROIs relate to individual differences in Common EF, then the MD ROIs should relate to a Common EF group difference score, supporting the idea of a unified underlying construct. This was done two ways. First, we conducted a series of t-tests using a dichotomous high / low Common EF variable, derived from the full EF latent variable model at a previous wave. Second,

Finally, we also addressed these questions using a more exploratory whole-brain approach to see if regions outside of the MD network are common to EFs or related to individual differences. We looked at whole-brain activation at the general group-level, as well comparing high EF to low EF, both at the task level and then across tasks, to see if regions outside of the MD network that are common to EFs or related to individual differences.

Method

Participants

Thirty young, healthy adults (mean age 28.1; 24-32 years; 14 males) participated in the study in exchange for monetary compensation. Participants were recruited from the University of Colorado Boulder Community Twin Sample (CTS). All participants were individual twins who had participated in a prior study on executive functions

approximately 7 years earlier. Participants were selected to have a Common EF score at least 1 standard deviation higher or lower than the mean in the prior wave of data collection.

Materials, Design, and Procedure

Three tasks were selected to tap the three EFs examined in the unity/diversity framework (Miyake & Friedman, 2012; Friedman & Miyake, in press): antisaccade to tap inhibition, keep track to tap updating, and number–letter to tap shifting. These tasks were selected because of their high loadings on their respective factors and their ease of modification for the scanner environment.

Task Design

Antisaccade. The antisaccade task was modified for use in the scanner from a speeded version we have used in prior studies (Friedman, et al., 2016) to maximize individual differences. All of the measurements of stimuli presented are for the projected screen. Participants viewed instructions and practiced the task outside the scanner prior to performing the full task in the scanner. The in-scanner task consisted of 2 runs, with 12 blocks of prosaccade or antisaccade trials, mixed with 6 fixation blocks in each run. Each block was preceded by a 500 ms blank, then instructions indicating block type ("TOWARD", "AWAY", or "FIXATION"), which remained on the screen for 2, 4, or 6 s. The fixation blocks consisted of a ¼" fixation cross in the center of the screen for 20 s. Each antisaccade and prosaccade block consisted of 5 trials; each trial consisted of a fixation cross (variable duration between 1-3 s), followed by a peripheral visual cue (a ¼" square) that would appear either to the left or right of the fixation for 232 ms. Immediately after the cue, the target stimulus (a 9/32" digit from 0 to 9 in a 7/16" squre)

appeared either on the same side as the cue for prosaccade trials, or on the opposite side for antisaccade trials, and remained on the screen for 150 ms before being masked by grey cross hatching. The mask remained on the screen for 1,650 ms, during which time the participant verbally reported the target. Antisaccade and Prosaccade blocks were counterbalanced within and across runs. Stimuli were presented in a fixed order to allow for individual differences analyses.

Keep track. The keep track task was adapted for an MRI environment from our previous version (see Friedman, et al., 2008 for a more thorough description of the task). Participants are given three or four categories for which they have to remember the most recent word in each category from a sequence of presented words. The categories for that trial remain on the screen throughout the process. In the current version of the tasks there is a "Get Ready" screen that is presented for 1500 ms, followed by a 500 ms inter-trial interval. Next, the three or four categories the participant needs to remember are presented for 2, 4, or 6 s. Then a sequence of 15 words, including both relevant and irrelevant words, was presented for 2 s each, followed by a 10 s verbal response period. Next there was a brief rest period (1500 ms) and another 500 ms inter-trial interval. A variable (20, 25, or 30 s) rest block would occur before the next block would start. There were 2 runs, each run including 3 blocks in which the participant had to maintain 4 categories, and 2 blocks in which the participant had to maintain 3 categories, for a total of 10 blocks. Each category contained 6 words. The categories and words within the categories were counterbalanced as much as possible across runs. Words did not appear across adjacent blocks.

Number–letter. The Number-letter task was adapted from our previous version of the task (adapted from Friedman et al., 2008). In this version there are four quadrants of a square where a number-letter pair (eg. 3E) was presented in one of the 4 quadrants. If the pair was presented in either of the upper two quadrants, then the participant was supposed to make a button press indicating whether the letter was a consonant (G, K, M, R) or a vowel (A, E, I, U). If the pair was presented in one of the lower two quadrants the participant would make a button press indicating whether the number was odd (3, 5, 7, 9) or even (2, 4, 6, 8). For each block, there would be a 10,000 s fixation, followed by a 500 ms blank, and then a 2 s, 4 s, or 6 s instruction period. After a 350 ms blank, the first stimuli would appear on screen for 3 s or until the participant responded using an MRI compatible button box. Then a second 350 ms blank occurred before the start of the next trial. There were 2 runs. Each run consisted of 8 mixed blocks, 4 fixation blocks, and 4 single-task blocks (2 "Letter" and 2 "Number" blocks). Mixed blocks consisted of trials were the stimuli were randomly presented in any of the 4 quadrants, where as Letter blocks consisted of trials were stimuli were only presented in the top two quadrants (letter judgment), and Number blocks consisted of stimuli only in the bottom two quadrants (number judgment). The order of the blocks was counterbalanced across the two runs.

MRI Parameters

MRI data were acquired on a Siemens 3Tesla MAGNETOM Trio with a 12-channel head coil located at the University of Colorado Boulder. Functional images used a T2*-weighted gradient echo, echo planar imaging (repetition time [TR] = 1900 ms, echo time [TE] = 25 ms, flip angle = 69°, 29 slices parallel to the AC-PC line, thickness = 3 mm, gap 1 mm, 64 x 64 in-plane resolution, in-plane FOV = 22 cm). A high-

resolution T1-weighted anatomical scan was collected for each participant to localize functional activity.

General Procedure

Tasks were administered using PsyScope X B51 (Cohen et al., 1993) on a Macintosh computer (Apple Computer, Cupertino, CA, USA). After the consent process, filled out safety screens as well as some self-report questionnaires. Then participants provided a urine sample to test for pregnancy in women and substance use and then completed a practice session before entering the MRI scanner. They were in the scanner for approximately an hour to complete the Antisaccade (inhibition), Keep track (updating), and Number-letter (shifting) tasks. Exclusion criteria included contraindications to MRI scanning, such as metal in the body or claustrophobia. All procedures were approved and carried out in accordance with the University of Colorado Boulder's Institutional Review Board.

Analyses

All analyses were carried out using SPM8 (Wellcome Department of Cognitive Neurology, UCL). All tasks were analyzed in an event-related design. Error trials were excluded for the Antisaccade and Number-letter tasks. For the Keep track task we did not exclude error trials as it is difficult to identify where the participant made an error within the block given only block-final recall. Rest was explicitly modeled as our task fixations for our baselines; therefore there was no need to regress out other forms of activation such as errors.

For the antisaccade task, prosaccade and antisaccade trials were modeled as the 232 ms cue plus the 150 ms target presentation. The 6-20 s fixation blocks served as

baseline. We had three regressors in our model, correct antisaccade trials, correct prosaccade trials, and rest fixation. Our main contrast was antisaccade versus fixation. We did look at antisaccade versus prosaccade, which is discussed in the supplement. Two subjects were excluded from the analysis of this task because they were performing at chance levels, presumably due to the increase in difficulty, and one for head movement.

Keep track trials consisted of each word as it was presented for 2 s. Baseline was composed of fixation blocks that were interspersed within the task. The contrast of interest for this task was between our two regressors, word events versus fixation blocks.

Number–letter trials started at the appearance of the cue and continued until the participant responded or until 3,000 ms passed. Our model included regressors for switch trials in mixed blocks, repeat trials in mixed blocks, repeat trials in number blocks, repeat trials in letter blocks, and rest blocks. The main contrast was switch trials within mixed blocks versus fixation. We did look at switch trials minus repeat trials within mixed blocks (i.e., local switch costs), and this contrast is discussed in the supplement. The first trials of each block were excluded because they were neither switch nor repeat.

For our individual differences analyses, we used the same subject-level model as we did for the ROI analyses for each task. These models consisted of regressors for the task condition, fixation, and 24 motion covariates, but we added a regressor at the group-level (between subjects) for Common EF group (coded as -.5 for low and .5 for high). We conducted separate analyses for the MD ROIs and the whole brain.

Regions of interest. Despite having a selected sample, we have a relatively small sample to test individual differences. Therefore we chose 27 ROIs in a frontoparietal network that is frequently activated across diverse cognitive tasks (from Federenko et al.,

2013). These ROIs were used as masks for contrasts of interest of each task at both the whole group-level and with Common EF as a covariate.

Whole-brain. Group-level analyses at the whole-brain level were corrected using a False Discovery Rate (FDR) to achieve an overall threshold of $q<.05$, controlling for multiple testing at the whole-brain level.

Analyses with Common EF as an independent variable were not corrected for multiple testing at a whole-brain level. Primary thresholds of $p<.001$, $p<.005$, and $p<.01$, were applied with corresponding cluster extent thresholds of k>5, k>10, and k>20 voxels, respectively. Clusters with the more relaxed thresholds were only shown if there was a contiguous cluster that met the p<.001 and k>5 thresholds.

Conjunction. In order to see what areas are common across each task type, we looked at overlap maps for all three tasks for our contrasts of interest. We also wanted to see if areas that are common across tasks at the whole group-level are the same as those that predict high or low EF. Therefore we also looked at an overlap map across all three tasks for regions predicting group differences. Each contrast for each task was thresholded at a *t*-distribution critical value that corresponded to a $p<.05$ level (not correcting for multiple testing at the whole-brain level). In order for a region to appear in the conjunction map, it had to be significant at a $p<.05$ level in each task; applying this threshold for each contrast resulted in an overall level of correction of $p<.000125$ (Fan et al., 2005; Wager et al., 2005).

Results and Discussion

Behavioral Results

Antisaccade accuracy significantly correlated with keep track accuracy and number–letter switch cost (*r*s=.51 and –.39 respectively), but keep track did not significantly correlate with number–letter (r=–.17). At the latent variable level, Shifting and Updating typically show a weaker correlation with each other than with Inhibition (Miyake & Friedman, 2012); therefore, at the individual task level, these results are consistent with the model from which these tasks were selected.

The high and low Common EF groups significantly differed on their in-scanner performance on all three tasks. The mean differences for antisaccade accuracy and keep track accuracy were 0.27 and 0.21, respectively, where individuals with higher Common EF were more accurate (antisaccade: M=0.62, SD=0.19; keep track: M=0.89, SD=0.08) than those with low Common EF (antisaccade: M=0.35, SD=0.12; keep track: M=0.69, SD=0.18), $t(25)=4.24$, $p<.001$ and $t(27)=4.07$, $p<.001$ respectively. The mean difference in number–letter switch cost was –187.60, where the high Common EF individuals showed a smaller switch cost (M=85.93, SD=68.25) than low Common EF (M=273.53, SD=163.44), $t(28)=–4.10$, $p<.001$.

We calculated a composite *z*-score for in-scanner behavior to compare to our Common EF latent factor. First we reverse-scored switch cost, then standardized each score for the three tasks, averaged the scores, and finally re-standardized our composite to create our measure of in-scanner Common EF. Common EF also significantly predicted mean differences in a composite *z*-score of in-scanner Common EF, with a mean difference of 1.66 (SD=0.20), $t(28) = 8.38$, $p<.001$.

ROI Analyses

Are Frontoparietal ROIs Active During All Three Tasks? We examined whether regions important for complex cognitive tasks in general are important for our three EF tasks at the group-level, using 27 regions of interest derived from the MD network (Fedorenko, Duncan & Kanwisher, 2013). Our task designs allowed for different contrasts. Both antisaccade and number–letter tasks included a difficult condition (antisaccade and switch trials) and an easier condition (prosaccade and repeat trials within the mixed block, respectively). However, the keep track task only had fixation blocks as the baseline. Therefore, a difference in difficulty of the contrasted conditions between keep track and the other two tasks exists. To ensure our results were not an artifact of baseline, such as increased power in the keep track contrasts due to task versus rest as opposed to hard task versus easy task, we focus on contrasts of the difficult task condition with fixation for all tasks in the following sections. Results from analyses with task-based baselines for the antisaccade and number–letter tasks are presented in the supplement.

As shown in Table 3.1, 16 of the 27 MD ROIs (59%) were significantly active across all 3 tasks. These ROIs (1, 2, 3, 7-11, 13-18, 21, 22) included bilateral Cerebellum I-IV and VI, right Precuneus-Lateral Occiptal Cortex (LOC), bilateral Frontal Operculum/Insula, bilateral Precentral Gyrus/Middle Frontal Gyrus (MFG), bilateral Supplementary Frontal Gyrus (SFG)/MFG, bilateral Brainstem, bilateral Thalamus, right Frontal Pole and right MFG.

Table 3.1. MD ROI activation by contrast

ROI Label	Anti-Fix	Switch-Fix	Keep-Fix
1 - VI cerebellum R	0.94**	0.36**	0.38**
2 - VI cerebellum L	1.39**	0.39**	0.40**
3 - I-IV cerebellum (bilateral)	1.36**	0.19**	0.25**
4 - Lateral occipital cortex / occipital fusiform gyrus L	0.20	0.22**	0.31**
5 - Precuneus / Lateral occipital cortex (superior division) L	0.59	0.97**	0.70**
6 - Lateral occipital cortex / occipital fusiform gyrus R	-0.10	0.15**	0.18**
7 - Precuneus / Lateral occipital cortex (superior division) R	0.89*	0.83**	0.63**
8 - Frontal operculum / Insula R	1.16**	0.31**	0.50**
9 - Frontal operculum / Insula L	1.15**	0.23**	0.47**
10 - Precentral gyrus / MFG R	1.11**	0.45**	0.46**
11 - Precentral gyrus / MFG L	0.63*	0.60**	0.73**
12 - SMA / SFG (bilateral)	0.41	0.48**	0.67**
13 - SFG / MFG / Precentral gyrus R	0.96**	0.63**	0.47**
14 SFG / MFG / Precentral gyrus L	0.78**	0.68**	0.48**
15 - Brainstem R	1.24*	0.23**	0.23**
16 - Brainstem L	1.28**	0.18**	0.18**
17 - Thalamus R	1.11**	0.23**	0.27**
18 - Thalamus L	1.07**	0.26**	0.29**
19 - Intracalcarine cortex L	0.82	-0.03	0.21**
20 - Intracalcarine cortex R	1.00	-0.14	0.19
21 - Frontal Pole R	0.86**	0.29**	0.56**
22 - MFG R	0.73*	0.43**	0.66**
23 - Frontal Pole L	0.11	0.23**	0.41**
24 - MFG L	0.14	0.56**	0.76**
25 - Cingulate gyrus (ant)	-0.64	0.08	0.33**
26 - Cingulate R	-0.69	0.09	0.28**
27 - Cingulate L	-0.73	0.09	0.25**

Note. Presented are unstandardized mean differences from 0 for each contrast versus fixation for each of the 27 Multiple Demand ROIs.
*p<.05, **p<.001.

The keep track task and number-letter task, but not the antisaccade task, showed significant activation for regions such as the left Precuneus, bilateral LOC/Occipital Fusiform Gyrus (OFG), bilateral Supplementary Motor Area (SMA), left Frontal Pole, and left MFG. This result could indicate that the antisaccade task requires fewer MD

network areas, but it is also possible that there was lower power to detect all regions due to fewer correct trials in the antisaccade analysis than in the other task analyses. Five regions (MD ROIs 19,20,25-27), which included the intracalcarine cortex and the cingulate cortex, showed activation only in the keep track task.

Do Individual Differences in Common EF Predict FP ROI Activation? We tested whether or not individual differences in Common EF predicted MD ROI activation for any of the tasks by testing for mean differences in activation for each ROI between Common EF groups. As shown in Table 3.2, Common EF groups differed in activation of 13 ROIs (or 48% of the MD network) in the keep track task, 8 ROIs (30%) in the number–letter task, and 2 ROIs (7%) in the ICC in the antisaccade task. The common activation across all three tasks for high Common EF – low Common EF contrast maps occurred only in these two ICC regions. Across all tasks, individuals with high Common EF activated ICC more strongly than their low Common EF counterparts.

Considering the conjunction of only the number–letter and keep track tasks, high Common EF individuals more strongly activated four other ROIs: left VI cerebellum, bilateral LOC/ occipital fusiform, and right SFG/MFG/precentral gyrus. With the exception of the ROI in the SFG/MFG/precentral gyrus, these ROIs are outside of the frontal and parietal cortex, and less often associated with executive control. As seen in Table 3.2, individuals with high Common EF uniquely activated bilateral precuneus in the number-letter task, and uniquely activated six other MD ROIs in the keep track task.

Table 3.2. MD ROIs mean differences in Common EF ability

ROI Label	Anti-Fix	Switch-Fix	Keep-Fix
1 - VI cerebellum R	0.44	0.50	0.52
2 - VI cerebellum L	0.50	0.91*	0.73*
3 - I-IV cerebellum (bilateral)	0.73	0.34	0.65
4 - Lateral occipital cortex / occipital fusiform	0.58	0.85*	0.84*

gyrus L			
5 - Precuneus / Lateral occipital cortex (superior division) L	0.24	0.78*	0.77
6 - Lateral occipital cortex / occipital fusiform gyrus R	0.72	0.81*	0.84*
7 - Precuneus / Lateral occipital cortex (superior division) R	0.12	0.96*	0.54
8 - Frontal operculum / Insula R	-0.02	0.32	0.81*
9 - Frontal operculum / Insula L	0.01	-0.08	0.61
10 - Precentral gyrus / MFG R	0.52	0.45	0.58
11 - Precentral gyrus / MFG L	0.40	0.32	0.79*
12 - SMA / SFG (bilateral)	0.37	0.61	0.79*
13 - SFG / MFG / Precentral gyrus R	0.44	1.00*	0.81*
14 SFG / MFG / Precentral gyrus L	0.16	0.65	0.68
15 - Brainstem R	0.42	0.57	0.86*
16 - Brainstem L	0.17	0.36	0.84*
17 - Thalamus R	0.30	0.32	0.69
18 - Thalamus L	0.44	0.20	0.74*
19 - Intracalcarine cortex L	0.95*	0.98*	0.72*
20 - Intracalcarine cortex R	0.82*	0.92*	0.84*
21 - Frontal Pole R	-0.23	0.25	0.70
22 - MFG R	-0.24	0.23	0.59
23 - Frontal Pole L	0.13	0.18	0.55
24 - MFG L	-0.00	0.04	0.59
25 - Cingulate gyrus (ant)	-0.30	-0.42	0.74*
26 - Cingulate R	-0.35	0.06	0.45
27 - Cingulate L	-0.26	-0.01	0.63

Note. Mean differences in standardized MD ROIs activation for each ROI are presented for contrast versus fixation for individuals grouped into high or low Common EF ability. The Common EF score was derived from a latent variable model in the previous wave of data collection, approximately 7 years prior to the current data.
*p<.05.

We also examined whether we would find the same results using the in-scanner Common EF *z*-composite, since these scores were more temporally proximal, although with more noise than the full 9-task model. In-scanner Common EF predicted seven ROIs in the keep track task (26%), three in the number–letter task (11%), and one in the antisaccade task (4%) (see Supplemental Table S3). No region was predicted in all three tasks by in-scanner Common EF, however multiple regions were predicted by at least two

tasks. As was the case for out-of-scanner Common EF, in-scanner Common EF predicted bilateral ICC activation for the number–letter task, and right ICC activation for the keep track task. Individuals with higher in-scanner Common EF showed increased activation in the bilateral I-IV cerebellar ROI for the antisaccade and keep track tasks, and left VI cerebellar activity for the keep track and number–letter tasks. Common EF predicted activation in four other regions for the keep track task that were also predicted by our out-of-scanner Common EF grouping variable.

Although similar patterns emerged for activation of ROIs with out-of-scanner and in-scanner Common EF variables, the out-of-scanner Common EF variable, which was based on a latent factor, was more related to MD ROI activation, despite being collected approximately seven years prior to the scans. This result provides evidence of the stability of Common EF at the latent level (see also Friedman et al., 2016).

Do Correlations Suggest a Common Underlying Construct? If MD activation reflects a common underlying process that drives activation, we should see correlations of activation levels within the same ROIs across the three tasks. No MD ROIs showed correlations in activation for all three pairwise possibilities, however there were some correlations in the level of activation within participants across specific pairs of tasks. In general, correlations across tasks ranged from r=–0.47 to r=0.58. Average correlations between keep track and the other two tasks were around M=0.23 to 0.24 (SD=0.18 to 0.19), while the average correlation between antisaccade and number–letter was M=–0.01 (SD=0.21).

Examining the pairwise correlations in Table 3.3, four ROIs correlated across the antisaccade and number-letter tasks. Three ROIs correlated across the antisaccade and

keep track tasks, and eight ROIs correlated across the keep track and number–letter tasks. However, these ROIs were different across pairs; only one ROI was involved in multiple correlations. Specifically, activation in the left VI cerebellum during the number–letter task positively correlated with its activation during the antisaccade and keep track tasks (rs=.58 and .48, respectively); however its activation during the antisaccade and keep track tasks did not correlate with each other (r=.20).

Table 3.3. Correlations in MD ROI activation

ROI Label	Antisaccade with Number–letter	Keep track with Number–letter	Antisaccade with Keep track
1 - VI cerebellum R	0.22	0.49**	-0.04
2 - VI cerebellum L	0.58**	0.48**	0.20
3 - I-IV cerebellum (bilateral)	-0.01	-0.06	0.00
4 - Lateral occipital cortex / occipital fusiform gyrus L	0.49**	0.14	0.04
5 - Precuneus / Lateral occipital cortex L	0.29	0.36	-0.15
6 - Lateral occipital cortex / occipital fusiform gyrus R	0.49**	0.14	-0.05
7 - Precuneus / Lateral occipital cortex R	0.33	0.24	-0.19
8 - Frontal operculum / Insula R	-0.02	0.23	0.03
9 - Frontal operculum / Insula L	0.04	0.00	-0.11
10 - Precentral gyrus / MFG R	0.37	0.40*	0.20
11 - Precentral gyrus / MFG L	0.35	0.43*	0.19
12 - SMA / SFG (bilateral)	0.39*	0.14	0.15
13 - SFG / MFG / Precentral gyrus R	0.35	0.48**	-0.08
14 SFG / MFG / Precentral gyrus L	0.36	0.38*	0.00
15 - Brainstem R	0.13	0.30	-0.07
16 - Brainstem L	0.03	0.10	-0.14
17 - Thalamus R	0.27	-0.16	-0.06
18 - Thalamus L	0.35	-0.21	-0.06
19 - Intracalcarine cortex L	0.35	0.20	0.49**
20 - Intracalcarine cortex R	0.25	0.20	0.51**
21 - Frontal Pole R	0.20	0.31	-0.19
22 - MFG R	0.09	0.24	-0.17
23 - Frontal Pole L	-0.22	0.17	-0.47*
24 - MFG L	0.17	0.28	-0.21

25 - Cingulate gyrus (ant)	0.24	0.23	0.01
26 - Cingulate R	0.12	0.51**	-0.13
27 - Cingulate L	0.12	0.46*	-0.04

Note. Presented are the correlations in activation across pairs of task contrasts versus fixation for each MD ROI.
*p<.05, **p<.001.

Correlations were higher within tasks (see Supplementary tables S3-S7). That is, individuals with high activation in one ROI during a task also tended to show higher activation in other MD ROIs during the same task. However, the between-task correlations suggest that those individuals did not necessarily show higher activation in the other tasks. Taken together, these correlational patterns suggest that the ROIs in the MD network work together during each task, but do not relate to individual differences in Common EF.

Whole Brain Analyses

The ROI analyses suggested that though many of the MD ROIs were active during the three EF tasks, they did not relate to individual differences in performance. Thus, we conducted exploratory whole-brain analyses to examine whether there were regions outside of the MD network associated with these tasks, at the group-level and individual differences level. In the following sections we focus on the conjunction maps at the group and individual differences levels; see the supplemental materials for maps for each task contrast.

Which brain regions are active at the whole-brain level across all three tasks? While the whole-brain maps at the group-level tell us about activation in each task, a question of interest is which regions are active across all three tasks. To answer this question, we calculated the whole-brain conjunction map for group-level activation. In order for a region to be included in the conjunction map, it had to be significant at the

p<.05 level for each task, resulting in an overall correction of p<.00013. As shown in Figure 3.1, we observed a large cluster of positive activation in the cerebellum, including bilateral portions of Crus I, II, and IV, bilateral middle and superior frontal gyri, insula, thalamus, and frontal pole (for list of all regions, see Table 3.4 & 3.5). Consistent with the results of the group-level activation in the ROIs, many of our positive regions overlapped with the MD ROIs. However, we also observed elevated activation across all three tasks in the cerebellar cluster, putamen, right superior temporal gyrus, and bilateral supramarginal gyrus.

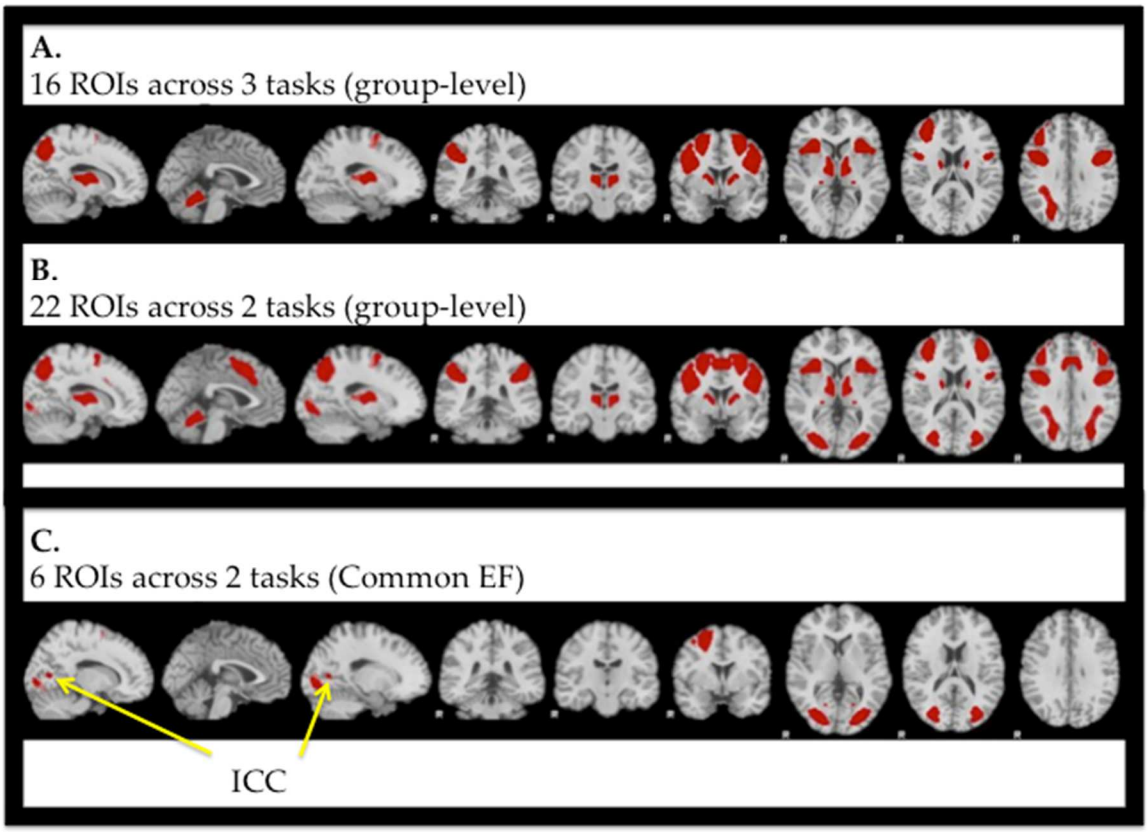

Figure 3.1. Panel A: 16 significant group-level MD ROIs across all three tasks versus fixation. Panel B: 22 significant group-level MD ROIs across two tasks (keep track and number-letter) versus fixation. Panel C: 6 MD ROIs that were more strongly activated by high Common EF individuals across two tasks (keep track and number-letter) versus fixation. High Common EF individuals more strongly activated two MD ROIs (bilateral intracalcarine cortex, ICC) across all three tasks versus fixation. These regions are indicated by the yellow arrows. No ROIs were associated with deactivation across multiple tasks.

Table 3.4. Clusters in group-level conjunction map

index	x	y	z	voxels	volume_mm3
1	-2	-16	32	52618	420944
2	-30	-80	-38	492	3936
3	32	-76	-38	666	5328
4	54	44	-10	4	32
5	2	-86	-10	18	144
6	-30	-60	-8	9	72
7	-22	-76	-4	80	640
8	56	38	4	56	448
9	54	-62	22	1010	8080
10	4	-72	0	3	24
11	62	6	2	3	24
12	52	-80	4	2	16
13	54	-76	8	2	16
14	-12	16	16	4	32
15	-42	22	48	2	16
16	-40	16	56	2	16
17	-34	-54	68	2	16

Note. Presented are significant clusters in our conjunction map showing group-level activation. No significant negative clusters. $p<.05$ for each contrast, resulting in an overall contrast of $p<.000125$.

What regions predict individual differences in Common EF in all 3 tasks? To examine whether the regions that are important for performing the three tasks predict individual differences in Common EF, we computed a similar conjunction map of areas that related to the Common EF grouping variable. As shown in Figure 3.2, high Common EF participants activated a cluster in the ICC and lingual gyrus, some clusters in the superior parietal lobule, cerebellum IV & V, temporal fusiform gyrus, bilateral superior temporal gyrus, and the LOC (for list of all regions, see Table 3.5). Most of these regions fell outside of the MD ROIs. Within the MD ROIs there was one small cluster in the right cerebellum, Crus I. These results generally suggest that the areas predicting Common EF are not those that are activated at the group-level, indicating that

regions that predict individual differences are outside of those that are necessary in order to complete EF tasks.

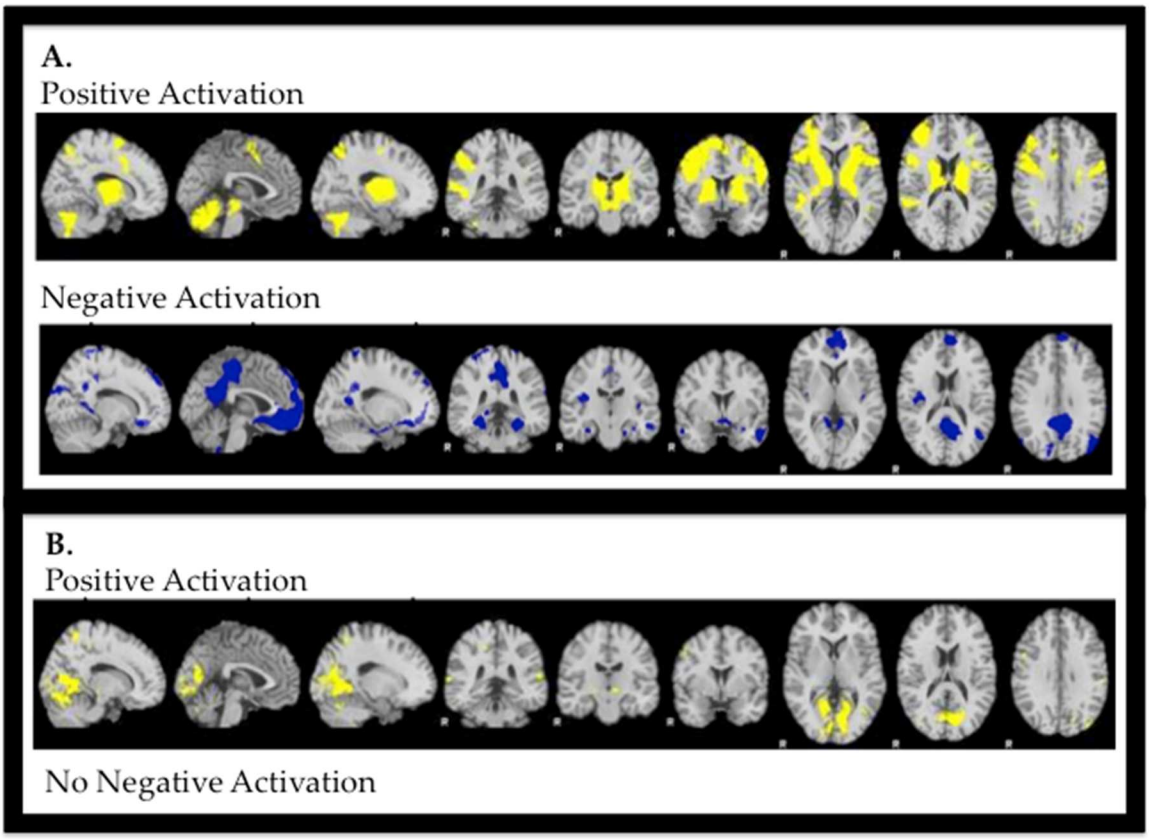

Figure 3.2. Conjunction map for task minus fixation. Panel A shows a conjunction map for the group-level activation for all three tasks versus fixation. Panel B shows a conjunction map for all three tasks versus fixation of regions predicted by Common EF ability. All tasks were thresholded at the $p<.05$ level, for a combined threshold of $p<.000125$ across each map.

Table 3.5. Clusters in conjunction map of Common EF ability

index	x	y	z	voxels	volume_mm3	numpeaks
1	-8	-70	-36	6	48	6
2	-24	-50	-34	2	16	2
3	-32	-46	-32	12	96	12
4	-32	-54	-28	3	24	3
5	12	-70	-14	93	744	93
6	-22	-62	-22	8	64	8
7	-10	-72	-20	5	40	5
8	36	-62	-20	2	16	2
9	32	-70	-12	63	504	63
10	-2	-88	-4	156	1248	156
11	-6	-70	8	900	7200	900

12	-26	-78	-8	3	24	3
13	18	-22	-4	15	120	15
14	-8	-16	-4	2	16	2
15	-4	-94	10	9	72	9
16	66	-30	10	57	456	57
17	-60	-34	14	14	112	14
18	-20	-64	22	2	16	2
19	-38	-84	36	16	128	16
20	-26	-26	36	6	48	6
21	52	0	46	3	24	3
22	8	6	48	9	72	9
23	-54	-6	50	2	16	2
24	-12	-54	62	21	168	21
25	14	-54	64	43	344	43
26	26	-54	66	10	80	10
27	10	-8	68	2	16	2

Note. Presented are significant clusters in our conjunction map where activation is associated with Common EF ability. No significant negative clusters. $p<.05$ for each contrast, resulting in an overall contrast of $p<.000125$.

General Discussion

The goals of this study were to better understand to what degree the MD network, a network that responds to a wide array of cognitive tasks, underlies Common EF and whether or not MD region activations predict individual differences in Common EF. A substantial number of MD ROIs were active across our three EF tasks, with participants activating the most ROIs for the updating tasks and the fewest for the inhibition task. However, two regions that were not active across all three tasks, the right and left ICC, were the only regions that were significantly predicted by individual differences in Common EF in all three tasks. Results from both the ROI and whole brain analyses suggested that individual differences in Common EF are related to activation in regions outside of the frontal cortex, with the exception of clusters in the frontal pole and in the middle frontal gyrus. The whole brain conjunction map also suggested that areas outside

of the MD network, such as the fusiform cortex, the superior temporal gyrus, and portions of the cerebellum, are important for individual differences in Common EF.

This study advances the literature on the neural correlates of Common EF in several ways. First, we focused on understanding the fMRI predictors of individual differences, which are less well studied than group mean activation. A frontoparietal network has been strongly associated with group-level activation across multiple types of EF tasks, however less is known about what regions predict individual differences in EF reliably across tasks. Second, we examined individual differences in functional activation during tasks selected to tap three important components of EF, based on a well-validated model. This model allowed us to focus on a Common EF factor, which is related to psychopathology, IQ, and various behaviors (Miyake & Friedman, 2012, Snyder, 2015). Prior reports with this model suggest substantial stability of individual differences (Friedman et al., 2016), which we leveraged to form another strength of the study: We used an extreme groups design to maximize Common EF variance in this relatively small sample. These groups were selected on the basis of a full latent variable model of nine tasks assessed in a sample of over 700 individuals. Although these groups were selected to be at least a standard deviation above or below the mean based on data collected 7 years earlier, they were still very different at the current time point; specifically they differed by 1.66 standard deviation units in their aggregate performance in the scanner. Thus, we were able to examine predictors of a stable individual difference. Taken together, our results indicate that areas that are commonly activated during EF tasks are not necessarily the same ones that predict individual differences in performance.

The MD Network and EF

We expected our tasks to activate many of the MD network regions, given that these MD ROIs were based on activation in a variety of cognitive tasks. Almost every MD ROI was significantly active in at least one of our three tasks, with the exception of the right ICC. Sixteen ROIs were active during all three tasks, demonstrating that the MD network overlaps with a higher-level EF network. Many of the regions of our group-level conjunction map also fell within or overlapped with the MD network. However, there was not complete overlap between the MD ROIs and activation from our EF tasks. Specifically, the ICC was not active in all three tasks, and other regions, such as the bilateral cingulate cortex, were not active in the shifting and inhibition tasks. Also, in our group-level conjunction map, many clusters extended outside of the MD ROIs and some regions fell outside of the MD network, such as the right cerebellum. The observed patterns suggest that while strongly related, the MD network is not synonymous with a Common EF network.

We then examined whether regions active during EF tasks at the group-level also influenced individual differences in EF in order to address two questions. One, do individuals with high and low Common EF ability differentially activate regions in the MD network? Those with high Common EF might differentially recruit particular regions that are active across all subjects. Two, do individuals with high or low Common EF recruit areas outside MD network? It is possible that those with high Common EF recruit additional regions outside of the core network which contribute to their higher performance, or that those with lower Common EF recruit additional regions to try and compensate for poorer performance.

To address the first question, we examined whether individual differences in Common EF predicted activation of each MD ROI. While Common EF ability predicted differences in activation in some MD ROIs, the majority of the ROIs were unaffected. The most consistent pattern was that individuals with high Common EF ability showed increased bilateral activation in the ICC compared to those with low Common EF ability across all 3 tasks, though at the group-level the right ICC was not active in any of the tasks and the left ICC was only active in the keep track task. This pattern indicates that individuals with high Common EF ability recruit the ICC, but individuals with lower Common EF ability do not. The ICC is a part of the primary visual cortex that has previously been associated with increased cognitive demand in participants in their early twenties (Stern et al., 2005).

We also found four ROIs that related to Common EF in two out of three tasks: left VI cerebellum, left and right occipital fusiform gyri, and right superior/medial frontal gyrus. These regions were significantly activated at the group-level, and therefore it seems that individuals with both high and low EF recruit these regions, but those with high Common EF ability activate them more strongly. It is not clear why activation of these same ROIs during the antisaccade task did not relate to Common EF. Although it is possible that the task requirements in that task were different, it is also possible that the antisaccade contrast had lower power because we focused only on correct trials.

In summary, there was limited evidence for activation levels of these ROIs across tasks being related to individual differences in Common EF ability. Moreover, an examination of the correlations of each ROI's activation across tasks provided little evidence for a common factor (i.e., perhaps unrelated to Common EF). Only one region

showed correlated activation in more than one pair of tasks. Left cerebellar activation during the keep track task was positively correlated with cerebellar activation during the antisaccade and the number–letter tasks, but cerebellar activation for the antisaccade task did not correlate with that for the switch task. The lack of correlation between tasks could be due to low variance, where everyone activated the ROIs to a similar extent, or to sufficiently different contexts across the tasks. In either case, this result suggests that there is not some unifying construct within the MD ROIs that reflects stable individual differences in EFs across tasks.

To answer the second question, whether individuals with high or low Common EF recruit areas outside MD network, we created a whole-brain conjunction map of areas that related to individual differences in Common EF across all three tasks. Many of the regions identified in the individual differences conjunction map, such as the LOC, occipital fusiform gyrus, lingual gyrus, and cerebellum, have previously been associated with cognition. Specifically these regions have previously been associated with object recognition and working memory load (LOC; Grill-Spector, Kourtzi, & Kanwisher, 2001; Taylor et al., 2004), face and word recognition (occipital fusiform gyrus; Rossion et al., 2013), visual processing of letters, and analysis of logical conditioning (lingual gyrus; Mechelli et al., 2000; Brunet et al., 2000). Given that many of the areas we found fell outside of the MD ROIs and our group-level conjunction map, individual differences research should not restrict the search to the areas significant at the group-level (Friedman & Miyake, 2016; Yarkoni & Braver,2010)—a network mask based upon which regions activate during a task might cause the researchers to miss areas important to individual differences.

Reineberg et al. (2015) also found that individuals with higher Common EF had expanded connectivity of a frontal-parietal resting state network to the Crus I and II of the cerebellum in an independent sample. This result was consistent with previous functional work that found increased functional connectivity to and activation in crus I/II predicted better working memory and EF performance (Bernard et al., 2013; Salmi et al., 2010; Stoodley & Schmahmann, 2009). Activation in the cerebellum has been associated with various EFs, yet the role of the cerebellum in cognition is still unclear (Stoodley et al., 2012; Koziol et al., 2013). One possibility is that the cerebellum may play a role in the automation of cognitive processes or inner speech involved with verbal working memory (Koziol et al., 2013).

Somewhat surprisingly, we found a relative lack of prefrontal cortical regions predicting individual differences in Common EF. The prefrontal cortex is integral to higher-level cognitive processes, and when damaged, EF deficits are often observed (for reviews see Alverz & Emory, 2006; Stuss & Alexander, 2000). Common EF predicted activation in a few regions in the frontal cortex. In our ROI analysis, Common EF predicted activation in the right MFG/precentral gyrus MD ROI in two out of three tasks. In our conjunction analysis, we found a region that overlapped with the posterior portion of that MD ROI and extended beyond it. We also found a frontal polar region outside of the MD network. However, most areas that related to individual differences in Common EF were more posterior.

One possible explanation for the relative lack of frontal regions related to individual differences is that everyone generally uses the PFC to complete the task, but those with better Common EF have a flexible enough frontoparietal network that they can

recruit additional regions outside of the PFC when tasks are sufficiently difficult. If tasks are easy, and everyone is performing at ceiling, all individuals might show approximately the same levels of activation, since those with higher EF do not need to recruit additional areas in order to perform better. However, if the task is difficult there might be more variance in brain activation, both in terms of how strongly an individual activates a given region, but also in terms of whether an individual recruits additional regions to help with the task. This pattern of spatially dissociable regions for group-level and individual differences is consistent with results described by Yarkoni & Braver (2010). Another potential explanation is that lower-level processing areas can be crucial for individual differences in more complex behaviors.

Working Memory, EF, IQ, and the MD Network

While Duncan and colleagues never explicitly argue that the MD network underlies executive functions, the MD network is a frontoparietal network that responds to a variety of cognitive tasks, and therefore seemed appropriate as candidate regions. One possible explanation for why we did not find that the MD network predicted individual differences in Common EF is that the MD network is more about intelligence than Common EF per se. The MD network activates in response to general fluid intelligence tasks such as those used by Bishop et al. (2008). They used both verbal and spatial problem solving tasks to tap g_f, which activated the MD network. In fact, their whole-brain corrected maps did not find any regions outside of the their network ROIs, suggesting high, if not complete, overlap between g_f and the MD network, at least at the group-level. However, to-date little research has focused on individual differences in the MD network.

While intelligence is correlated with EFs (Carpenter, Just, & Shell, 1990; Engle, Tuholski, Laughlin, & Conway, 1999; Friedman et al., 2006; Salthouse, Fristoe, McGuthry, & Hambrick, 1998), the Common EF factor is not the same as *g*. Friedman et al. (2008) found that the genetic correlation between Common EF and full-scale IQ was only 0.57, indicating substantial genetic separability between the two constructs. Moreover, EFs predict variance above and beyond intelligence in behaviors such as attention problems and self-restraint (Friedman et al., 2007; Friedman et al., 2011; see Friedman & Miyake, 2016, for a review).

Although Englehardt et al. (2016) found high genetic overlap between a Common EF factor and intelligence, their hierarchical Common EF factor had very high loadings for Updating and Working Memory subfactors, in contrast with the Common EF factor we examined here (see Friedman et al., 2008, for the analogous hierarchical model). Their result is thus very consistent with the general finding that intelligence is closely related to updating abilities(Friedman et al., 2006) and working memory, which are closely related constructs (Schmiedek, Hildebrandt, Lövdén, Lindenberger, Wilhelm, 2009). For example, Colom and colleagues (2015) found a strong correlation (r=0.86) between latent variable measures of working-memory and fluid intelligence. Updating, at the general level, is a combination of Common EF and an Updating-Specific factor, which both are equally related to IQ (Friedman et al, 2008). Therefore, the MD network is likely tapping a neural network for working-memory and/or fluid intelligence that is related to Common EF, but also Updating-Specific ability. If so, then we would expect to find partial overlap between our EF tasks and the MD network. However, we might expect greatest overlap for the updating task.

Consistent with this idea, the keep track task involved the most MD ROIs at the group-level and individual differences level. This could indicate increased power in the keep track task, or it could indicate that the MD network is more closely related to updating working memory than EF in general. Unfortunately we could not directly test if an Updating-Specific latent factor predicted more MD ROIs than a Common-EF factor, given that our sample was selected for Common-EF. (These individuals also varied to some extent on the other EFs and IQ, but these variations were confounded with Common EF differences.)

Limitations

One limitation of the study is that it was likely underpowered for whole-brain individual differences analyses, despite the increased power afforded by the selected sample. We focused on ROIs for this reason, but we also presented exploratory whole-brain analyses. Moreover, although our sample was larger than the average individual differences study in contemporary neurocognitive literature (27-30 subjects compared to 15-20; Yarkoni & Braver, 2010), it is possible that it was still underpowered to detect neural correlates of individual differences, even with ROIs. Thus, it may be the case that areas within the MD network significantly predict individual differences in a larger sample. However, these effects may be relatively weak, given that the selected sample we used may also have inflated our effect sizes compared to a population sample.

In terms of the group-level effects, a potential source of difference between our study and Fedorenko et al.'s (2013) study was the increased variance of our sample. Our subjects were selected based on being at least one standard deviation above or below the mean for a Common EF factor, whereas the Fedorenko et al., (2013) sample, from which

the ROIs were derived, presumably followed a normal distribution with a majority of the participants around the mean of Common EF ability. It is possible that individuals high or low in Common EF show different neural patterns than those closer to the mean, decreasing our power to significantly detect activation in some of the ROIs. If individuals who differ in EF ability have different neural patterns, then regions most related to individual differences in EF would likely fail to activate at the group-level. Indeed we observed that particular pattern for the ICC ROIs, which were related to Common EF across all three tasks. However we did observe some other MD regions related to individual differences in Common EF ability for particular tasks or pairs of tasks.

Another limitation is that we selected participants on the basis of Common EF, so we did not have sufficient variance in Updating-Specific, Shifting-Specific, or intelligence factors to test the different neural patterns between the constructs. Therefore we could not test the possibility that the MD network is more related to Updating or intelligence than Common EF. A balanced design with equal numbers of individuals with high and low IQ (or Updating-Specific or Shifting-Specific) within the high and low Common EF groups, or a sufficiently large sample size to allow for estimating of each variable would allow us to tease apart the contributions of intelligence and multiple EFs. We are currently collecting data for such a study.

Conclusion

In a selected sample of individuals with high and low Common EF ability, we demonstrated that group-level activation across three EF tasks overlaps with the MD network. However, high and low Common EF individuals did not differentially recruit any of these regions across all three tasks. The only region within the MD network that

was more strongly activated in individuals with high Common EF across all three tasks was bilateral intracalcarine cortex, which was not active at the group-level across these three tasks. Exploratory whole-brain conjunction analyses also suggested that individuals might recruit regions outside the MD network to boost task performance. The results suggest that EF tasks recruit areas that overlap with the MD network. While the MD network and a Common EF network seem to share neural correlates distributed throughout the brain, they also have regions that seem to be unique to each system. However, these common and unique regions at the group-level do not seem to be important for individual differences in Common EF. These results are consistent with Yarkoni and Braver's (2010) suggestion that the areas important for individual differences are not necessarily be those that are active at the group-level. In terms of the current study, the MD network may be necessary to complete EF tasks, but individual differences may be more related to non-MD areas that are differentially recruited to augment performance.

CHAPTER 4
Phenotypic Interplay of Sleep, Depression, and Executive Functions

Atypical sleep and depression are associated with executive dysfunction, and atypical sleep is related to increased depression, but the exact nature of how sleep, depression, and executive functions (EFs) work together to influence each other remains unknown. In this study, I investigate how sleep duration and depression might both contribute to executive dysfunction, as well as the degree to which genes and environment influence the relationships among these three variables. To do this I use data on depression, sleep duration, and EFs from a longitudinal twin study from early adolescence to young adulthood (approximately ages 12 to 23 years). Specifically, I first investigate whether individual differences in sleep duration are associated with depression and EF abilities, and the longitudinal relations among these variables. Next I ask whether sleep duration explains the relationship between depression and EF, or if depression explains the relationship between sleep duration and EF. Then I examine how stable these genetic and environmental influences are across time, and when new genetic and environmental factors arise during different developmental stages. Lastly, I examine the genetic and environmental correlations between sleep duration, EF, and depression to better understand the nature of the relationships between these variables.

Depression and Sleep

A good deal of research supports the link between sleep characteristics and depression. In fact, sleep is often disordered in those with depression to the point that it is a criterion for Major Depression Disorder (MDD) as diagnosed using the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-V*; American Psychiatric Association, 2013). Also, sleep characteristics, such as rapid eye movement sleep, have

been proposed an endophenotype – intermediate phenotype on the path between genes and disorder diagnosis – for MDD (Hasler et al., 2004; Modell & Lauer, 2007). Insomnia also has a bidirectional relationship with depression. Insomnia has often been considered a secondary disorder to depression, however there is also research suggesting that having insomnia increases the risk of depression in follow-ups 1 year to 34 years later (see Baglioni et al., 2011 for a review). Other sleep characteristics, such as sleep duration (which will be the focus of this dissertation), have an established, but less clear-cut association with depression (van Mill et al., 2010; Zhai et al., 2015; Raniti et al., 2016).

Individual differences in sleep duration have been inconsistently associated with depression. Some research finds an association between short sleep duration and depression (e.g., Park et al., 2010). Other studies, such as a meta-analysis by Zhai et al. (2015), show increased relative risk of depression with both long and short sleep durations (relative risk 1.42 and 1.31, respectively) compared to "normal" sleep durations. And yet other studies find no relationship between depression and sleep duration (e.g., Supartini et al., 2016). These studies differ in nationalities of the samples, how depression and sleep duration are measured, as well as the age of the samples, indicating that the relationship between depression and sleep duration is possibly influenced by these factors.

Problematic sleep is associated with depression in childhood and throughout adolescence. For example, a variety of parent-reported sleep problems were moderately correlated with depression symptoms in 8 year olds (Gregory et al., 2016). Both longitudinal and cross-sectional studies suggest short sleep duration and sleep disturbances are prospective predictors of adolescent depression (Breslau, et al., 1996;

Chang et al., 1997; Roberts & Duong, 2014). I add to this literature by assessing whether short, long, or both types of sleep durations influence depression at three different time points from adolescence to young adulthood. I also assess whether changes in sleep duration influence depression at a later age, as well as whether changes in depression influence sleep duration throughout development.

Executive Functions, Sleep, and Depression

EFs are related to sleep (Boonstra, Stins, Daffertshofer, & Beek, 2007) and depression (see Snyder, 2013 for a review). EFs are higher order cognitive processes that help individuals regulate thoughts and behaviors in order to achieve goals. They are associated with later life outcomes, such as education and occupation (Best et al., 2009; Miller et al., 2012; Valliente et al., 2013), and are often impaired in those with psychopathology (Snyder et al., 2015). In fact, EF is often suggested as an endophenotype for multiple forms of psychopathology, including MDD (e.g., Hasler et al., 2004).

Multi-component EF model. Most studies use various measures of EF, and often equate one task to one type of EF, despite task impurity. Task impurity refers to the problem that an EF task may actually tap non-EF processes, and even multiple EF processes, rather than just the EF of interest. For example, the Stroop task, which is considered an EF task, involves lower level visual processing, word-reading, color processing, as well as other cognitive processes, all of which are not EFs, but contribute to the performance on that task. The Stroop task was designed to assess individual differences to inhibiting the reading response, yet all of these other processes are contributing to the final score. However, the current study uses the Unity and Diversity

model of EF (Miyake et al., 2000; Miyake & Friedman, 2012), which focuses on latent variables. Latent variables allow for purer measures of EF because task-specific processes and measurement error are removed. EF is measured by a nine-task battery, which includes 3 tasks selected to tap response inhibition, 3 tasks designed to assess updating working memory, and 3 tasks targeting set-shifting.

These tasks result in 3 latent variables (see Figure 4.1; reproduced from Friedman et al., 2016): a Common EF variable that explains covariance across all 9 tasks, an Updating-Specific variable that explains residual covariance across the 3 updating tasks (i.e., not explained by the Common EF factor), and a Shifting-Specific variable that explains residual covariance across the 3 shifting tasks. There is no Inhibition-Specific variable after accounting for Common EF, because there was no inhibition task covariance leftover after accounting for covariance explained by the Common EF factor.

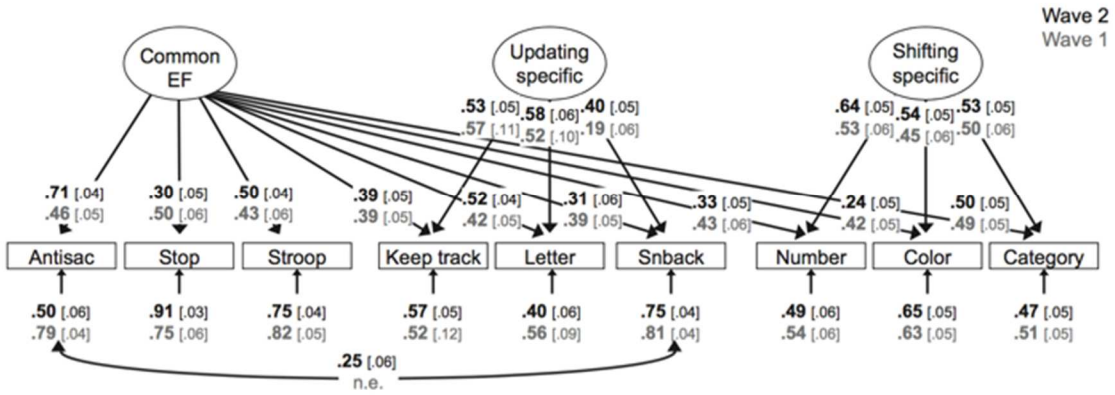

Figure 4.1.
Bifactor latent variable model of executive functions (EF) data with standardized loadings for late adolescence (Wave 1 in grey; approximately age 17) and early adulthood (Wave 2 in black; approximately age 23) reproduced from Friedman et al., 2016. Numbers on arrows are standardized factor loadings, those under the smaller arrows are residual variances, and those on curved double-headed arrows are interfactor correlations. Numbers in brackets are standard errors. There is a Common EF latent variable on which all nine EF tasks load, as well as two "nested" latent variables on which the updating and shifting tasks, respectively, also load. The Common EF variance is isomorphic with the Inhibiting latent variable, so there was no inhibiting-specific variance at either time point. Because the Common EF factor captures the variance common to all three EFs, the Updating-Specific and Shifting-Specific factors capture the variance that is unique to Updating and Shifting, respectively. Hence, they are uncorrelated with the Common EF factor and with each other. All parameters were statistically significant (p .05).

Antisac antisaccade; Stop stop-signal; Letter letter memory; Snback spatial *n*-back; Number number–letter; Color color–shape; Category category-switch; n.e. not estimated at Wave 1.

The three EF latent factors differentially relate to important constructs, such as psychopathology and intelligence. The inhibition or Common EF factor has previously been negatively associated with attention problems and conduct disorder (Friedman et al., 2007; Young et al., 2009; Gustavson et al., 2015), while both the Common EF and the Updating-Specific factor have been positively associated with IQ (Friedman et al., 2006). Interestingly, the Shifting-Specific factor has been positively associated with problematic behaviors, such as Attention Deficit/Hyperactivity Disorder, and Behavioral Disinhibition (Miyake & Friedman, 2012; Herd et al., 2014). These results demonstrate that the factors uniquely relate to disorders and problematic behavior, and therefore the different factors might differentially relate to sleep and depression.

EF and Depression. A review on the relationship between depression and other measures of EF by McClintock and colleagues (2010) reveals a mixed relationship between MDD and cognitive impairments. While there is a significant body of research that suggests those with MDD have impaired cognition, it may be influenced by age (older individuals are more affected), type of EF assessment, severity, medication, and comorbidity. A meta-analysis and review by Snyder (2013) found that there is a robust relationship between MDD and measures of EF, with effect sizes ranging from relatively small (d=.32) to large (d=.97). These findings prompted the speculation that depression might be related to Common EF, given that the three factors differentially related to psychopathology.

Confirming this speculation, ongoing research by Friedman (presented in 2016; reproduced in Figure 4.2) show that, in the longitudinal sample examined here,

depression symptom scores at age 12 negatively predicted Common EF and Updating Specific latent factors at age 17. Age 17 depression scores negatively predicted both Common EF at age 17 and Updating at age 23 controlling for previous depressive symptoms and EF abilities. A similar, but slightly stronger, pattern emerged when using lifetime MDD diagnosis at age 12, and past year MDD diagnosis at ages 17 and 23 instead of symptom scores, with the exception that age 12 lifetime diagnosis no longer predicted Updating-Specific abilities at age 17.

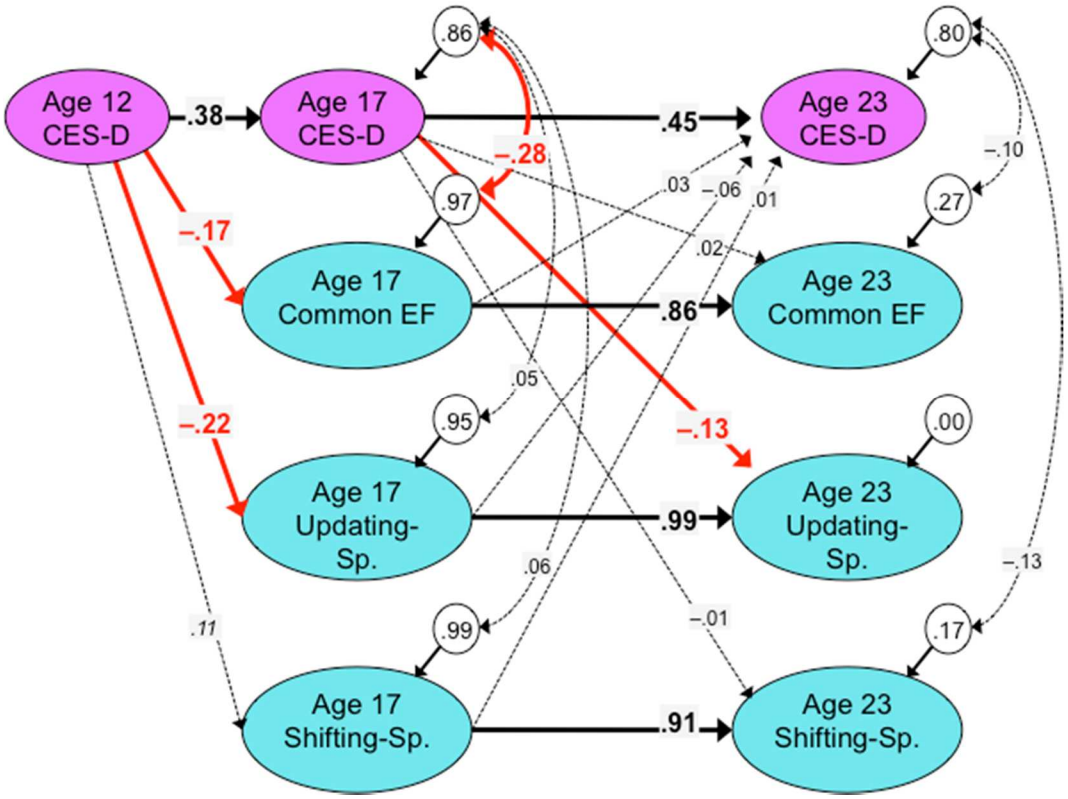

Figure 4.2.
Cross-lag model of the effects of depression symptoms on EF. Reproduced from Friedman et al. (in preparation). Red lines = $p<.05$. Black Bold lines = Stability effects with $p<.05$. CESD = Center for Epidemiological Studies-Depression score. Common EF = Common EF latent factor. Updating-Sp. = Updating-Specific latent factor. Shifting-Sp. = Shifting specific latent factor.

If the relations between this model of EF and depression reflect variance shared with sleep, then I expect individual differences in sleep duration to be related to Common

EF abilities, but not Shifting-Specific abilities. Given this recently completed study with EF and depression, it is possible that I will observe a relationship between sleep duration and Updating-Specific abilities as well. However, if sleep duration is only related to depression and not EF, then I might not observe these patterns.

EF and Sleep. Sleep characteristics, such as duration and quality, are also associated with level of EF (Dahl, 1996; Anderson et al., 2009; Virta et al., 2013). For example, children who get less sleep tend to perform more poorly in school (Dewald et al., 2010). While school performance is not a pure measure of EF, EF is necessary for and correlated with school performance (Best et al., 2011), and the integration and variety of EFs necessary for school performance is consistent with Common EF abilities. However, as school performance is correlated with IQ, and IQ is related to both Common EF and Updating-Specific abilities, then school performance should also be related to an Updating-Specific latent factor. Therefore, I hypothesize that sleep duration will most likely be related to Common EF and possibly Updating-Specific abilities.

However, at a slightly later age, approximately age 14, sleepiness, but not sleep duration, was related to cognitive function as assessed by portions of the Behavior Rating Inventory of Executive Function and Delis-Kaplan Executive Functioning System (Anderson et al., 2009). The latent variable model of EF will allow me to get a purer measure of Common EF ability than is possible with any one measure, and detect a relationship between sleep duration and EF abilities, if one exists in later adolescence. Therefore, I will use the bifactor latent variable model of EF, to better understand the relationship between individual differences in sleep duration and EF throughout early adolescence into young adulthood.

The relationship between EF and sleep duration may be influenced by age. A meta-analysis by Dewald and colleagues (2010) found a relationship between sleepiness, sleep duration, and sleep quality and school performance in children aged 8-18. These effects were stronger for younger adolescents than older adolescents, which suggests that the relationship between sleep duration and EF might change with age. Another study, which used a subset of our sample, found that stable variance in sleep problems, as assessed by the Child-Behavioral Checklist, at ages 4,5,7, and 9-16 did not predict EF at age 17. However, children who showed improvement in sleep problems over time showed better Common EF in adolescence compared to those who continued to have sleep problems (Friedman et al., 2009). This indicates that changes in sleep, could be related to EF. Taken together, the literature is unclear on whether the relationship depends on age. I will start to address the idea that sleep duration differentially influences cognition depending on age by using our longitudinal data to see if relations between sleep and EFs differ across ages, and if changes in sleep across time influence EF abilities, or if changes in EF abilities influence sleep duration.

It is possible that those with low EF are more likely to experience depression or atypical sleep. It is also plausible that depression influences typical sleep durations and cognitive abilities. For example, Naismith and colleagues (2008) found that insomnia (early or late) in participants with MDD was associated differentially with depression and cognition. Early insomnia (trouble falling asleep) related to poorer global cognition and depression severity, whereas late insomnia (trouble falling back asleep in the early morning) related to poorer verbal fluency and memory, later depression onset, and symptom severity. The authors suggest that sleep disturbances might be a modifiable risk

factor for cognitive decline in people with depression. Given the limited research examining sleep, EF, and depression in the same model, I propose to run a series of path models to better understand the relationship between these three variables over time.

Current Study

The available data are part of the Longitudinal Twin Study (LTS). The depression measures include symptom scores on the Center for Epidemiologic Studies – Depression (CES-D) scale, and lifetime diagnosis from the Diagnostic Interview Schedule (DIS; for participants 18 and older) or Diagnostic Interview Schedule for Children (DISC; for participants under 18). These measures were collected at three different time points (see Table 4.1 for information on which measures were collected at each age).

Table 4.1 Available Measures at each Time Point

	Early Adolescence: age 12	**Late Adolescence: age 17**	**Henry Ford: age 21**	**Early Adulthood: age 23**
EF	-	9 EF tasks	-	9 EF tasks
Depression	CESD DISC	CESD DISC / DIS	CESD	CESD DIS
Sleep	Weekday Weekend	Weekday Weekend Typical Last night	Weekday Weekend Insomnia*	Weekday Weekend

Note. CESD = Center of Epidemiological Studies –Depression scale scores; DISC = Diagnostic Interview Schedule for Children (administered to participants under 18); DIS Diagnostic Interview Schedule (administered to participants 18 and older); DIS and DISC represent lifetime diagnosis of Major Depressive Disorder at a given wave of data collection.

I examined both CES-D symptom scores and lifetime diagnoses, because the LTS is a population-based sample, which has relatively low rates of clinical endorsement, particularly at the younger time points. However, I expect subclinical levels of depression to be associated with sleep duration and EF. For EFs the full battery of nine tasks was

available at two time points: late adolescence (approximately 17), and early adulthood (approximately 23).

Multiple waves of sleep data were also collected through 3 different sources: a health questionnaire that was part of LTS, the Jessor survey (Jessor & Jessor, 1977) which was part of the Center for Antisocial Drug Dependence (CADD), and a separate online study focusing on insomnia that was partially funded by the Sleep Research Center at the Henry Ford Health System. For the remainder of the paper, these datasets are referred to as the Health survey, the Jessor survey, and the Henry Ford survey, respectively. Refer to Table 4.1 for when each survey was administered during development.

At the level of the indicators, sleep duration and depression tend to negatively correlate (see Table 4.2 for correlation matrix). A more robust pattern emerges with depression symptoms compared to lifetime diagnosis, likely due to low rates of clinical depression at this age in a population sample.

Table 4.2 Phenotypic Correlations between Depression and Sleep Duration

	CES-D 12	CES-D 17	CES-D 23	MDD 12 LT DX	MDD 17 LT DX	MDD 23 LT DX	health typ 17	health last 17	jessor day 12	jessor end 12	jessor day 17	jessor end 17	jessor wk 22	jessor end 22	HF day 21	HF end 21	HF typ day 21
CES-D 12																	
CES-D 17	**.33**																
CES-D 23	**.18**	**.41**															
MDD 12 LT DX	**.46**	.16	**.25**														
MDD 17 LT DX	**.18**	**.39**	**.20**	**.51**													
MDD 23 LT DX	**.12**	**.29**	**.40**	.23	**.42**												
health typ 17	.01	**-.12**	-.07	.04	**-.11**	-.06											
health last 17	-.03	**-.14**	**-.12**	.11	-.08	-.09	**.42**										
Jessor day 12	**-.11**	**-.11**	**-.09**	.07	**-.11**	**-.11**	**.31**	**.17**									
Jessor end 12	**-.12**	**-.08**	**-.10**	.21	-.07	**-.14**	**.13**	**.11**	**.45**								
Jessor day 17	-.02	**-.10**	-.06	.06	**-.14**	-.04	**.75**	**.29**	**.25**	**.12**							
Jessor end 17	**-.08**	**-.11**	-.03	.08	-.10	.04	**.32**	**.28**	**.11**	**.11**	**.32**						
Jessor day 23	.01	-.07	-.07	**.24**	-.03	**-.15**	**.24**	**.12**	**.16**	**.13**	**.23**	**.13**					
Jessor end 23	.01	**-.08**	-.06	**.50**	.02	**-.12**	**.12**	**.14**	.06	**.18**	**.13**	**.29**	**.42**				
HF day 21	.02	.06	.04	.04	-.02	-.06	**.16**	.07	**.12**	.02	**.22**	**.10**	**.32**	**.12**			
HF end 21	-.02	.01	-.02	.12	.10	-.04	**.15**	.08	**.13**	**.10**	**.19**	**.16**	**.28**	**.29**	**.42**		
HF typ day 21	-.02	**-.12**	**-.08**	.06	**-.13**	**-.16**	**.23**	**.13**	**.16**	.06	**.27**	.08	**.43**	**.20**	**.47**	**.26**	
HF typ end 21	-.07	**-.12**	**-.11**	**.25**	.00	**-.12**	**.23**	**.13**	**.13**	**.13**	**.21**	**.17**	**.28**	**.35**	**.21**	**.62**	**.44**

Note. Bold = $p<.05$. All continuous variables were age and sex regressed before correlations were run. CES-D = Center for Epidemiological Studies-Depression score; MDD LT DX = Major Depressive Disorder Lifetime Diagnosis; day = weekday; end = weekend; Typ = sleep duration on a typical night; Last = sleep duration last night; Health indicates that the sleep duration came from the health survey during the LTS study; HF indicates that the sleep duration came from the Henry Ford Sample; Jessor indicates sleep duration came from the Jessor questionnaire.

Given the benefits of structural equation modeling, I examined whether there is evidence for a latent variable of sleep duration. If there is evidence for a latent factor of sleep, it might be for sleep in general, including both weekday and weekend sleep duration. Most of the sleep durations positively correlate with each other; however, stronger relationships emerge within a given age for weeknight and weekend night sleep, rather than across ages for separate weekday-specific and weekend-specific latent factors. Combining the age 21 and 23 variables, I estimated separate weekday-specific and weekend-specific latent variables, to help explore potential differential contributions of sleep during the week, and less restricted sleep during the weekends. Figure 4.3 depicts the latent variable models of sleep duration estimated with these data. However, I did not see substantial increases in association with these latent variables, so they were dropped from further analyses.

Figure 4.3.
Latent variable models for sleep with standardized loadings. Panel A. Latent variable model of age 17 sleep. Last sleep = how much a person slept the previous night, Typ Sleep = how much the

person typically slept. Panel B. Latent variable for Weekday sleep during early adulthood (approximately ages 21-23). Panel C. Latent variable for Weekend sleep during early adulthood (approximately ages 21-23). Health indicates that the sleep duration came from the health survey during the LTS study; HF indicates that the sleep duration came from the Henry Ford Sample; Jessor indicates sleep duration came from the Jessor questionnaire.

Method

Sample

Our participants were from the Longitudinal Twin Sample (LTS) at the University of Colorado Boulder. The sample was comprised of 857 twins (54% MZ (463)), from 402 families. The sample was 51% female (437). The LTS twins are same-sex twin pairs born between 1986 and 1990. The families were initially recruited through the Colorado Department of Health. The initial sample was 86.6% Caucasian, 8.5% Hispanic, 0.7% African- American, 1.2% Asian, and 2.9% other, which approximates the ethnic and racial composition of Boulder County, Colorado as whole during the 1990's according to the United States Census (Rhea, Gross, Haberstick, & Corley, 2013). Data collection is currently ongoing.

Materials

Depression The following depression measures were collected at multiple waves, with mean ages 12.43(0.37), 17.26(0.64), 22.81(1.28).

Lifetime Diagnosis. The Diagnostic Interview Schedule (DIS) is a structured, diagnostic interview, was used for participants over 18, and the child version, Diagnostic Interview Schedule for Children (DISC) was used for individuals under 18. Then DSM-IV criteria were applied to the responses, to come up with a dichotomous, case / control diagnosis for MDD.

Depression Symptoms. Depression symptoms were assessed using the 20-question

Center for Epidemiologic Studies – Depression (CES-D) scale (Radloff, 1977). The scale assesses how often a person experiences depressive symptoms over the past week on a scale of 0 (rarely or none of the time) to 3 (most or all of the time). In order to calculate depression symptoms, CES-D items were reverse coded where appropriate. Then when at least 16 out 20 questions were answered, summing the raw scores and multiplying that sum by 20 resulted in a total score. Last, I took the square root of that score to achieve a better distribution. Please see Table 4.3 for descriptive statistic information and raw values for CES-D and lifetime diagnosis.

Table 4.3 Descriptive Statistics for Executive Function (EF) and Depression Measures

	Mean	**Case/ Control**	**SD**	**Min**	**Max**	**N**	**Age (SD)**
EF age 17	na	na	na	na	na	786	17.26 (0.64)
EF age 23	na	na	na	na	na	749	22.84 (1.29)
CESD age 12	9.68	na	7.68	0	48	716	12.43 (0.37)
CESD age 17	9.45	na	7.50	0	47	795	17.26 (0.64)
CESD age 21	11.91	na	9.23	0	50	761	21.06 (2.02)
CESD age 23	11.06	na	8.94	0	46	752	22.81 (1.28)
DISC age 12	na	10/709	na	na	na	719	12.43 (0.37)
DISC/DIS age 17	na	59/738	na	na	na	797	17.26 (0.64)
DIS age 23	na	99/664	na	na	na	763	22.81 (1.28)

Note. CESD = Center of Epidemiological Studies Depression Scale scores; DISC = Diagnostic Interview Schedule for Children (administered to participants under 18); DIS Diagnostic Interview Schedule (administered to participants 18 and older); na = not applicable to this category. DIS and DISC represent lifetime diagnosis of Major Depressive Disorder at a given wave of data collection. As EFs are estimated at the latent variable level for Common EF, Updating-Specific, and Shifting-Specific EF, there are no descriptive statistics at each wave.

Executive Functions EF information was collected at two different waves with mean ages (17.26, 22.84). Participants completed 9 EF tasks, 3 of which are designed to tap response inhibition, 3 to tap updating working memory, and 3 that tap set-shifting between sub-tasks. Please see Friedman et al. (2008; 2016) for a comprehensive description of the tasks included in our EF battery. All sleep, EF variables, and CES-D scores were regressed on sex and age at time of appropriate assessment.

Sleep

Health. In the Health questionnaires sleep duration was assessed with the questions, "How many hours do you typically sleep at night?" and "How many hours of sleep did you get last night?" The participants then wrote in the number of hours for each question. See Table 4.4 for descriptive statistics for all sleep measures.

Table 4.4 Descriptive Statistics for Sleep Duration

	Mean	SD	Min	Max	N	Age (SD)
Health						
Typical: age 17	7.49	1.15	3	13	715	17.26 (0.64)
Last night: age 17	7.17	1.62	1	13	714	17.26 (0.64)
Jessor						
Weekday: age 12	8	31.30%	5	11	820	13.12 (1.82)
Weekend: age 12	10	23.30%	5	11	817	13.12 (1.82)
Weekday: age 17	8	36.80%	5	11	682	17.20 (0.57)
Weekend: age 17	9	25.80%	5	11	681	17.20 (0.57)
Weekday: age 23	7	36.70%	5	11	755	22.28 (1.28)
Weekend: age 23	8	33.90%	5	11	755	22.28 (1.28)
Henry Ford: age 21						
Weekday 1	7.14	1.22	3.5	11	703	21.07 (2.02)
Weekend 1	7.84	1.47	3	12.5	712	21.07 (2.02)
Weekday 2	7.97	1.55	3	12	710	21.07 (2.02)
Weekend 2	8.52	1.47	4	13	714	21.07 (2.02)
Insomnia*	.24 (140/587)	.	.	.	727	21.07 (2.02)

Note. Raw scores before age and sex are regressed out. Health refers to a health survey that was administered at the same time as EF measurements in the LTS study. Jessor refers to a survey that was administered to assess basic demographic, health, sleep, and other information as part of the CADD studies. Sleep variables in the Jessor study were truncated (5 hours of sleep or less, 11 hours of sleep or more). Henry Ford refers to a sleep study that was conducted in conjunction with the Henry Ford Health System. Typical sleep refers to amount typically slept in the past month at different waves of the EF study. Last night sleep refers to how much the participant slept the night preceding the EF test. Weekday 1 and Weekend 1 refers to sleep that was calculated from the time a participant woke up and went to sleep during weekday and weekends respectively. Weekday 2 and Weekend2 refers to the average number of hours slept reported by the participant. Chronotype = Continuous Chronotype score. Insomnia = case/control status for insomnia diagnosis. *This sleep measure was available, but not focused on in Studies.

Jessor. In the Jessor surveys participants were asked, "About how many hours of sleep do you usually get each week night?" followed by "How about on weekend

nights?" and they selected from the options: 5 or less, 6, 7, 8, 9, 10, 11 or more, or, would rather not answer.

Henry Ford. The sleep survey in the Henry Ford data set assessed sleep duration 2 ways. They asked, "During the past month, at what time did you TYPICALLY get up ON WEEKDAYS? (Please be sure to write 'AM' or 'PM' in the last box)" and "During the past month, at what time did you TYPICALLY go to bed ON WEEKDAYS? (Please be sure to write 'AM' or 'PM' in the last box)". The participant wrote in the hour, minute, and AM or PM, each one in a separate box for each question. Sleep duration could usually be calculated from this information, if it was filled out correctly (answered all parts, did not indicate that they slept on average 23 hours per day, etc.). Participants were also asked, "During the past month, thinking about your average WEEKDAY, how long did you ACTUALLY sleep, EACH night (or your longest sleep period if you work a night shift or rotating shift)?**"** to which they wrote in the hours and the minutes they slept into separate boxes. Weekend sleep durations were assessed in the same manners. Please see Table 4.4 for descriptive statistics.

During the Henry Ford study, insomnia and sleep related anxiety were also measured. For the insomnia measure, participants indicated how often in the past month they had: difficulty falling asleep, difficulty staying asleep, and having non-refreshing sleep. The response options were "never", "sometimes", or "always". If a participant answered something other than "never" to at least one of the questions, he/she was asked a series of follow-up questions about duration of problems and extent of interference during the daytime. Participants met DSM-IV-TR criteria for insomnia if they a

minimum of one problem "usually" or "always" for at least a month, with at least "somewhat" interference (4th ed.; *DSM-IV*; American Psychiatric Association, 2000). The anxiety score was measured using the Ford Insomnia Response to Stress Test (FIRST). The FIRST is a 9-item scale that assessed if a person had difficulty sleeping under circumstances such as, "Before an important meeting the next day" and "After an argument". The response options ranged from 1 to 4 point scale (1 = not likely; 2 = somewhat likely; 3 = moderately likely; 4 = very likely). This scale was then summed for a total score (Drake et al., 2004; Drake, Friedman, Wright & Roth, 2011)

Categorical Sleep Variable. A categorical sleep variable was created to account for non-linear relationships between sleep duration and depression and EF. Typically nonlinear, quadratic trends in the data are captured by squaring the linear term, which results in larger values for both the positive and negative extreme values in a dependent variable, compared to both positive and negative values closer to the mean. Then an association can be captured between the dependent variable (e.g. depression) and both positive and negative extremes for the independent variable (e.g. sleep duration). However, the cross-lag models use sleep duration both as a dependent and independent variable, depending on the age, and when it is a dependent variable, may violate assumptions of normality. In order to avoid violating this assumption, I created a dichotomous sleep variable. Sleep was classified as either typical sleep (6.5 to 9.4 hours of sleep for weekdays and 6.5 to 10.4 hours of sleep for weekends) or atypical sleep (less than 6.5 or greater than 9.5 hours for weekday sleep and less than 6.5 or greater than 10.5 for weekday sleep). These thresholds were based on visual inspection of within-wave sleep duration and depression symptoms scatterplots and are consistent with previous

literature (short (<7), medium (7-9), and long (≥9); Watson et al., 2010; Watson, et al., 2012). Overall, the categorical sleep variable and the quadratic sleep term resulted in similar associations with depression symptoms. Descriptive information for these variables is available in Table 4.5.

Table 4.5. Counts for Categorical Sleep Variable

Atypical (high/low) /Typical N's	**Age 13**	**Age 17**	**Age 21**	**Age 23**
Health - 17				
typical sleep	-	161 (127/34) /554	-	-
last night sleep	-	240 (205/35) /474	-	-
Jessor				
weekday 12	229 (46/183) /591	-	-	-
weekend 12	193 (91/102) /624	-	-	-
weekday 17	-	147 (123/24) /535	-	-
weekend 17	-	160 (121/39) /521	-	-
weekday 23	-	-	-	221(207/14) /534
weekend 23	-	-	-	163(150/13) /592
Henry Ford - 21				
weekday calculated	-	-	177 (83/94) /533	-
weekend calculated	-	-	118 (52/66) /595	-
weekday typical	-	-	190 (171/19) /513	-
weekend typical	-	-	140 (109/31) /570	-

Note. Atypical sleep / Typical Sleep counts. Weekday Typical Sleep = 6.5 to 9.4 hours. Weekend Typical Sleep = 6.5 to 10.4 hours. Weekday Atypical sleep = < 6.5 or > 9.5 hours. Weekend Atypical Sleep = < 6.5 or >10.5.

Weekend – Weekday Sleep Difference. In addition to assessing sleep duration's relationship with EFs and depression, I also wanted to know if consistency in sleep

predicted differences in EFs and depression depending on the amount of sleep. For example, some people might have short weekday sleep durations and then catch up on sleep on the weekend, or they just have consistently shorter sleep durations. These two situations could predict differences in depression and EFs. For example, the person who catches up on sleep might perform better on EF tasks because he or she makes up for lost sleep on weekends compared to the person who does not. Or the person who has a more consistent short sleep duration might have a lower sleep drive, and therefore need less sleep and have better EFs. So I created a sleep difference variable and an average sleep variable in order to test these possibilities. The sleep difference variable assumed that most individuals sleep more on weekends than weekdays, and so I subtracted weekday sleep duration from weekend sleep duration. I also calculated the average amount of sleep by weighting weekday sleep by 5 (for 5 weekdays), and weekend sleep by 2 (for 2 weekend days), and then dividing it by 7 (all the days in the week). Negative difference scores indicate that the person reported sleeping more during the week than on weekends, while positive scores indicate longer weekend sleep durations.

Analyses

Analyses were completed in Mplus 7.4 (Muthen & Muthen, 2012). Models used all available phenotypic data; non-independence (due to the inclusion of co-twins) was corrected with the type= COMPLEX option, which clusters on family. For models with only continuous data, robust maximum likelihood (MLR) was used, and all variables were residualized on within-wave age and sex. For models with ordinal diagnoses, mean and variance adjusted weighted least squares (WLSMV) estimation (delta parameterization) was used, and within-wave age and sex was included as covariates.

While sex effects are not typically observed with EF measures, they can be for depression and sleep. Therefore, throughout the analyses I looked for evidence of sex effects and followed up when necessary. To assess model fit I primarily used comparative fit index (CFI) and root-mean-square error of approximation (RMSEA), with the criteria of CFI > 0.95 and RMSEA < 0.06 as indicators of good fit, since Chi-square is sensitive to sample size (Hu & Bentler, 1998). Fit indices are presented in Figure notes or summarized in Table notes.

Results

Aim 1 – Are individual differences in sleep duration associated with depression?

Given previous research, I tested whether depression has a linear or quadratic relationship with sleep at our adolescent and young adult time, using regression. To test for nonlinear relationships I ran regressions with both a linear term and the square of the linear term to test for quadratic effects. Results from these regressions are presented below in Table 4.6.

Table 4.6 Beta Estimates between Depression and Sleep Duration

Linear Model	Sleep Variable	**CESD - 12**	**CESD - 17**	**CESD - 21**	**CESD - 23**	**MDD - 12**	**MDD - 17**	**MDD - 23**
Weekday - 12	linear	**-0.09**				0.06		
	quadratic	**0.12**				0.06		
Weekend - 12	linear	**-0.09**				0.21		
	quadratic	0.07				-0.02		
Weekday - 17	linear		**-0.12**				**-0.14**	
	quadratic		**0.14**				0.08	
Weekend - 17	linear		-0.08				-0.10	
	quadratic		0.07				0.02	
Typically sleep - 17	linear		**-0.13**				**-0.09**	
	quadratic		**0.09**				0.03	
Last night - 17	linear		**-0.12**				-0.03	
	quadratic		0.05				0.06	

Weekday - 21	linear			**-0.23**				**-0.13**
	quadratic			**0.19**				**0.08**
Weekend - 21	linear			**-0.17**				**-0.12**
	quadratic			**0.19**				0.00
Weekday - 23	linear				**-0.10**			**-0.17**
	quadratic				**0.17**			**0.09**
Weekend - 23	linear				-0.05			-0.11
	quadratic				0.08			0.02
insomnia	linear	0.10	0.22	0.39	0.26			

Note. Presented are the standardized beta estimates from regressions where linear and quadratic sleep duration predicts depression. Bold = $p<.05$. CESD = Center of Epidemiological Studies Depression Scale scores; MDD = Lifetime Diagnosis for Major Depressive Disorder at a given wave of data collection. Weekday = Weekday sleep duration. Weekend = Weekend sleep duration.

Generally I found linear relationships where shorter weekday and weekend sleep durations predicted more depression symptoms at ages 12 and 21. The fewer hours an individual reported that they typically slept and slept the night before testing in the health questionnaire also predicted more depression symptoms. Fewer hours of weekday sleep duration predicted more depression symptoms at age 17, but after controlling for quadratic sleep weekend sleep does not predicted depression. After controlling for nonlinear sleep, fewer hours of weekday sleep at age 23 predicted more depression symptoms.

Nonlinear weekday sleep at 12, 17, 21, and 23, and nonlinear weekend sleep at age 21 predicted more depression symptoms, while the quadratic effects for weekend sleep were marginally significant at all other ages.[5] I also saw an association between the amount 17 year olds reported they typically slept, but not how much they slept the night before testing, and depression. As seen in Figure 4.4, all of these relationships indicated

[5]At ages 21 and 23, I removed the CES-D symptom that related to sleep from the scores: all of the associations that were significant with the complete CES-D scores remained significant. CES-D scores with and without the sleep item were correlated at $r>0.99$.

that both fewer and more than a typical 7-9 hours of sleep and 7-10 hours on weekends, were associated with increased depression symptoms. The plots show that both short sleep duration and long sleep duration are associated with increased depression, however, at the level of depression symptoms, those who report very short sleep seem to have the highest depression symptoms.

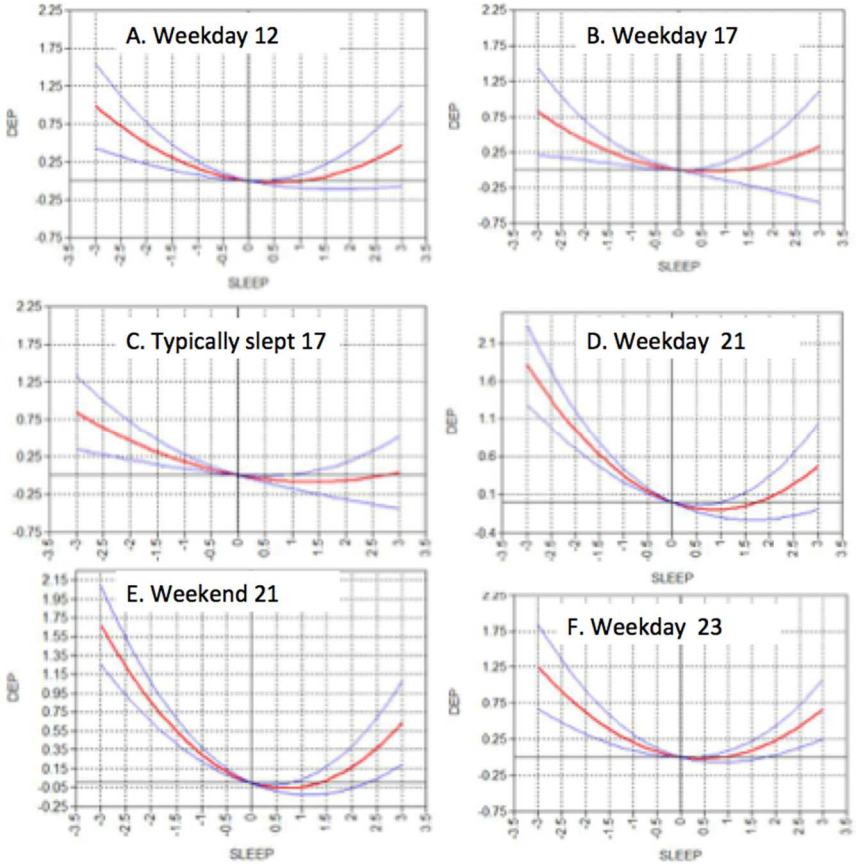

Figure 4.4 Plots of Quadratic Sleep on Depression Symptoms. Y-axis: Dep: standardized depression symptom scores as measured by the CES-D questionnaire. X-axis: Sleep: standardized quadratic function of sleep duration.

As previously mentioned, there is a strong relationship between insomnia and depression. To examine whether the relationship between sleep duration and depression was an artifact of the relationship between insomnia and depression, I included insomnia in the model when it was available at age 21. I found that even when controlling for insomnia, less sleep still predicted more depression symptoms for the measures that

previously had a significant relationship. Likewise, anxiety about future events, or rumination about past events while lying in bed could have a direct influence on sleep duration. So I ran models including a Ford Insomnia Response to Stress Test (FIRST) score, which assesses a person's ability to fall asleep after or before stressful situations (Drake et al., 2004). Again, all relationships that were previously significant at age 21 remained after including this covariate.

MDD

In addition to depression symptom scores, depression diagnosis was also available at ages 12, 17, and 23 for these subjects; therefore I tested whether linear and quadratic sleep predicts the stricter criteria of diagnosis. Overall I saw the same pattern as with depression symptoms: less sleep predicted diagnosis linearly and atypical sleep was associated with more depression. Only 10 participants at age 12 met the criteria for depression; neither linear nor quadratic sleep predicted at this age, and so after these regressions, age 12 MDD diagnosis was dropped from further analyses. Less weekday sleep, controlling for quadratic sleep, predicted increased depression at ages 17 and 23. Both fewer hours and more hours of sleep than typical at age 23 also predicted depression. As there was no age 21 time point for depression diagnosis, the relationship between age 21 sleep and MDD diagnosis was not assessed. When controlling for quadratic sleep, reported typical sleep in the health survey at age 17 and weekend sleep at age 23 no longer significantly predict MDD. Overall the phenotypic, within-wave regressions and plots indicate that shorter sleep durations and atypical sleep were associated with increased depression symptoms.

Aim 2 – What are the longitudinal relationships between sleep duration and

depression?

Insomnia is a known predictor of depression; across a wide range of ages, individuals with insomnia are twice as likely to report depression than those without sleep problems during follow ups 1-34 years later (Baglioni et al., 2011). Therefore I wanted to test the longitudinal relationships between sleep duration and depression to see if similar patterns were observed. These longitudinal relationships were assessed within a cross-lag model framework (see Figure 4.5 for an example), with both a linear sleep term and a categorical sleep term to account for the non-linear trends in the data. A cross-lag model was used because it provides information about prospective relationships while controlling for earlier time points, and whether within-wave relations exist even after controlling for previous measures or are due to earlier measures. The categorical sleep term was a binary variable with one category for "typical" sleep (6.5-9.4 hours on weekday nights, and 6.5-10.4 hours on weekend nights) and the other for "atypical" sleep (<6.5 or >9.4 hours on weekday nights and <6.5 or >10.4 on weekend nights). Refer to methods section above for more information on the categorical sleep variable.

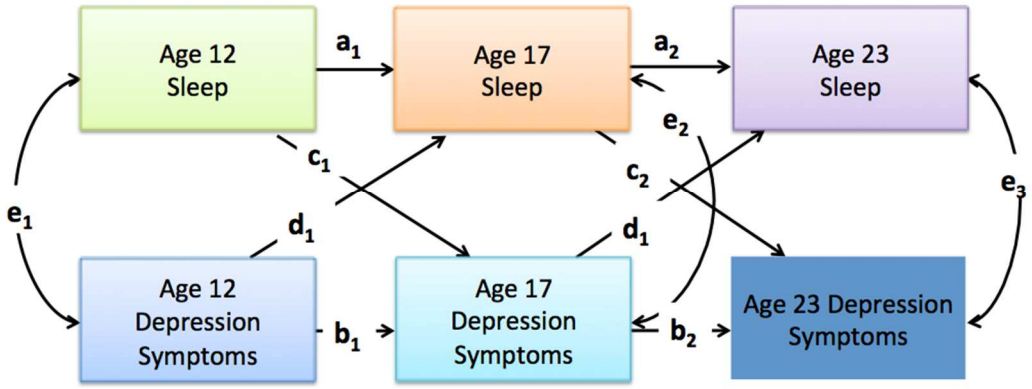

Figure 4.5 Example of a longitudinal cross-lag model between sleep and depression over 3 waves of data. a_1 and a_2 represent the effects of sleep at time t on sleep at time t+1; b_1 and b_2 represent the effects of depression at time t on depression at time t+1; c_1 and c_2 represent the effects of sleep at time t on depression at time t+1, controlling for depression at time t; d_1 and d_2 represent the effects of depression at time t on sleep at time t+1 controlling for sleep at time t; e_1 represents

the correlation between sleep and depression at time t; e_2, and e_3 represent residual correlations between sleep and depression within a given time point; residuals not depicted.

The first step in creating the cross-lag models was to establish a model for the categorical and linear sleep variables. I modeled all of the autoregressive paths, sleep from time t to sleep at time t+1, for both the linear and categorical variables respectively. I also modeled all of the within-wave effects with correlations between linear and categorical sleep at age 12, and then residual correlations at all other waves. Then I used fit indices with the strongest effects to add in either cross-lag paths between linear sleep and categorical sleep, or regressions of sleep at time t on sleep at time t+2 (or more), until adequate fit was achieved. The final models are presented in Figure 4.6 for both weekday (Panel A) and weekend sleep (Panel B).

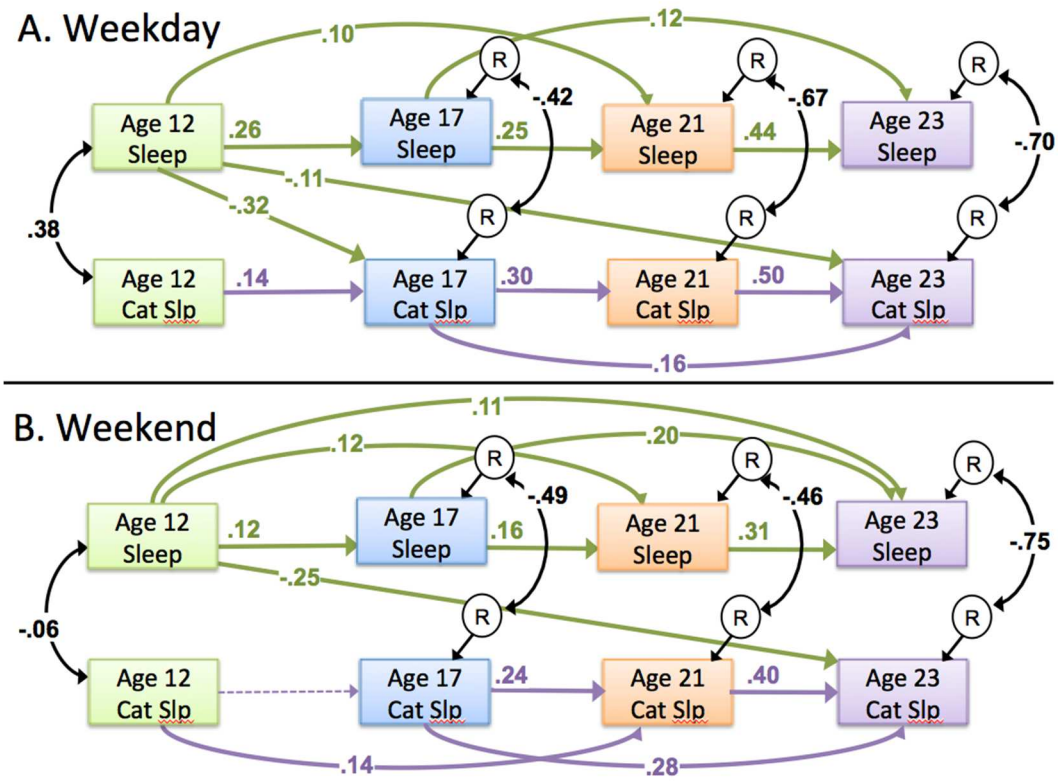

Figure 4.6 Cross-lag model of the effects of linear sleep on categorical sleep. Panel A depicts weekday sleep; CFI=.992, RMSEA=.030. Panel B depicts weekend sleep; CFI=.980,

RMSEA=.036. Solid lines = $p<.05$. Black bolded lines = Correlations within age with $p<.05$. Cat Slp = Categorical Sleep (Weekday: Atypical (<6.5 or >9.5)/Typical (6.5-9.5); Weekend: Atypical (<6.5 or >10.5)/Typical (6.5-10.5)). Longitudinal cross-lag paths more than one time point removed were modeled, but are not pictured here if they were not significant.

There were within-wave relationships between linear and categorical sleep at all ages, for both weekday and weekend sleep. At age 12, linear weekday sleep correlated positively with categorical sleep, meaning that those who slept more on weekdays at age 12 were more likely to have atypical weekday sleep. All other ages saw the reverse pattern where shorter sleep duration was associated with atypical sleep (sleeping more or less than typical for weekday or weekend sleep). This association appears to get stronger as people age. Both weekday and weekend sleep saw a longitudinal association where fewer hours sleep at age 12 predicted atypical sleep at age 23. The weekday model also saw that less age 12 sleep predicted atypical sleep at age 17. Overall, sleep duration is associated with whether or not someone sleeps more or less than average within a given time point, but sleep duration in early adolescence seems to also predict longitudinal sleep patterns.

In addition to these cross-lag paths, earlier sleep duration seems to predict later duration, and earlier sleep category predicts later sleep category, over and above the autoregressive paths one time point removed. Both weekday and weekend sleep models indicated positive relationships between age 12 sleep duration and age 21 sleep duration, and age 17 sleep duration and age 23 sleep duration. Similarly, they both indicated age 17 categorical sleep predict age 23 categorical sleep, where atypical sleep at a younger age was associated with atypical sleep at a later age, over and above what was already predicted through the autoregressive paths. Weekend sleep showed additional effects of early sleep on later sleep where age 12 sleep duration positively predicted age 23 sleep

duration, and age 12 categorical sleep positively predicated age 21 categorical sleep. This could indicate that age 12 sleep patterns are particularly influential on later sleep patterns.

When adding in the other behavioral variables (e.g. depression or EF) to our cross-lag sleep models I allowed for longitudinal cross-lag paths, where earlier sleep duration predicted all subsequent measures of depression symptoms (and vice versa). If cross-paths were more than 1 time point removed from the predictor and not significant, then those lines were excluded from the figures for ease of visual interpretation. Also, as the sleep models change very little with addition of the depression or EF variables, I will focus on the relationship between sleep and depression and sleep and EF when describing cross-lag models.

The cross-lag model between weekday sleep and depression symptoms showed that age 12 sleep predicts age 17 depression scores (see Figure 4.7). Both the linear and categorical measures predicted CESD, such that less sleep and atypical sleep predicted higher endorsement of depression symptoms. Again, shorter sleep duration at age 12 predicted more depression at age 21, after controlling for age 12 and age 17 categorical sleep, age 17 linear sleep and age 12 depression. Similarly, increased depression at age 17 predicted less sleep at age 21, controlling for age 12 and 17 sleep duration. In contrast, those with atypical sleep at age 17 are less depressed at ages 21 and at age 23, after holding constant previous depression and sleep. Similarly, atypical sleep at age 21 predicts less depression at age 23, holding constant age 21 depression and previous waves of sleep. Interestingly, age 17 sleep duration and age 21 sleep duration have opposing effects on age 23 depression symptoms, where more sleep at 17 predicts more depression at age 23, but less sleep at age 21 predicts more depressive symptoms.

All autoregressive paths for depression were significant. Consistent with simple regressions, there were significant, negative correlations between within-wave linear sleep and depression at ages 12, 17, and 21, where fewer hours of sleep were associated with more depression symptoms. There were positive correlations between categorical sleep and depression at ages 17, 21, and 23, where atypical sleep was associated with increased depression symptoms. Inconsistent with the simple regressions, within the weekday cross-lag model age 12 categorical sleep and age 23 sleep duration did not show within-wave correlations with depression.

To summarize, shorter sleep durations at ages 12 and 21 predict more depression at ages 17, 21, and 23, but longer sleep durations at age 17 predicts more depression symptoms at age 23, controlling for previous waves of sleep, and accounting for earlier depression. Also atypical sleep at ages 12 predicts more depression symptoms at ages 17, but atypical sleep at 17 and 21 predicts less depression at ages 21 and 23, holding previous depression and sleep constant. Some of these changes in direction of association could be due to changing circumstance with increased flexibility in college and post-university life.

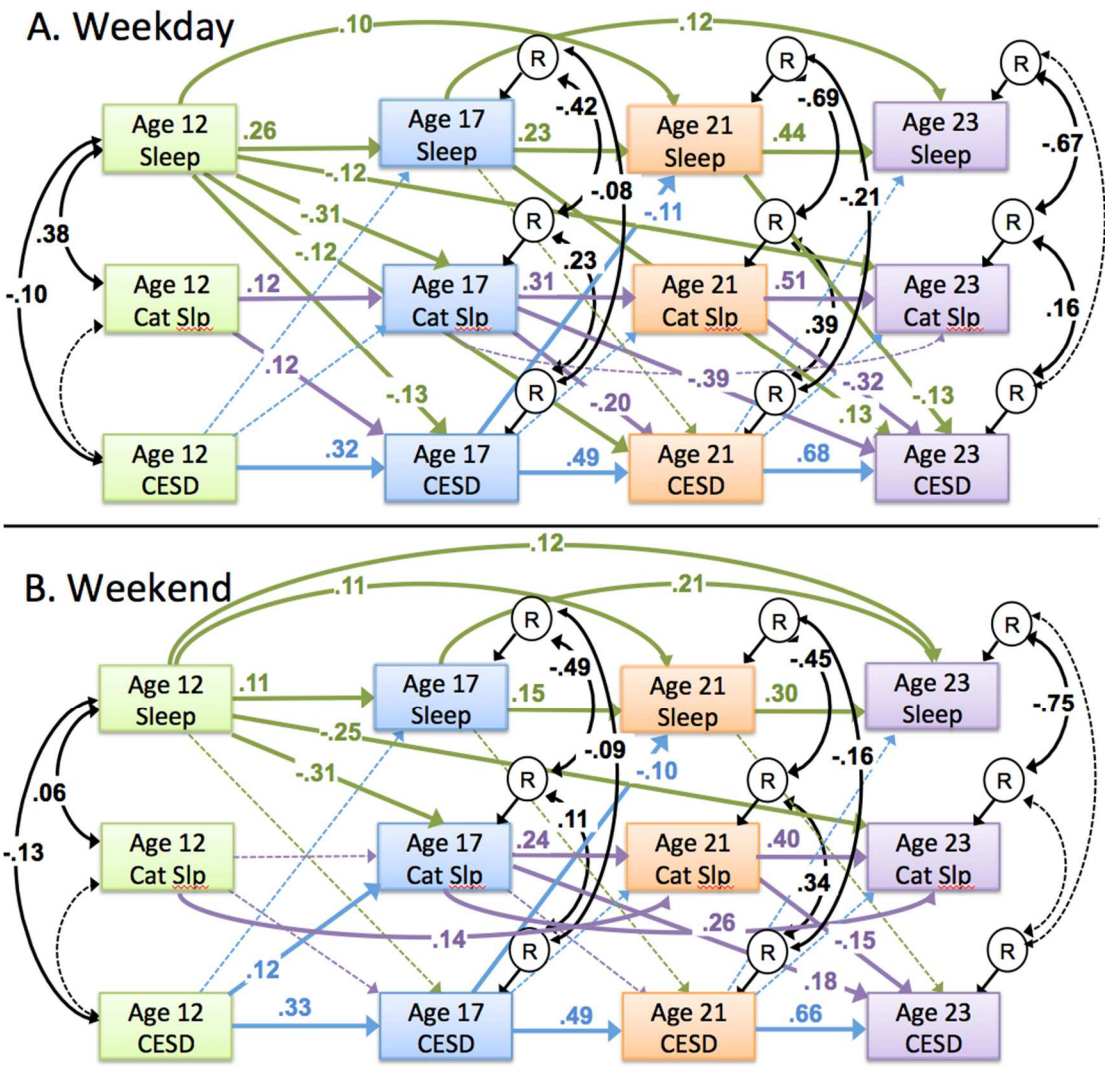

Figure 4.7. Cross-lag model of the effects of depression symptoms on weekday sleep. Panel A depicts weekday sleep; CFI=.988, RMSEA=.037. Panel B depicts weekend sleep; CFI=.972; RMSEA=.048. Solid lines = $p<.05$. Black bolded lines = Correlations within age with $p<.05$. All dashed lines $p>.05$. CESD = Center for Epidemiological Studies-Depression score. Sleep = linear sleep duration. Cat Slp = Categorical Sleep (Weekday: Atypical (<6.5 or >9.5)/Typical (6.5-9.5); Weekend: Atypical (<6.5 or >10.5)/Typical (6.5-10.5)). Longitudinal cross-lag paths more than one time point removed were modeled, but are not pictured here if they were not significant.

The weekend cross-lag model with depression symptoms showed a different pattern from weekday sleep. Age 12 sleep did not predict depression at all, but age 12 depression did predict atypical sleep at age 17. Again, earlier depression had an effect on later sleep, where more depression symptoms at age 17 predicted shorter sleep duration at 21, holding previous sleep duration constant. Inconsistent with weekday sleep, atypical

weekend sleep at age 17 predicted more depression symptoms at age 2; but consistent with weekday sleep, atypical sleep at age 21 predicted fewer depression symptoms, holding age 21 depression and sleep duration constant.

All autoregressive paths but one were significant; sleep category for weekend sleep at age 12 did not significantly predict age 17 sleep category status. Within-wave correlations were very similar to weekday sleep, with one exception. Shorter sleep durations were associated with increased depression symptoms at ages 12, 17, and 21, while atypical sleep was associated with more depression at ages 17 and 21, but not 23.

Longitudinal relationships between sleep and depression diagnosis were also assessed with cross-lag models, however depression diagnosis was not available at age 21. Also as previously mentioned, due to low rates of depression diagnosis at age 12, MDD was left out of the model at that time point. See Figure 4.8, Panel A for the final weekday model.

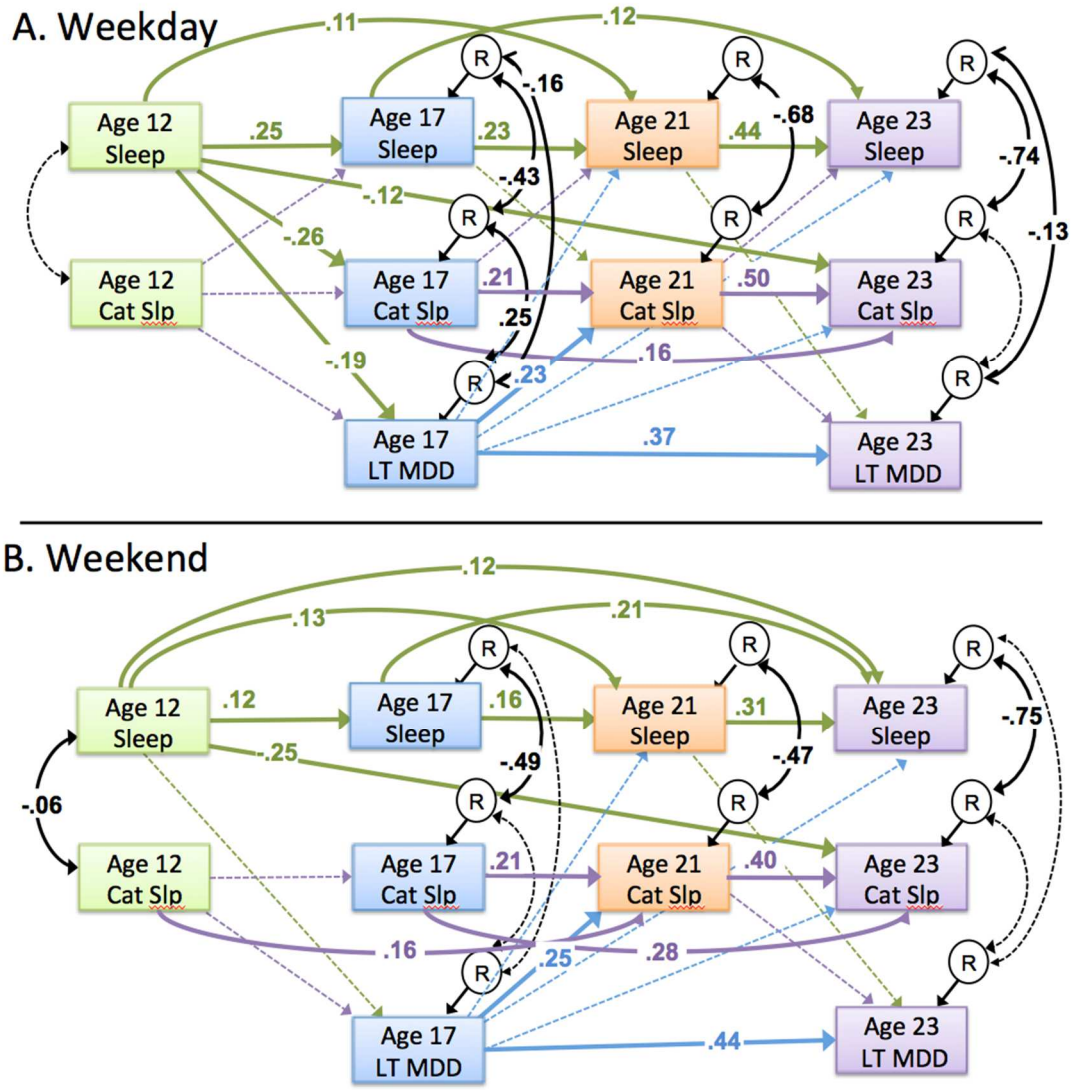

Figure 4.8. Cross-lag model of the effects of depression diagnosis on sleep.
Panel A depicts weekday sleep; CFI=1.000, RMSEA=0.000. Panel B depicts weekend sleep; CFI=.985; RMSEA=.033. Solid lines = p<.05. Black bolded lines = Correlations within age with p<.05. All dashed lines p>.05. Green lines = linear sleep was the predictor variable. Purple lines = categorical sleep was the predictor variable. Blue lines = depression symptoms were the predictor variable. Sleep = linear sleep duration. MDD = Lifetime Diagnosis of Major Depressive Disorder.

As expected, when I saw results, either less sleep, or atypical sleep was associated with depression diagnosis. Fewer hours of sleep at age 12 predicted depression at age 17, holding constant (a)typical sleep status. Age 17 MDD predicted too much or too little sleep at age 23, accounting for previous waves of both categorical and linear sleep. The autoregressive path between age 17 and age 23 depression was significant. Within each

of the age 17 and 23 waves, fewer hours of sleep were associated with depression. Age 17 atypical sleep was also associated with depression.

Overall a similar pattern emerged for weekend sleep (see Figure 4.7, Panel B), however fewer hours of weekend sleep at age 12 was only marginally associated with MDD at age 17. As seen in weekday sleep, depression diagnosis at age 17 predicted sleeping too much or too little at age 23. Weekend sleep showed some differences from the weekday model in the within-wave correlates. Linear and categorical weekend sleep at age 17 no longer associated with MDD diagnosis. Similarly, linear sleep at age 23 was no longer related to MDD.

With a few exceptions, fewer hours of sleep and atypical sleep (sleeping too much or too little) were related to both depression symptoms and diagnosis longitudinally. However, some anomalies from late adolescence to early adulthood (e.g. atypical weekday sleep at age 17 predicting fewer depression symptoms at age 21). These anomalies could reflect the drastic change in environment, from living with parents with a set schedule, to college with a more flexible schedule and less oversight or delayed circadian phase around the time of puberty, which often peaks in young adulthood (Hagenuer, Perryman, Lee & Carskadon, 2009). More often earlier sleep (particularly weekday sleep) was predictive of later depression; however depression at age 17 consistently predicted age 21 sleep patterns across models. Increased depression symptoms predicted fewer hours of sleep, and MDD diagnosis predicted atypical sleep. When earlier sleep predicted later depression, typically shorter sleep durations predicted more depression, but categorical sleep differentially predicted depression depending on age.

Aim 3 – Is there a relationship between sleep duration and EFs?

Given relationships previously observed between sleep deprivation, EFs, and sleep duration and cognitive measures such as academic performance, I predicted I would see shorter sleep durations and atypical sleep status associated with lower Common EF and Updating-specific abilities at the latent level, and potentially shorter sleep duration and atypical sleep status associated with better Shifting-specific abilities. However, as it is possible that students with lower EFs differ in sleep drive, are not involved in as many school activities, skip early classes, schedule later classes, or are unemployed, and therefore sleep more or more regularly than their higher EF counterparts. In order to test for overall linear trends and for the possibility of both long and short sleepers, similar analyses were conducted at the phenotypic level between sleep duration and our latent variable model of EF. Again, I included both a linear sleep variable and then the square of that variable to test for quadratic effects.

Updating-specific abilities, but not Common EF or Shifting-specific abilities were related to linear sleep duration, after controlling for nonlinear sleep. I found less sleep at age 12 predicted better Updating-specific abilities, but more sleep at age 21 (weekday and weekend respectively) was associated with better Updating-specific abilities at ages 17 and 23 (see Table 4.7).

Table 4.7 Linear and Quadratic Regression Estimates between EFs and Sleep Duration

Sleep Variable		Common EF	Updating	Shifting
		Age 17 EFs		
Age 12 weekday	linear	0.06	**-0.16**	-0.01
	quadratic	-0.06	-0.02	0.03
Age 12 weekend	linear	-0.00	0.05	-0.03
	quadratic	-0.10	-0.01	-0.02
Age 17 weekday	linear	0.02	-0.02	-0.02
	quadratic	-0.14	0.13	**0.17**
Age 17 weekend	linear	0.02	0.02	0.06
	quadratic	-0.12	0.05	**0.23**
Age 17 Typical	linear	-0.02	-0.02	-0.09
	quadratic	-0.09	0.13	**0.14**
Age 17 Last Night	linear	0.08	-0.07	0.01
	quadratic	0.00	-0.06	0.04
Age 21 weekday	linear	-0.03	**0.13**	0.07
	quadratic	-0.00	-0.02	0.08
Age 21 weekend	linear	-0.02	0.11	-0.02
	quadratic	-0.10	0.14	**0.18**
		Age 23 EFs		
Age 21 weekday	linear	-0.04	0.09	-0.04
	quadratic	0.02	-0.06	0.00
Age 21 weekend	linear	0.02	**0.13**	-0.07
	quadratic	-0.04	-0.01	0.06
Age 23 weekday	linear	-0.07	0.02	-0.07
	quadratic	**-0.13**	-0.09	0.05
Age 23 weekend	linear	0.04	0.02	-0.06
	quadratic	-0.06	-0.06	0.04

Note. Presented are the beta estimates from regressions where both linear and quadratic sleep duration predict latent EFs. Bold = $p<.05$.

When it comes to non-linear effects of sleep, I hypothesized that for different reasons, both those getting too little sleep and too much sleep might have disrupted EFs, and therefore tested the quadratic relationship with EF as well. Both Common EF and Shifting-specific factors showed a positive relationship with quadratic sleep duration. Too much or too little age 17 weekday, weekend, and typically reported sleep all predicted better Shifting-specific abilities at age 17. Additionally better age 17 Shifting-

specific abilities related to more extreme weekend sleep at age 21. Taken together this indicates a consistent relationship between both too little and too much sleep at age 17 and better Shifting-specific ability. However, age 23 weekday sleep negatively predicted age 23 Common EF, indicating that worse Common EF was associated with too much and too little sleep. The trade-off seen here between Common EF and Shifting-specific abilities has been previously reported (Miyake & Friedman, 2012) and might reflect competition between goal maintenance and flexibility to switch between goals (Herd et al., 2014).

The longitudinal relationships between EFs and sleep duration were also assessed, because chronic partial sleep deprivation might lead to long-term decrements in EFs; alternatively, better EFs might lead to a more regimented sleep scheduled, either imposed by parents or individuals themselves, and therefore better Common EF and Updating-specific abilities could be associated with typical sleep durations. Again, longitudinal relationships were assessed in a cross-lag model that included both the linear and a categorical sleep term. EFs were simultaneously estimated in weekday and then weekend models, as shown in Figure 4.8; however, the results are presented in figures with each EF latent factor by itself. Again, I removed non-significantcross-lag paths more than one time point removed (eg. age 12 sleep predicting age 23 EFs). Residual correlations for the Updating-specific factor at age 23 could not be estimated because the residual variance for that factor was fixed at zero. This was done because the Updating-specific factor at age 17 almost perfectly predicts the Updating-specific factor at age 21.

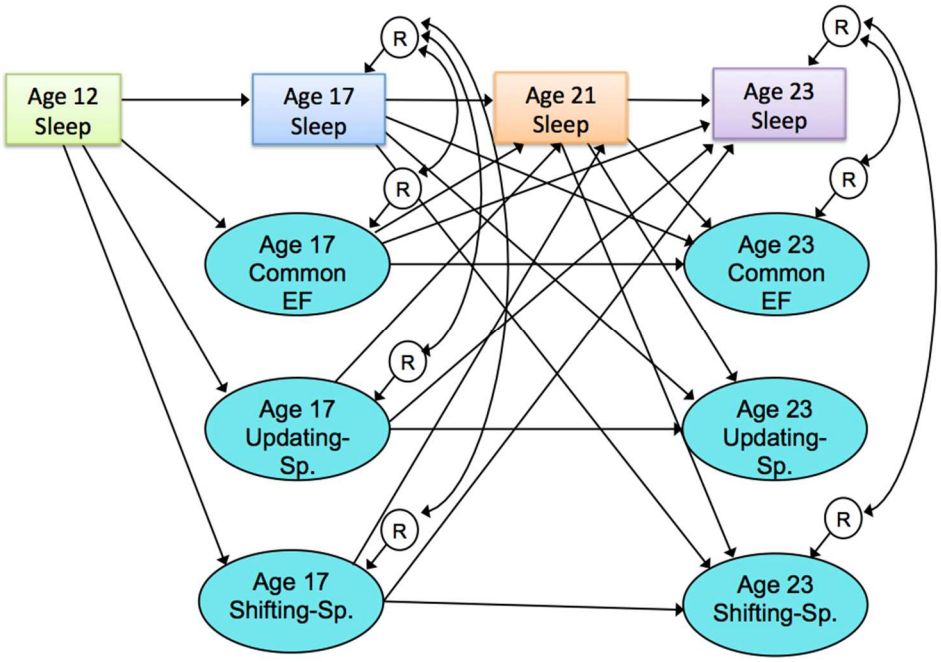

Figure 4.9. Cross-lag model of the effects of sleep on EFs.
Single headed arrows = regression path. Double headed arrows = correlation. Sleep = linear sleep duration. Common EF = Common Executive Function. Updating-Sp. = Updating Specific factor. Shifting-Sp. = Shifting Specific factor.

Better Common EF at age 17 predicted atypical weekday sleep at age 21, and fewer hours of weekday sleep at age 23 (see Figure 4.10, panel A). In general the weekday model shows significant residual correlations between linear and categorical sleep at each time point, but only the residual correlation between linear sleep and Common EF and categorical sleep and Shifting-specific abilities were significant at age 17. As expected, all autoregressive EF paths were significant in both the weekday and weekend sleep models.

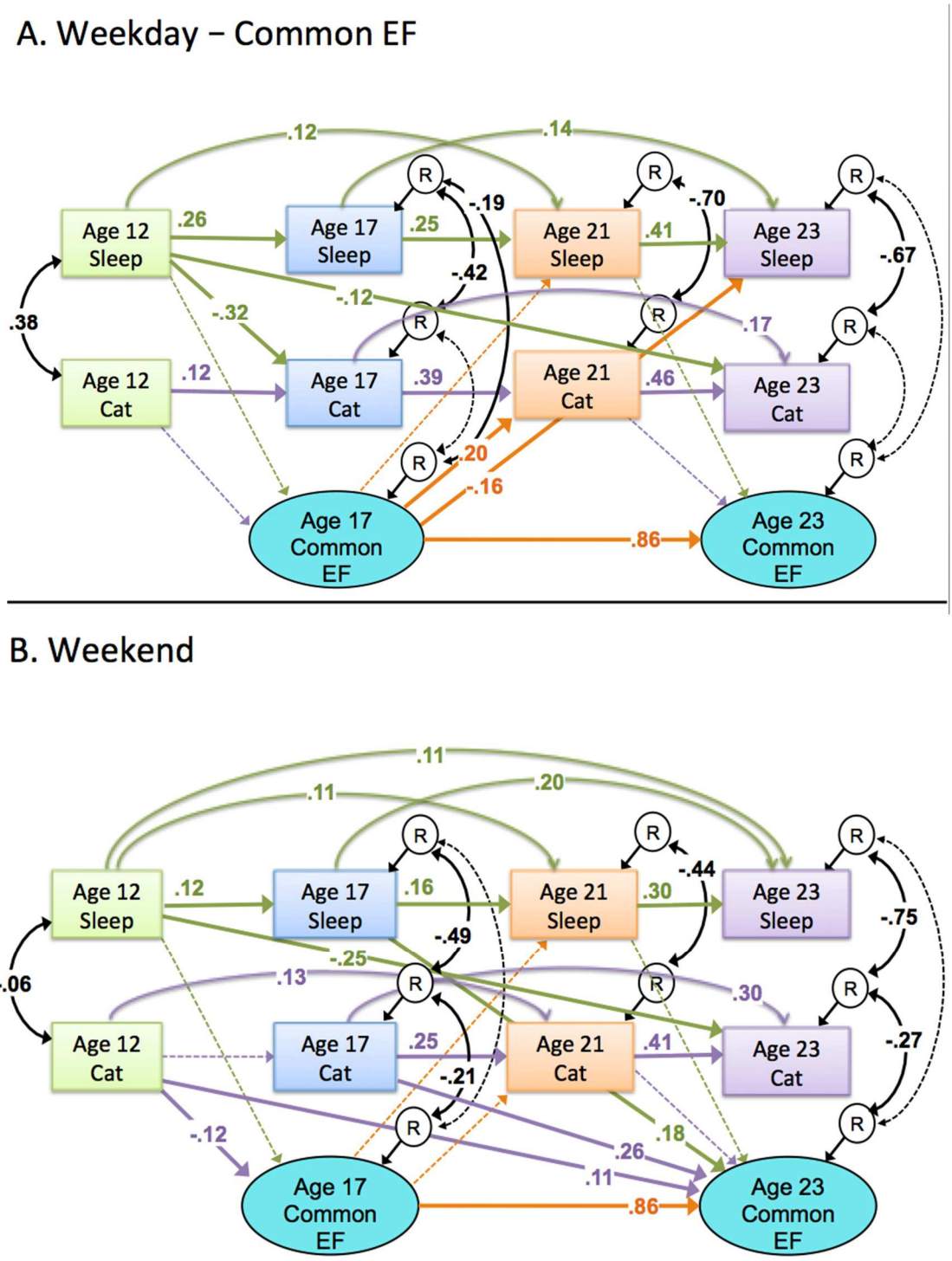

Figure 4.10. Cross-lag models of the effects of linear and categorical sleep with a Common EF latent factor. Panel A depicts weekday sleep; CFI=.957, RMSEA=0.026. Panel B depicts weekend sleep; CFI=.971; RMSEA=.022. Solid lines = $p<.05$. All dashed lines $p>.05$. Green lines = linear sleep was the predictor variable. Purple lines = categorical sleep was the predictor variable. Orange lines = EFs were the predictor variables. Sleep = linear sleep duration. Cat = Categorical sleep. Common EF = Common Executive Function.

Weekend sleep showed a different pattern of results from weekday sleep for Common EF. For example, atypical weekend sleep at ages 12 and 17 predicted better Common EF (see Figure 4.10, panel B). However typical weekend sleep at age 12 predicted better Common EF at age 17 while controlling for previous sleep durations. Holding categorical sleep and previous depression constant, longer sleep durations at age 17 predicted better Common EF at age 23. Common EF was the only EF to show within-wave correlations between weekend sleep and an EF: Both age 17 and 23 typical sleep was associated with better Common EF controlling for previous sleep and Common EF.

Within the same weekday model, Updating-specific abilities and Shifting-specific abilities were also examined. Updating results are presented in Panel A of Figure 4.11. Controlling for age 12 sleep, typical sleep and shorter sleep durations at 17 predicted better Updating-specific abilities at age 23, while better Updating-specific abilities at age 17 predicted typical weekday sleep and longer sleep durations at age 21. Similarly, longer weekday sleep durations predicted better Updating-specific abilities at 23; however typical sleep at age 21 predicted worse Updating-specific abilities at 23.

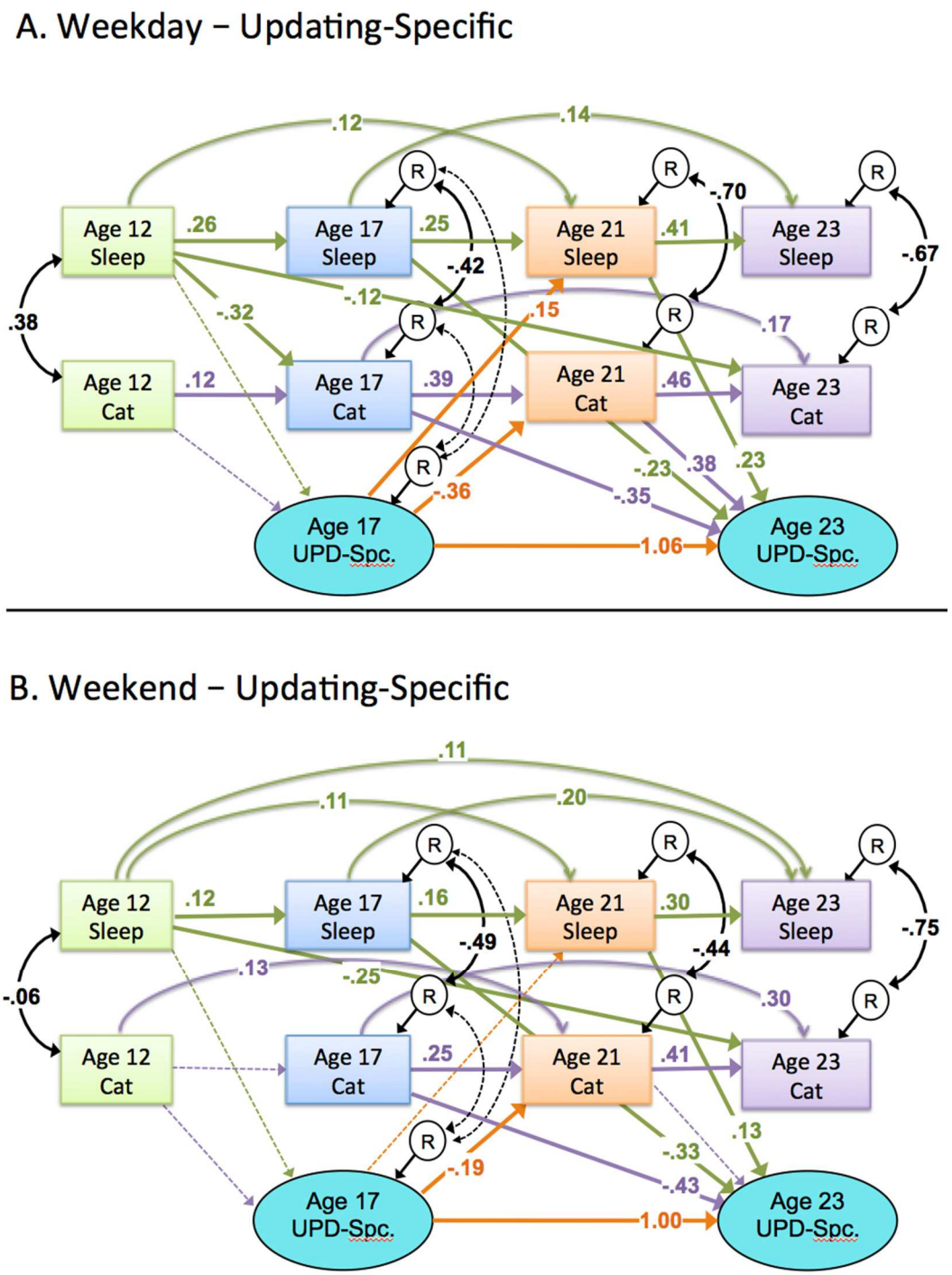

Figure 4.11. Cross-lag models of the effects of linear and categorical sleep with an Updating-Specific latent factor. Panel A is the model with weekday sleep. Panel B is the model with weekend sleep. Solid lines = $p<.05$. All dashed lines $p>.05$. Green lines = linear sleep was the predictor variable. Purple lines = categorical sleep was the predictor variable. Orange lines = EFs were the predictor variables. Sleep = linear sleep duration. Cat = Categorical sleep. UPD-Spc. = Updating-Specific abilities.

Weekend and weekday sleep showed similar patterns with Updating-specific abilities (see Panel B of Figure 4.11). Holding previous sleep constant, better age 17 Updating-specific abilities predicted typical sleep at age 21, while typical sleep at 17 predicted better Updating-specific abilities at 23. Shorter sleep durations at age 17 predicted better Updating-specific abilities at age 23, but shorter sleep durations at 21 predicted worse Updating-specific abilities at 23. There could be a number of explanations for this reversal in patterns with weekday and weekend sleep, ranging from environmental to biological (such as an increased sleep drive in early adulthood that's associated with better Updating-specific abilities), however more research will be needed to determine why this flip occurs.

The same flip exists in weekday sleep with Shifting-specific abilities: Shorter sleep durations at age 17, but longer sleep durations at 21 predict better Shifting-specific abilities at age 23 (see Figure 412). Also similar to previous EFs, better Shifting-specific abilities at 17 predict typical sleep at 21. Consistent with weekday sleep and Updating-specific abilities, after controlling for previous Shifting-specific abilities and sleep, typical weekday sleep at 21 predicts worse Shifting-specific abilities at 23. This same pattern is seen with age 17 typical weekday sleep predicting worse Shifting-specific abilities at 23, however the opposite pattern was seen with Updating-specific abilities for that time point. Atypical sleep at ages 17 and 21 predicting worse shifting specific abilities at 23.

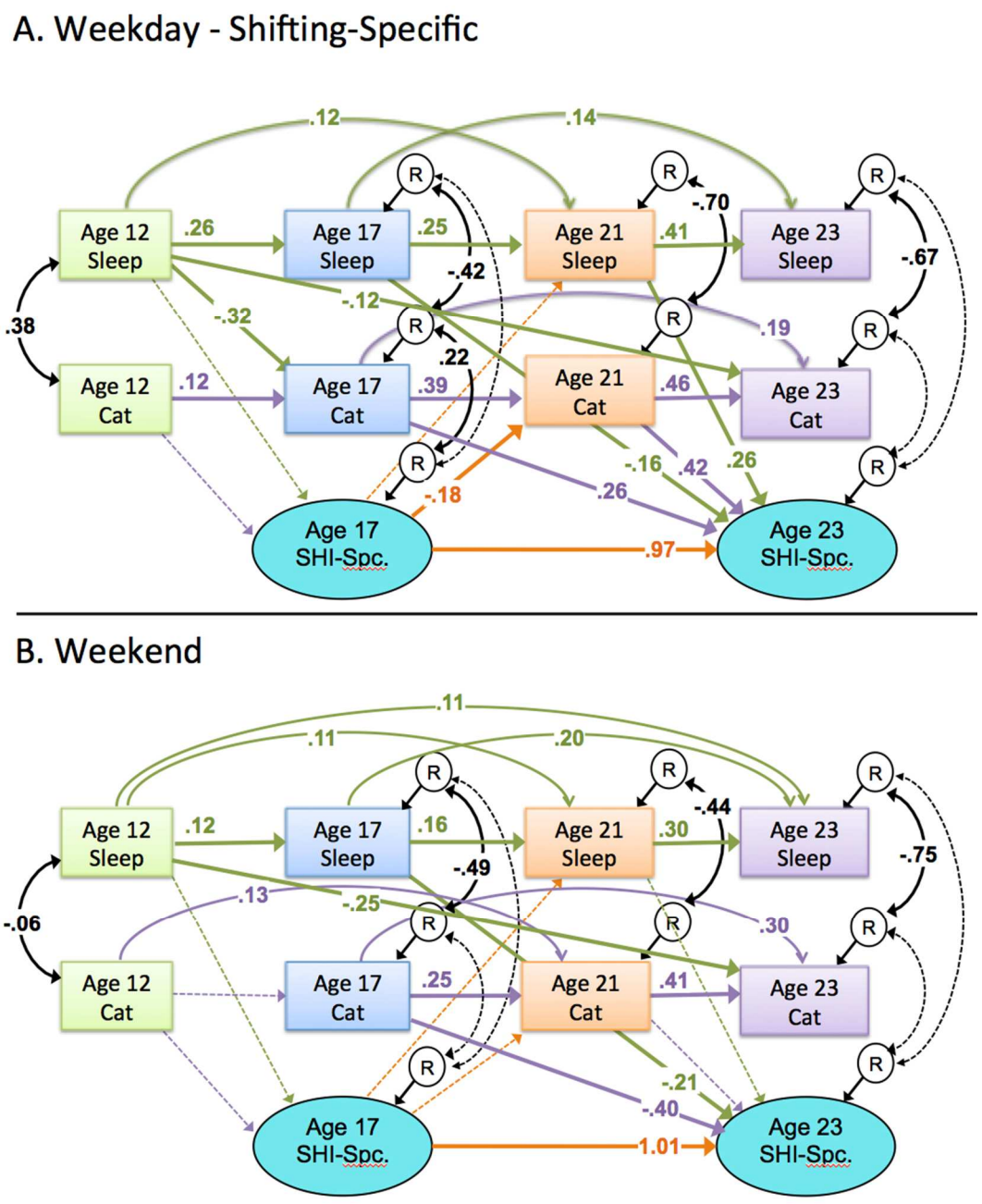

Figure 4.12. Cross-lag models of the effects of linear and categorical sleep on a Shifting-Specific latent factor. Panel A is the model with weekday sleep. Panel B is the model with weekend sleep. Solid lines = $p<.05$. All dashed lines $p>.05$. Green lines = linear sleep was the predictor variable. Purple lines = categorical sleep was the predictor variable. Orange lines = EFs were the predictor variables. Sleep = linear sleep duration. Cat = Categorical sleep. SHI-Spc. = Shifting-Specific abilities.

The cross-lag relationships between weekend sleep and Shifting-specific abilities were fewer than with weekday sleep (see Figure 4.12). What remained was consistent

with weekend sleep and Updating-specific abilities, but inconsistent with weekend sleep and Common EF: Shorter sleep duration and typical weekend sleep at age 17 predicted better Shifting-specific abilities at age 23. While Common EF and Shifting-specific abilities often show opposing patterns with maladaptive behaviors, it was somewhat unexpected that Common EF and Updating-specific abilities would show opposing patterns. This will be covered more extensively in the discussion.

Sleep Consistency

In addition to understanding sleep duration's relationship with depression and EFs, I also wanted to know if different patterns of sleep predicted differences in depression or EFs. For example, people with better EFs or less depression might stay up later working during the week and then catch-up on sleep on the weekends or might consistently make sure they get enough sleep no matter if it is a weekday or weekend. So, I ran within-wave regressions using a difference score between weekend sleep and weekday controlling for the average amount of sleep (see methods for a description on how the average sleep and sleep difference variables were created; results presented in Table 4.8).

Table 4.8 Standardized Beta Estimates from Models with Sleep duration Average and Difference Scores

Weekend-Weekday	Sleep Variable	Age 12	Age 17	Age 21	Age 23
CESD					
Age 12	Average	**0.20**			
	Difference	**-0.19**			
Age 17	Average		**0.21**		
	Difference		**-0.21**		
Age 21	Average			**0.43**	
	Difference			**-0.32**	
Age 23	Average				0.12
	Difference				-0.11
EFs 17					
Common EF	Average	-0.11	-0.03	0.04	
	Difference	0.05	0.07	-0.01	
Updating-Specific	Average	0.17	-0.01	**-0.26**	
	Difference	0.03	0.00	**0.21**	
Shifiting-Specific	Average	0.03	-0.02	-0.07	
	Difference	-0.05	0.02	-0.02	
EFs 23					
Common EF	Average			0.02	0.09
	Difference			0.04	0.06
Updating-Specific	Average			**-0.21**	-0.02
	Difference			**0.21**	0.03
Shifting-Specific	Average			0.10	0.12
	Difference			-0.12	-0.11

Note. Presented are the standardized beta estimates from regressions where sleep average and sleep duration difference scores predict depression symptoms and EFs. Bold = $p<.05$. CESD = Center of Epidemiological Studies Depression Scale scores; EFs 17 = EF abilities at age 17; EFs 23 = EF abilities at age 23.

Holding constant average sleep duration, those who slept less on weekends (and more on weekdays) showed more depression symptoms at age 12, 17, and 21. The only EF to show a relationship with weekend-weekday sleep differences was Updating-specific abilities. Again, after controlling for average sleep duration, catching up on sleep on the weekends during college (age 21) was associated with better Updating-specific abilities at both ages 17 and 23.

Aim 4 – What is the relationship between depression, sleep, and EF?

The final aim of this chapter was to look at the relationship between sleep, depression, and EFs in the same model in order to better understand how these variables affect each other. I ran a series of mediation models with both sleep and depression symptoms measured at the same age and the closest wave of EF data (eg. age 12 depression and sleep and age 17 EFs). I ran these models with both the linear and categorical sleep variables to better understand the non-linearity of the relationships. Bootstrap confidence intervals were also calculated, and in cases of partial indirect effects, the confidence intervals for the direct effect did not include zero. When I found evidence for indirect effects, depression was mediating the modest relationship between EFs and sleep (see Table 4.9 for results).

Table 4.9. Indirect Effects of Sleep on EFs Through Depression

		Common EF			**Updating-Specific**			**Shifting-Specific**		
Sleep Variable		Total	Indir	Direct	Total	Indir	Direct	Total	Indir	Direct
		Age 17 EFs								
Age 12 Weekday	linear	0.07	**0.02**	0.05	**-0.16**	**0.03**	**-0.19**	-0.03	-0.02	-0.01
	categ	-0.00	-0.01	0.01	-0.00	-0.01	0.01	0.04	0.01	0.03
Age 12 Weekend	linear	0.03	0.02	0.01	0.05	0.02	0.03	-0.03	-0.02	-0.02
	categ	**-0.13**	-0.01	**-0.12**	**-0.13**	-0.01	0.01	0.02	0.01	0.01
Age 17 Weekday	linear	-0.06	0.01	-0.07	0.05	0.00	0.05	0.08	0.00	0.08
	categ	**-0.14**	**-0.04**	-0.10	**-0.14**	-0.01	0.11	**0.17**	0.00	**0.17**
Age 17 Weekend	linear	0.00	0.02	-0.02	0.02	0.01	0.01	0.10	-0.00	0.09
	categ	-0.11	-0.01	-0.10	-0.11	-0.00	0.02	**0.21**	0.00	**0.21**
Age 17 Typical	linear	-0.05	0.02	-0.07	0.03	0.01	0.02	-0.05	0.00	-0.05
	categ	-0.09	**-0.04**	-0.05	0.11	-0.01	0.12	0.13	-0.00	0.13
Age 17 Last night	linear	0.02	0.03	-.02	-0.04	0.01	-0.04	0.01	-0.00	0.01
	categ	-0.12	-0.01	-0.11	0.02	-.00	0.02	0.03	0.00	0.03
Age 21 Weekday	linear	0.01	0.02	-0.01	0.05	0.00	0.05	0.07	-0.07	0.08
	categ	0.07	-0.02	0.09	-0.15	-0.00	-0.15	0.03	0.01	0.02
Age 21 Weekend	linear	-0.04	0.01	-0.05	0.10	0.00	0.09	0.02	-0.00	0.02
	categ	-0.06	-0.01	-0.05	-0.04	-0.00	-0.04	0.14	0.00	0.14
		Age 23 EFs								
Age 21	linear	-0.03	**0.03**	-0.06	0.06	0.01	0.05	0.02	0.00	0.02

Weekday	categ	0.02	-0.03	0.04	-0.06	-0.01	-0.05	0.10	-0.00	0.10
Age 21 Weekend	linear	0.01	**0.02**	-0.02	0.12	0.01	0.11	-0.03	0.00	-0.03
	categ	-0.04	-0.02	-0.02	-0.03	-0.01	-0.02	0.11	-0.00	0.11
Age 23 Weekday	linear	**-0.15**	-0.00	**-0.14**	-0.11	-0.01	-0.11	-0.09	0.00	-0.09
	categ	-0.08	-0.02	-0.06	**-0.19**	**-0.02**	**-0.17**	-0.03	-0.00	-0.03
Age 23 Weekend	linear	-0.02	0.00	-0.03	0.04	0.00	0.04	-0.03	0	-0.03
	categ	-0.11	-0.01	-0.11	0.03	-0.01	0.04	0.05	-0.00	0.06

Note. Presented are the standardized beta estimates from mediation analyses where depression mediates the relationship between sleep duration and latent EFs. Bold = $p<.05$. Linear = linear sleep variable. Indir = Indirect effect. Categ = categorical sleep variable.

The following results are for models that had both a significant total effect between sleep and EF, and a significant indirect effect of sleep through depression symptoms. A significant partial indirect effect was observed, but the overall effect was minimally changed (it went from .16 to .18.) and is not particularly notable.

Instead of a partial indirect effect, it seems that the effect of atypical age 17 sleep to predict worse age 17 Common EF, controlling for linear sleep effects, is mediated through depression at that age. In other words, once accounting for depression, sleeping too much or too little does not predict worse Common EF.

Results also show a partial indirect effect at age 23 of atypical weekday sleep predicting worse Updating-specific abilities, through depression, holding linear sleep duration constant. So at least some of the relationship between sleeping too much or too little at age 23 and worse Updating-specific abilities at age 23 is mediated through level of depression at that age.

Both sleep and depression, and depression and EFs have stronger total relationships than sleep and EFs, yet, those relationships do not seem to be due to their relationships with EFs and sleep, respectively. It is the relatively weaker relationship

between sleep durations and EFs that is at least partially mediated by depression symptoms. I observed some significant indirect effects of sleep on Common EF and Updating-specific abilities through depression at a given age, but none with Shifting-specific abilities. However, the total amount of phenotypic variance explained between sleep and EFs was small to begin with, and the indirect effects are even smaller and often did not result in a substantial amount of change despite being significant. Therefore, these results should be interpreted cautiously until replicated.

Discussion

Individual differences in sleep duration were meaningfully related to both depression and EFs, both within a given time point and longitudinally. When significant, regressions including both linear and categorical sleep indicate that shorter sleep durations, and sleeping more or less than typical, are associated with increased depression, better Shifting-specific abilities, worse Common EF, and mixed associations with Updating-specific abilities. With the exception of longer sleep predicting worse Updating-specific abilities at age 12, these associations were all in the hypothesized directions.

After controlling for previous levels of sleep and depression or EFs in longitudinal models the patterns of association sometimes became less clear. Some consistent patterns emerged across longitudinal models though. For example, shorter weekday sleep durations at age 12 significantly predicted both increased depression symptoms and depression diagnosis at age 17. Also, more age 17 depression symptoms

predicted fewer hours of both weekday and weekend sleep at age 21, while MDD diagnosis at age 17 predicted atypical sleep at age 21. Patterns also emerged across the EF models, such that Updating- and Shifting-specific abilities often showed more similar patterns with each other than with Common EF. For example, after controlling for previous sleep, shorter sleep duration and typical weekday and weekend sleep at age 17 predicted both better Updating- and Shifting-specific abilities, but longer weekend sleep duration and typical weekend sleep predicted worse Common EF. Similarly, better age 17 Updating- and Shifting-specific abilities predicted typical weekday sleep at 21, but worse age 17 Common EF predicted typical weekday sleep at 21. A trade-off between Common EF and the other two EFs with sleep seems to be relatively consistent across models.

A portion of the associations seen in the cross-lag models, particularly those specific to each model, are inconsistent from what would be predicted by the within-wave regressions. In fact, more associations emerge in the EF models than would be expected from those regressions. There are a few potential reasons for differences. First, sleep could have different associations with depression or EFs longitudinally due changing environmental and developmental factors between measurements. Second, the cross-lag models take into account previous sleep patterns, allowing for more complex associations to emerge across time. For example, in the weekday model, better updating- and Shifting-specific abilities predict typical sleep at age 21, but typical sleep at age 21 predicts worse updating- and Shifting-specific abilities at age 23. The cross-lag model was used to assess longitudinal relations precisely because it examines them in the context of factors that matter: previous sleep durations and depression or EFs. If previously shorter sleep duration at age 12 have induced depression at age 17, and then age 17 depression is

influencing age 21 sleep (such as is the case with weekday sleep durations and depression symptoms), that is more informative compared to a simple association between age 12 sleep duration and age 21 depression.

To expand upon the simple regressions, I ran mediation models that included both categorical and linear sleep durations, depression symptoms, and EFs. There were theoretical motivations for any of the three phenotypes to act as a mediator between the other two. For example, depression could underlie abnormalities to sleep duration and decreased EFs. Alternatively, worse EF abilities could cause mismanagement of time leading to less sleep and worse decision-making leading to depression. Short sleep duration could also impact emotional affect and executive abilities. When significant indirect effects were observed, there were indirect effects of sleep duration on EFs through depression. While significant, these indirect effects were often small in magnitude and often only partial mediations. Given that the total effects of sleep on EFs were also small to begin with, these results may not be particularly meaningful.

Overall, individual differences in sleep duration predict variation in both depression and EFs. Both linear and categorical sleep durations, when individuals exhibit extreme sleep duration (sleeping more or less than the majority of other people), differentially predict depression and EFs. When it comes to understanding sleep durations relationships with EFs and depression longitudinally, sleep duration seems to have a bidirectional effect with both phenotypes. Sometimes previous sleep duration predicts later depression (or EFs), sometimes those phenotypes predict later sleep duration. While depression, EFs, and sleep duration all relate to each other, no one variable seems to substantially mediate the relationship of the other two. This could potentially indicate that

other factors, unstudied here, are responsible for these associations. The next step is to decompose the variation between sleep durations and these phenotypes into genetic, shared, and non-shared environmental factors. This can indicate where to look for variables that influence these relationships.

CHAPTER 5

Genetic and Environmental Relationships between Sleep duration, EFs and Depression

Understanding the relationships between sleep duration, EF, and depression at the phenotypic level is important, but understanding the extent to which genes and environment influence these variables and their relationships can provide additional insight into how these variables affect each other. For example, if environmental influences shared between sleep duration and depression are driving the phenotypic associations, then identifying and targeting environments that influence both should help improve both. However, shared genetic influences could also be contributing to the phenotypic association between sleep duration and EFs or depression.

While a significant amount of sleep research focuses on interventions to target sleep habits and environment (e.g., Morin, Culbert, & Schwartz, 1994; Irwin, Cole & Nicassio, 2006; Meltzer & Mindell, 2014), the heritabilities of sleep characteristics are also well supported by previous research (for a review see Barclay & Gregory, 2013). For sleep duration in particular, there have been discrepancies in results, particularly as age, sample size, and measurement of sleep duration varied. For example, a study conducted by Barclay and colleagues (2010) found shared environment, rather than genetics, to be important for sleep duration in adulthood. However, other studies found sleep duration to be moderately heritable in both childhood ($h^2 = 0.52$, Sletten et al., 2013) and adult samples ($h^2 = 0.44$, Partinen et al., 1983).

During the transition from adolescence to young adulthood, pubertal onset and hormonal changes accompany changes in sleep (Hagenauer, Perryman, Lee, &

Carskadon, 2009). Notably, circadian phase is delayed as adolescents transition towards a more evening chronotype, with females starting this delay process and peaking a year earlier than males. This delay in circadian phase is observed in a number of other cultures and is conserved in other mammalian species, all of which points to genetic contributions to these changes in sleep. While there could be genetic effects that contribute to adolescent sleep duration before pubertal onset, I expect heritability of sleep to increase with age in the LTS sample as more genetic contribute to the phenotype, and for more genetic stability between waves after the peak circadian phase delay in the late teens.

Depression is also moderately heritable (heritability estimate = .37; Sullivan, Neale, Kendler, 2000). A meta-analysis of age-related changes in heritability indicates that the heritability in depression significantly increases over time from approximately age 10 through age 30 years (Bergen, Gardner, & Kender, 2007). As individuals age, there may be environmental changes that allow for genes to have a larger impact on depression, or new genetic influences that come online to affect sleep, and sleep's relationship with both EF and depression.

As both depression and poor sleep are heritable, it is possible that some of the same genes are contributing to both poor sleep and depression. In fact, one study found that the relationship between children's self-reported depression symptoms and their sleep problems (reported by their parents) seemed to be mostly genetic in nature (r_A = .64 at age 8; Gregory et al., 2016). However, that sample had been mostly selected to be high in anxiety at age 7, and it is unclear if this pattern holds at later time points in development. It would be helpful to know if there are genetic, shared environmental or non-shared

environmental correlations between depression and sleep characteristics at later time points as well.

EFs, particularly as measured by the bifactor latent variable model of EF, are highly heritable (Friedman et al., 2009; Miyake & Friedman, 2012). The latent factor constructs of Common EF, Updating-Specific, and Shifting-Specific have heritability estimates of 98%, 100% and 74% respectively in our sample at age 17 (Friedman et al., 2009). While the phenotypic relationship between depression and EF is supported by a number of studies (McClintock, et al., 2010; Snyder, 2013), less is known about the genetic relationship between depression and EF. Using the latent variable model of EF, the study described earlier by Friedman (2016) provides evidence that this relationship is genetic in nature, where genes that promote lower depression scores at age 17, promote higher EF at age 23 (reproduced in Figure 5.1). A genetic relationship between EF and depression, in the absence of an environmental relationship, suggests that there may be common genetic risk, where the same genes influence both EF abilities and levels of depression.

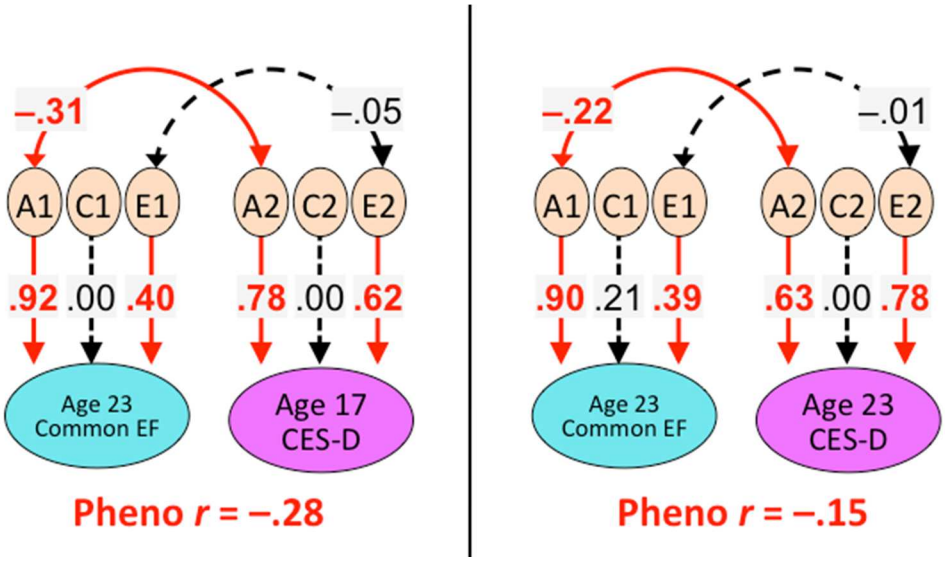

Figure 5.1. Genetic relationships between Common EF and Depression Scores. Reproduced from Friedman, et al., in prep. Red lines = p <.05. CESD = Center for Epidemiological Studies-

Depression score, modeled as a latent variable. Common EF = Common EF latent factor. A = genetic effects. C = Shared environmental effects. E = nonshared environmental effects.

In addition, new genetic and environmental factors could alter the relationship between sleep duration, depression, and cognition during different developmental phases, such as the onset of puberty, to moving out and starting college, a new job, or a new family. Depending on what the relationship is at a given stage, and whether it is genetic or environmental in nature could influence how best to treat sleep problems, depression, or cognitive dysfunction.

Current Study

This study first assessed the heritability of sleep duration at 4 time points from early adolescence (age 12) to young adulthood (age 23). This was done for both weekday and weekend sleep duration, first as linear, then as categorical variables. The categorical variable was created to capture nonlinear effects between sleep duration and depression and EFs in that, both those who slept more or less than typical showed increased depression or different relationships with EF over and above nonlinear sleep duration.

As individuals undergo a lot of developmental changes relevant to depression, sleep duration, and EFs during the transition from adolescences to young adulthood; it is likely that new genetic or environmental factors become relevant during that time. Although it is unclear if those new influences would jointly influence both sleep duration and EFs or sleep duration and depression. Alternatively, given the stability of EFs at the latent variable level and the consistency of earlier sleep to predict later sleep at the phenotypic level, it is plausible that there are genetic or environmental influences that are consistent across this time period. So, to better understand the genetic and environmental influences of sleep duration, I asked whether genetic or environmental variation at earlier

waves was shared with later waves of sleep duration or if new influences became important throughout development.

Similarly, the genetic and environmental influences on weekday sleep could be consistent with weekend sleep, or there could be unique contributions to each. Circadian rhythm and sleep drive would be consistent across weekdays and weekends, but differences in weekday and weekend schedules could lead people to give in to or fight these biological processes. For example, people with a delayed circadian rhythm might have to force themselves to wake up early during the week due to obligations, resulting in shorter sleep durations, but then can sleep in on the weekends, resulting in more typical, longer sleep durations. So while that individual would show different weekday, weekend patterns, the same genetic influences could be influencing both. Therefore, Cholesky decompositions were used to determine if weekday and weekend sleep had shared or unique influences acting on them within each time point.

After the more in-depth characterization of linear sleep duration, univariate heritability estimates for categorical sleep were conducted. These analyses were followed up by a bivariate Cholesky decomposition between linear sleep duration and categorical sleep duration, to better understand if the same or unique influences were contributing to both.

Method

Please see Chapter 4 Methods for Sample and Materials.

Analyses

It is plausible that as new hormonal and environmental factors occur, the heritability of sleep duration may change. So the first set of genetic analyses estimate the

heritability of sleep in our sample at each age using an ACE twin models. Second, I examined the stability of the genetic influences on sleep, as well as whether new, unique genetic influences come online throughout development using multivariate modeling (a series of bivariate Cholesky decompositions; see Figure 5.2).

The bivariate Cholesky decomposes the genetic (and environmental) contributions by utilizing cross-twin, cross-trait correlations – for example, the correlation between one twin's depression score with the co-twin's sleep duration. When cross-trait twin correlations are greater for MZ than for DZ twin pairs, genetic factors are implicated in the covariation across traits. If cross-trait twin correlations are approximately equal for MZ and DZ twin pairs, then environmental factors are implicated in the covariation across traits. In Figure 5.2, A1 represents the genetic contributions to individual differences in weekday sleep duration. Path a_{21} represents the extent that those genetic influences also predict weekend sleep duration, and therefore the genetic stability between phenotypes. Path a_{22} represents the new, or unique genetic influences weekend sleep duration.

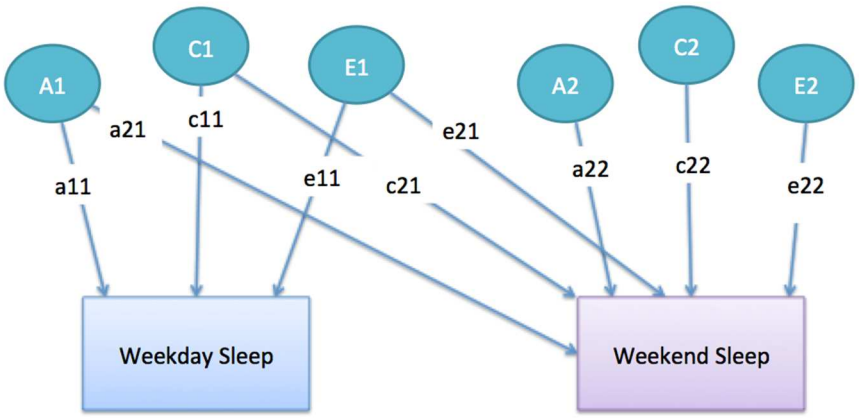

Figure 5.2 Example of a bivariate Cholesky decomposition with A (additive genetic), C (shared environment), and E (non-shared environment) loadings.

Importantly, the order of the variables in a Cholesky decomposition can matter. In the models between time points, naturally the younger time points will precede the later time points. However, when looking across phenotypes, the ordering of the variables for the trivariate Cholesky, determines what is partialled out. For example, if Common EF is placed first in a Cholesky decomposition, with sleep second, and depression third, then the a_{23} path will represent the genetic contributions of sleep independent of EF that predict depression, and a_{22}*a_{23} is the covariance between sleep and depression independent of EF. These differ depending on the order of the variables.

The cross paths from a bivariate Cholesky decomposition indicate where the covariance is coming from (A, C, or E). This can be used to compute the genetic (and environmental) correlations with the following equation: $r_A = \frac{a_{11}a_{21}}{\sqrt{a_{11}^2 * (a_{21}^2 + a_{22}^2)}}$.

I also examined genetic and environmental correlations between sleep and depression at each age using a correlated factor solution of the Cholesky decomposition, explained above. I obtained correlations between our genetic A components (r_A), shared environmental C components (r_C), and non-shared environmental E components (r_E). If a genetic correlation between sleep and depression occurs at age 12, but not during young adulthood, it could be partially explained if individuals with a high genetic loading for depression already suffer from depression at age 12, but environmental factors contribute to depression or poor sleep habits by age 23. Therefore, these patterns might have important implications for how best to approach treatments at various ages.

Results

Aim 1- Is sleep duration heritable across development?

Heritability estimates for sleep duration have been inconsistent in the literature,

with some studies finding duration to be heritable, but others not. I ran univariate ACE twin analyses for both weekday and weekend sleep duration at each time point (see Table 5.1). In general, using ACE twin models, sleep seems to be moderately heritable with a^2 estimates ranging from .15-.4.[6] While not explicitly tested, trends in the data suggest that the heritability of sleep duration seems to increase with age in this sample, with weekend sleep being more consistently heritable than weekday sleep.

Table 5.1. Univariate Heritability Estimates for Linear Sleep Duration

Linear Sleep		rMZ	rDZ	A	C	E
Age 12	weekday	0.30	0.21	0.19	0.11	**0.70**
	weekend	0.13	0.07	0.13	0.00	**0.86**
Age 17	weekday	0.31	0.15	0.30	0.00	**0.69**
	weekend	0.37	0.19	**0.38**	0.00	**0.63**
	typical	0.31	0.17	0.29	0.02	**0.68**
	last night	0.41	0.21	**0.40**	0.01	**0.59**
Age 21	weekday	0.17	0.02	0.15	0.00	**0.85**
	weekend	0.29	0.06	**0.26**	0.00	**0.74**
Age 23	weekday	0.43	0.24	**0.38**	0.05	**0.57**
	weekend	0.33	0.16	**0.33**	0.00	**0.67**

Note. Presented here are the univariate heritability estimates for linear weekday and weekend sleep duration at each age. Boldface type indicates $p<.05$ according to chi-square difference tests. A = additive genetic heritability estimates. C = shared environmental estimates. E = Non-shared environmental estimates. rMZ = monozygotic twin correlations. rDZ = dizygotic twin correlations. Model fit indices ranged from CFI= .951-1.000; RMSEA= .000-.035. Models with poor fit included: Typ 17: χ^2= 13.08(6), p=.042; CFI = .710; RMSEA= .081. Last night 17: χ^2= 12.44(6), p=.053; CFI = .808; RMSEA= .077. Weekday 23: χ^2= 12.56(6), p=.051; CFI = .863; RMSEA= .075. Weekend 23: χ^2= 10.62(6), p=.101; CFI = .780; RMSEA= .063.

Reasons for these trends could stem from the fact that as individuals progress from adolescence to young adulthood, a sleep schedule is less likely to be as strictly imposed, whether it is because the individual left home for college or the parent has permitted them more flexibility in determining their own schedule (Carskadon, Acebo &

[6] ACE models were used instead of AE models because age 12 suggested the possibility of shared environmental influences on sleep duration and I felt it was important to allow for potential common shared-environmental influences in bivariate Cholesky decompositions with other phenotypes. However, if AE models are used, then all A estimates become statistically significant.

Jenni, 2004). Then the sleep schedule, and duration, become the individual's decision and can be determined by their biological sleep drive, delayed circadian phase, and priorities such as socialization, activity level, or work ethic (Hagenauer, Perryman, Lee, & Carskadon, 2009). Puberty triggers hormones that also influence sleep duration, and perhaps genetic variation relevant to those hormonal changes are not apparent until closer to age 17 (Knutson, 2005).

Do the same genetic variants that contribute to early sleep contribute to later sleep?

While it useful to understand whether genetic and environmental influences are present for sleep duration at a given age, I wondered whether the same genetic variants are shared between earlier and later sleep.

Table 5.2 Cholesky decomposition loadings for sleep duration at earlier waves to later waves.

Model		a11	a21	a22	c11	c21	c22	e11	e21	e22
12 to 17	weekday	0.46	0.41	0.38	0.31	0.00	0.00	**0.83**	0.08	**0.82**
	weekend	0.28	0.43	0.42	0.21	-0.07	0.00	**0.93**	0.00	**0.8**
17 to 21	weekday	0.51	0.15	0.33	0.19	-0.06	0	**0.83**	**0.24**	**0.9**
	weekend	0.54	0.05	0.47	0.11	0.05	0	**0.83**	**0.2**	**0.85**
21 to 23	weekday	**0.45**	**0.63**	0	0.12	0.2	0	**0.9**	**0.15**	**0.74**
	weekend	**0.51**	**0.57**	0	-0.16	0.33	0	**0.85**	0.06	**0.75**
12 to 21	weekday	0.41	0.3	0.33	0.36	-0.08	0	**0.84**	0.03	**0.89**
	weekend	0.41	0.33	0.22	0.35	0.08	0	**0.84**	0	**0.92**
12 to 23	weekday	0.43	0.29	0.52	0.33	-0.02	0.25	**0.84**	0.04	**0.76**
	weekend	0.26	0.51	0	0.25	-0.25	0	**0.93**	**0.12**	**0.82**
17 to 23	weekday	**0.55**	0.18	0.59	0	0.05	0.2	**0.83**	**0.16**	**0.74**
	weekend	**0.53**	0.08	0.43	0.28	0.33	0	**0.8**	**0.2**	**0.81**

Note. Standardized path loadings for bivariate Cholesky decompositions between earlier sleep and later sleep. a11 is the genetic loading unique to weekday sleep. a21 represent the shared genetic variance between weekday and weekend sleep. a22 represents the genetic variation unique to weekend sleep. C11, c21, and c22 = shared environmental loading. e11, e21, and e22 = non-shared environmental loadings. Model fit indices ranged from: χ^2= 11.57(17), p=.825 to χ^2= 34.95(17), p=.006; CFI= .787-1.000; RMSEA= .000-.072, with almost every model having at least one of the two model fit indices within acceptable ranges.

There was little evidence (see Table 5.2) for shared genetic variation between earlier sleep and later sleep, except at age 21 to 23. This was present for both weekday and weekend sleep. Also it seems as if there were significant genetic influences on age 17 sleep when modeled with age 23 sleep duration. Consistently there were non-shared environmental influences that were unique to weekday and weekend sleep. Non-shared environmental influences were common to both weekday and weekend sleep at age 17 with age 21 and age 23. However, only weekday sleep showed common non-shared environmental influences between age 21 sleep duration and age 23 sleep duration. These results indicate that the phenotypic associations for earlier sleep with later sleep might initially be attributable to common non-environmental influences starting at 17, but then at age 21 shared genetic influences also contribute.

Is weekday sleep related to weekend sleep within-wave?

The different phenotypic relationships that weekday and weekend sometimes showed with depression and EFs begged the question: Do weekday and weekend sleep duration share a genetic or environmental etiology or are there unique contributions working on both? Therefore, I used bivariate Cholesky decompositions with weekday and weekend sleep at each wave. Results are presented in Table 5.3.

Table 5.3 Bivariate Cholesky Decomposition from Weekday to Weekend Sleep

Weekday to Weekend	a11	a21	a22	c11	c21	c22	e11	e21	e22
Age 12	**0.45**	0.15	**0.30**	0.31	0.15	0.00	**0.83**	**0.40**	**0.84**
Age 17	0.50	0.33	0.51	0.21	-0.06	0.00	**0.84**	**0.20**	**0.77**
Age 21	**0.41**	**0.47**	0.20	0.00	0.00	0.00	**0.92**	**0.27**	**0.82**
Age 23	**0.60**	**0.55**	0.00	0.26	-0.16	0.00	**0.76**	**0.18**	**0.80**

Note. Presented are standardized path loadings from weekday to weekend sleep. Boldface type indicates $p<.05$ according to chi-square difference tests. a11 is the genetic loading unique to weekday sleep. a21 represent the shared genetic variance between weekday and weekend sleep. a22 represents the genetic variation unique to weekend sleep. c11, c21, and c22 = shared environmental loading. e11, e21, and e22 = non-shared environmental loadings. Model fit indices

ranged from: χ^2= 7.86(17), p=.969 to χ^2= 31.97(17), p=.015 CFI= .928 to 1.000; RMSEA= .000 to .067.

Overall, there is genetic variation unique to weekday sleep and some shared genetic variation that also contributes to weekend sleep starting in early adulthood. Genetic influences becoming more influential around age 21 is consistent with the pattern observed between sleep duration at earlier waves with sleep duration at later waves. Also consistent with the previous analyses, there seems to be no contributions from shared environmental influences (C), and strong evidence for both unique and common non-shared environmental influences between weekday and weekend sleep.

Is categorical sleep heritable?

The phenotypic analyses showed that in addition to linear sleep duration, categorical sleep duration, sleeping more or less than typical (<7 or >9.5 for weekdays and 10.5 for weekends), has interesting relationships with depression and EFs. Better understanding the genetic and environmental influences on this categorical sleep variable can better inform the phenotypic relationships. Therefore I estimated heritability using a univariate twin analysis, results shown below in Table 5.4.

Table 5.4 Heritability and Environmental Estimates for Categorical Sleep Duration

Categorical sleep		rMZ	rDZ	A	C	E
Age 12	weekday	0.26	0.34	0.00	**0.30**	**0.70**
	weekend	0.12	0.22	0.00	0.17	**0.83**
Age 17	weekday	0.55	0.14	**0.53**	0.00	**0.48**
	weekend	0.57	0.37	**0.41**	0.16	**0.43**
	typical	0.50	0.37	0.25	0.24	**0.51**
	last night	0.58	0.41	**0.33**	0.25	**0.42**
Age 21	weekday	0.18	0.04	0.16	0.00	**0.84**
	weekend	0.46	0.01	**0.40**	0.00	**0.60**
Age 23	weekday	0.54	0.12	**0.50**	0.00	**0.50**
	weekend	0.44	0.19	**0.43**	0.00	**0.57**

Note. A = additive genetic heritability estimates. C = shared environmental estimates. Boldface

type indicates $p<.05$ according to chi-square difference tests. rMZ = monozygotic twin correlations. rDZ = dizygotic twin correlations. Model fit indices ranged from CFI: .958-1.000; RMSEA: .000-.045. Models with poorer fit included Weekend 17: $\chi2= 8.29(3)$, p=.04; CFI=.840; RMSEA=.101. Weekend 23: $\chi2= 9.28(3)$, p=.03; CFI=.481; RMSEA=.104. Weekday 23: $\chi2= 5.83(3)$, p=.12; CFI=.898; RMSEA=.070. Weekday 17: $\chi2= 4.76(3)$, p=.19; CFI=.929; RMSEA=.058.

Categorical sleep shows almost the exact same heritability pattern as linear sleep duration with the exception weekday categorical sleep is heritable at age 17, and weekend sleep at age 21 is only marginally significant. Again, additive genetic influences seem to start to come online in early adulthood, and any suggestion of shared environmental influences decrease with age. As this was a univariate model, linear sleep duration was not controlled for, and therefore, the results probably reflect some of the linear sleep heritability. The similarity in genetic and environmental etiology shared by linear and categorical sleep could be due to the fact that the same genetic variants or environments influence both.

Do categorical sleep and linear sleep share a genetic / environmental etiology?

Naturally, the next step was to examine whether or not linear and categorical sleep share a genetic or environmental etiology using bivariate Cholesky decompositions including both linear and categorical sleep. Plausibly the genes or environments that influence an individual's sleep duration are the same that influence whether that amount is more or less than typical. Yet, when including quadratic or categorical sleep into many of the models, linear sleep was still significant, indicating that they explained unique variance at the phenotypic level. This could indicate unique genetic or environmental influences on both types of sleep. Thus I ran bivariate Cholesky decompositions with linear sleep and categorical sleep within each time point.

As seen in Table 5.5, weekday linear and categorical sleep share a substantial amount of genetic variation at age 23, but not at any of the other time points. More often, linear and categorical sleep have overlapping non-shared environmental influences (E), as represented by the e21 path. This is seen for weekday sleep at ages 12, 17, and 23, as well as weekend sleep at ages 17 and 23. The amount of sleep a 17 year old estimates he "typically" gets and number of hours slept the night before testing, also shows this pattern. In other words, non-shared environmental factors that differentiate twins from each other influence both an individual's sleep duration, and whether or not he/she sleep more or less than a typical person.

Table 5.5 Bivariate Cholesky Decomposition of Linear and Categorical Sleep

Linear to Cat		a11	a21	a22	c11	c21	c22	e11	e21	e22
Age 12	weekday	0.35	-0.19	0.00	**0.41**	0.54	0.00	**0.84**	**0.29**	**0.59**
	weekend	0.35	0.31	0.00	-0.13	0.32	0.00	**0.93**	-0.14	**0.78**
Age 17	weekday	**0.57**	-0.48	0.52	0.11	0.15	0.00	**0.82**	**-0.23**	**0.43**
	weekend	**0.57**	-0.35	0.53	-0.10	0.41	-0.01	**0.82**	**-0.25**	**0.37**
	typical	**0.56**	-0.46	0.01	0.14	0.53	0.01	**0.81**	**-0.19**	**0.47**
	last night	-0.38	0.58	0.00	**-0.43**	0.25	**0.43**	**0.82**	**-0.34**	**0.31**
Age 21	weekday	0.30	0.04	0.00	0.21	0.57	0.01	**0.93**	0.00	**0.68**
	weekend	**0.47**	-0.44	0.01	0.16	0.33	0.00	**0.87**	0.03	**0.70**
Age 23	weekday	**0.54**	**-0.70**	0.02	-0.35	-0.08	0.00	**0.77**	**-0.46**	0.29
	weekend	0.03	-0.48	-0.16	**-0.51**	0.38	0.00	**-0.86**	**0.79**	-0.02

Note. Standardized path loadings from a bivariate Cholesky decompositions between linear and categorical sleep. Boldface type indicates $p<.05$ according to chi-square difference tests. a11 is the genetic loading unique to weekday sleep. a21 represent the shared genetic variance between weekday and weekend sleep. a22 represents the genetic variation unique to weekend sleep. c11, c21, and c22 = shared environmental loading. e11, e21, e22 = non-shared environmental loadings. Model fit indices ranged from: $\chi^2= 18.11(14)$, p=.202 to $\chi^2= 36.20(14)$, p=.001; CFI= .901 to .964; RMSEA= .038 to .094.

Often, linear sleep and categorical sleep show significant unique non-shared environmental influences that differentiated the two types of sleep from each other. Occasionally, linear sleep showed substantial unique additive genetic influences or unique shared-environmental influences. The amount of sleep the night before testing at

age 17 showed shared-environmental effects (environmental factors that make twins more similar to each other), that differentiated between sleep duration and typical vs. atypical sleep.

Aim 2 – Do depression and sleep duration share a genetic etiology?

Previous research has been conducted on the heritability of depression in a subsample of our data for both depression symptoms and diagnosis (h^2 = .30-.45 at age 22; Johnson et al., 2014). Similar heritability estimates were seen in the full LTS sample, with lower estimates at ages 12 and 17. After the initial confirmation of heritabilities of depression and sleep duration, variation was decomposed into unique and common genetic and environmental contributions. Results are shown in Tables 5.6 and 5.7.

Table 5.6 Bivariate Cholesky decompositions between linear sleep and depression symptoms

Lin Sleep & Dep		**a11**	**a21**	**a22**	**c11**	**c21**	**c22**	**e11**	**e21**	**e22**
Age 12	weekday	0.42	-0.05	0.32	-0.35	0.20	0.46	**0.84**	-0.02	**0.80**
	weekend	0.20	0.14	0.28	-0.29	0.51	0.00	**0.94**	-0.00	**0.80**
Age 17	weekday	**0.55**	-0.04	0.74	-0.01	0.00	0.00	**0.83**	**-0.10**	**0.66**
	weekend	**0.62**	-0.17	**0.73**	0.00	0.00	0.00	**0.79**	-0.01	**0.67**
	typical	**0.56**	-0.13	**0.73**	-0.18	-0.03	0.00	**0.77**	-0.05	**0.67**
	last night	**0.62**	-0.18	0.72	-0.08	0.01	0.00	**0.83**	-0.06	**0.67**
Age 21	weekday	**0.35**	-0.52	0.06	0.17	0.26	0.00	**0.92**	**-0.13**	**0.80**
	weekend	**0.49**	-0.14	0.53	0.12	0.21	0.00	**0.86**	**-0.18**	**0.79**
Age 23	weekday	**0.61**	0.01	**0.59**	-0.22	0.03	-0.00	**0.76**	-0.08	**0.80**
	weekend	**0.57**	-0.12	0.59	0.00	0.00	-0.00	**0.82**	0.01	**0.80**

Note. Bivariate Cholesky decomposition estimates between depression symptoms (CES-D score) and linear sleep duration. Bold = p<.05 as determined by chi-square differences tests. a11 paths = variation unique to linear sleep, a21 paths = proportion of shared variation between sleep and depression; a22 paths = variation unique to depression. c11, c21, and c22 = shared environmental loading. e11, e21, e22 = non-shared environmental loadings. Model fit indices ranged from χ^2= 8.99(17), p=.941 to χ^2= 23.19(17), p=.143; CFI .951 to 1.000; RMSEA .000 to .043.

Table 5.7 Bivariate Cholesky decompositions between depression symptoms and

categorical sleep

Cat Sleep & Dep		a11	a21	a22	c11	c21	c22	e11	e21	e22
Age 12	weekday	0.0	0.00	0.00	**0.58**	0.05	0.55	**0.82**	-0.03	**0.70**
	weekend	0.16	0.17	0.01	**-0.56**	-0.01	0.39	**0.81**	0.06	**0.82**
Age 17	weekday	**0.63**	0.51	0.01	-0.31	0.49	0	**0.72**	0.13	**0.49**
	weekend	**0.69**	-.011	0.56	0.18	0.49	-0.00	**0.71**	0.13	**0.43**
	typical	**0.65**	0.43	0.01	-0.27	0.56	0.00	**0.71**	0.12	**0.49**
	last night	**0.71**	0.07	**0.57**	-0.03	0.50	0.02	**0.7**	0.13	**0.41**
Age 21	weekday	**0.62**	0.34	0.22	0.00	0.00	0.00	**0.79**	0.13	**0.82**
	weekend	**0.62**	0.11	**0.62**	0.00	0.00	0.00	**0.79**	**0.26**	**0.53**
Age 23	weekday	0.60	0.22	0.62	-0.00	0.00	0.00	0.80	-0.05	**0.57**
	weekend	**0.60**	0.38	0.59	0.00	0.00	0.00	**0.80**	-0.07	**0.50**

Note. Bivariate Cholesky decomposition estimates between depression symptoms (CES-D score) and categorical sleep duration. Bold = $p<.05$ as determined by chi-square differences tests Bold = $p<.05$ as determined by chi-square differences tests. a11 paths = variation unique to depression, a21 paths = proportion of shared variation between categorical sleep and depression; a22 paths = variation unique to categorical sleep. c11, c21, and c22 = shared environmental loading. e11, e21, e22 = non-shared environmental loadings. Model fit indices ranged from χ^2= 10.03(14), p=.760 to RMSEA:.000-.032; CFI: .961-1.000; Four models had less than optimal model fit: last night 17: χ^2= 23.70(14), p=.050, RMSEA = .059, CFI=.915; typical 17: χ^2= 25.39(14), p=.030, RMSEA=.064; CFI=.897; weekday 17: χ^2= 26.28(14), p=.024, RMSEA=.066; CFI=.892; weekend 17: 25.46(14), p=.030, RMSEA=.064, CFI=.872.

While depression and sleep duration often show a significant phenotypic relationship, there was no evidence for shared genetic variation. However, there seemed to be substantial unique genetic and non-shared environmental influences for both depression and sleep duration. Likely the phenotypic associations seen between sleep and depression can be explained by a combination of genetic and environmental effects. Bivariate Cholesky decompositions were also run between MDD and sleep duration, however, no significant relationships between sleep and MDD emerged.

Aim 3 – Do EFs and sleep duration share a genetic etiology?

The prefrontal cortex and executive functions develop substantially from adolescence to well into adulthood. During that same period, individuals undergo

changes in circadian phase and sleep pressure as a result of hormonal influences (Hagenauer & Lee, 2013). Genetic variation related to these hormonal changes could contribute to changes in both sleep duration and EFs. Alternatively, the changes in environment as a result of the changes in sleep patterns could be then causing the relationship between sleep duration and EFs. A series of bivariate Cholesky decomposition were run to see if shared or unique genetic or environmental influences contributed to sleep duration and EFS. However, given small to modest phenotypic relationships between EFs and sleep in the LTS sample, I only ran analyses for relationships that were significant in the phenotypic within-wave regressions. Cholesky decompositions between linear sleep duration and EFs are presented below in Table 5.8.

Table 5.8 Bivariate Cholesky Decomposition of EFs and Linear Sleep Duration

EF	Sleep Variable	a11	a21	a22	c11	c21	c22	e11	e21	e22
					EFs 17					
UPD	weekday 12	**1.00**	-0.10	0.32	0.00	-	0.00	-0.09	**0.75**	-0.01
UPD	weekday 21	**1.00**	0.10	0.26	0.00	-	0.00	0.04	**0.89**	0.00
					EFs 23					
UPD	weekend 21	**0.99**	**0.19**	0.00	0.00	-	0.00	-0.14	0.52	0.76

Note: Bivariate Cholesky decompositions of EFs and linear sleep duration. All EFs were modeled simultaneously. UPD = Updating-specific abilities. a11 is the genetic loading unique to UPD. a21 represent the shared genetic variance between UPD and linear sleep. a22 represents the genetic variation unique to sleep. c11, c21, and c22 = shared environmental loading. e11, e21, e22 = non-shared environmental loadings. c21 was dropped from models to allow for convergence. Model fit indices ranged from: χ^2=461.468(400), p=.018 to χ^2=526.859(399), p=.018; RMSEA: 0.024 to .0033; CFI: 0.951 to 0.970.

Overall, the genetic contributions to Updating-specific were not shared to significant extent with linear sleep duration. As expected, shared environmental influences did not contribute in any meaningful way in these phenotypes. Surprisingly, there were no unique non-shared environmental contributions to either EFs or sleep duration, but the non-shared environmental influences were significant. However, the

non-shared environmental estimate for the Updating-specific latent factor is essentially zero, and while the model tries to estimate the proportion of non-shared environmental variance shared between sleep duration and Updating-specific abilities, there is nothing to really be shared. Therefore the E for sleep duration can go into e21 or e22, and so the e21 paths are really reflecting unique environmental variation for sleep duration. Due to model convergence issues, only analyses with linear sleep were presented.

General Discussion

Although previous studies have found mixed results for heritability estimates of sleep duration, I expected sleep duration to be heritable in our sample at least in adulthood. While genetic variation probably contributes to sleep in early adolescence, I found that the proportion of variance explained by genetic effects becomes statistically significant in late adolescence, around age 17. Linear and categorical sleep showed similar heritability patterns. Separate heritabilities were estimated for weekday and weekend sleep, with weekend sleep being more frequently heritable across ages. There are multiple possibilities for why this pattern of results is observed. New genetic factors could be coming online in late adolescence and early adulthood. Alternatively, flexibility in determining one's own sleep schedule could allow genetic influence to be detected. Given that weekend sleep was more frequently heritable than weekday sleep and that sleep schedule can be more flexible on weekends, these results suggested that the ability to set one's own sleep schedule might allow for genetically influenced biological processes to have a stronger impact on sleep duration.

In addition, when looking for genetic and environmental contributions shared between earlier sleep duration and later sleep duration, only evidence for shared genetic

variation existed in young adulthood (age 21 to age 23). At younger ages, environmental factors seem to be more important for phenotypic similarities between earlier sleep and later sleep. Again, this pattern could be observed if more controlled environments, such as imposed bed and wake times, mask genetic effects, or if new genetic effects come online at age 21 that are at least somewhat stable until age 23. To disentangle these two possibilities more research will need to be done with more detailed measures.

Weekday and weekend sleep showed some differences in heritability patterns, yet I expected the same genetic variation to be contributing to both weekday and weekend sleep. After all, while sleep patterns might differ between weekdays and weekends, biological processes such as sleep pressure and circadian rhythm do not care what day it is. Indeed, in young adulthood (ages 21 and 23), a substantial proportion of genetic variation that contributes to weekday sleep also contributes to weekend sleep. While not statistically significant (most were marginally significant), the estimates for unique contributions to weekend sleep were larger than expected, and might be significant if shared environmental estimates were dropped from the models. Genetic influences attributable to some other phenotype (such as activity level or sociability) might be contributing to weekend sleep more than weekday sleep. Non-shared environmental influences play a large role in weekday and weekend sleep. Both unique and common non-shared environments contribute to weekday and weekend sleep duration. More detailed sleep studies would be necessary to identify which environments are shared between weekday and weekend sleep and which are unique to each.

Since categorical sleep captures the extremes of the same continuum as linear sleep duration, I expected a substantial proportion of genetic variation to be shared

between linear and categorical sleep. However some different phenotypic relationships emerged between linear and categorical sleep with depression and EFs, hinting that there might not be as much genetic overlap as anticipated. In line with the phenotypic results, there was little evidence of the same genetic variation contributing to both linear and categorical sleep. The majority of the time there were also significant non-shared environmental contributions to both linear and categorical sleep, with a significant proportion of that variance being shared between them at ages 17 and 23. The strong phenotypic relationship between linear and categorical is mostly explained by non-shared environmental factors. Intriguingly there are substantial non-shared (and even one shared) environmental factors that distinguish categorical sleep from linear sleep. Perhaps specific environments are influencing people who would normally have short sleep duration to have a very short sleep duration. Further research is necessary to identify what those environments might be. Regardless, the differences indicate that it would be important to include both linear sleep duration and categorical sleep duration in future analyses.

Next I examined whether there were unique or shared genetic (or environmental) influences with depression and sleep duration. The same overall pattern was observed for depression symptoms with both linear and categorical sleep; there were unique genetic contributions to each, and unique non-shared environmental contributions to each, but no shared genetic variation. There was some evidence for common non-shared environments contributing to both depression symptoms and linear sleep duration at ages 17 and 21. While a bidirectional relationship between sleep and depression exists, it does not appear to be due to a shared genetic etiology in young adolescence and early adulthood.

Lastly, the genetic and environmental etiology of EFs and sleep duration were examined for significant phenotypic results from the within-wave regressions. Given that the phenotypic relationships were relatively weak, I did not have grand expectations for shared genetic influences. However, given the high heritability of EFs at the latent level, some portion of the genetic influences that contribute to EFs could also contribute to sleep duration. Indeed, shared genetic influences contributed to both Updating-specific abilities at age 23 and age 21 weekday sleep duration, but not with any of the other time points. Interestingly, there were not significant unique non-shared environmental contributions to either trait, but significant proportions of non-shared environmental influences contributed to Updating-specific abilities and sleep durations.

Chapter 6

General Discussion

This dissertation presented four studies to further examine EFs and their relationship with psychopathology and sleep duration. Executive dysfunction is so commonly associated with psychopathology and problematic behaviors that it is often suggested as an underlying factor and a potential endophenotype. However, most studies that examine EF as an endophenotype, use impure measures of one EF with task specific components or a measure that conflates multiple types of EFs. Therefore, I wanted to examine EFs as a potential endophenotype using the Unity and Diversity model of EFs, in which the latent EF variables have been shown to be highly heritable, stable, more pure measures of EFs.

First, I tried to use genetic risk for psychopathology to predict EFs (Chapter 2) – if EFs are an endophenotype for psychopathology, some portion of the genetic variation that contributes to psychopathology should also contribute to EFs. Second, I began to develop a functional neural network for Common EFs based on the Unity and Diversity model of EFs to be tested in the future as a potential endophenotype (Chapter 3). Third, I further characterized the phenotypic relationships between individual differences in sleep duration, depression, and EFs (Chapter 4). Last, I examined the genetic and environmental etiology of sleep duration in adolescence and young adulthood and asked to what degree genetic and environmental factors were shared with EFs and depression. A shared genetic etiology between EF and sleep duration, and sleep duration and depression, would suggest the possibility for intermediate phenotype relationships between the variables.

Major Contributions of the Studies

EFs are cognitive abilities that allow individuals to achieve goals and navigate novel situations, and underlie everyday behaviors. When EFs are disrupted, it allows for problematic behaviors to occur. Behaviors are considered problematic when they are inconsistent with or undermine higher-level goals. And different collections of problematic behaviors (or symptoms) get classified as different forms of psychopathology. While there are biological differences between disorders, they are initially diagnosed from clusters of symptoms. Rumination is a symptom of depression, and can stem from the inability to inhibit a thought, to clear it from working memory, or successfully switch to a different line of thinking. For reasons such as this, it is easy to see why EFs might be proposed as an endophenotype. The next step is to test how well EFs fit the criteria of an endophenotype for various disorders. While none of these studies were designed to explicitly test whether or not EFs are an endophenotype, the results have important implications and take steps towards furthering research on this topic.

The first study does not rule out the possibility of EFs as an endophenotype for the five forms of psychopathology examined, but it does tell us that larger sample sizes are necessary when using polygenic risk scores. Despite deep phenotyping in the testing sample, polygenic risk for psychopathology did not significantly predict EFs after controlling for multiple testing. This was in part due to the fact that the Unity and Diversity model of EF, as currently measured (lab-based cognitive testing), did not produce larger effect sizes than psychopathology diagnosis or symptoms themselves. If they had, then a smaller sample size might not have been an issue. But given the observed effect sizes, we had low power from insufficient sample sizes. While EFs might be more

proximal to genetic effects than symptoms of psychopathology, EFs are a complex phenotype in and of themselves. So while EFs are highly related to psychopathology and are arguable more proximal to genetic effects, perhaps less complex phenotypes would be better candidates as endophenotypes. Alternatively, neural activation evoked by EFs might be more proximal to genetic variation compared to accuracy and reaction time associated with EFs and therefore a better potential endophenotype.

The second study started developing a functional neural network based on the Unity and Diversity model of EF, in part for future use as a potential endophenotype. Neural activation is arguably more proximal to genetic effects than reaction time and accuracy measures. I examined whether the overlapping neural activation from three of the EF tasks in the unity and diversity model battery produced results that would be interchangeable with previously identified frontoparietal networks based on other cognitive tasks. While there was a substantial amount of overlap, there were also a number of differences, which makes using these other frontoparietal networks in place of this EF network unadvisable. In addition, some areas important for individual differences in Common EF fell outside of the network identified at the group level as necessary for completion of the tasks. This suggests that when using a neural network as an endophenotype, it should not be restricted to just those areas identified at the group level. Areas outside of the group level, which are important for individual differences, should also be examined.

The third study examined the relationships between sleep duration and EFs and sleep duration and depression at the phenotypic level in our sample from adolescence through young adulthood. When significant, associations were in the expected directions:

Shorter sleep duration and sleeping more or less than typical were associated with more depression and worse Updating-specific abilities at later ages. Longitudinally, sleep duration seems to have a bidirectional relationship with EFs and depression; sometimes earlier sleep predicts later EF (or depression), and other times earlier EF (or depression) predicts later sleep duration. This indicates that sleep duration is highly interconnected with these two phenotypes over time. When these three phenotypes are put into one model (within a given time point), it seems as if depression mediates the relationship between sleep duration and EFs. However, the relationship between sleep duration and EFs is usually small to begin with, and while significant, the indirect effects of sleep on EFs through depression do not often change that relationship in a meaningful way. While these three phenotypes all influence each other, it does not seem as if any set of relationships is due to the interference of the third at this time in development. The series of analyses in study 3 set the groundwork for the fourth study.

The final study in this dissertation assessed the genetic and environmental etiology of both linear and categorical sleep duration. Sleep duration showed significant genetic and non-shared environmental effects in late adolescence and young adulthood. I further characterized genetic and environmental contributions of sleep duration. While there was some evidence that a portion those genetic effects were stable in young adulthood, across weekday and weekend sleep duration in young adulthood, and between linear and categorical sleep, it was not as to the extent I had predicted. More often common non-shared environments contributed to the different types of sleep duration. Next I asked to what degree were those genetic and environmental influences shared with depression and EFs. A substantial proportion of shared genetic variation between

phenotypes would suggest the possibility of sleep duration as an endophenotype for EF or depression. However, there was little evidence for shared genetic contributions, and instead more evidence for unique genetic contributions to each phenotype. Interestingly, unique non-shared environments were important for sleep duration and depression, but common non-shared environments were important for sleep duration and Updating-specific abilities. This implies that while sleep duration has bidirectional relationships with EFs and depression, it probably will not meet criteria for an endophenotype for either phenotype if more thoroughly tested. However, future studies should look to disentangle the environmental influences that contribute to these phenotypic relationships.

Concluding remarks

In conclusion, this body of work suggests that while there is a significant relationship between EFs, more research needs to be conducted to determine if EFs are indeed an endophenotype that will aid in the identification of genetic variation contributing to problematic behaviors and diseases. Even if it turns out that EFs are not ideal endophenotypes, important information can still be gleaned from better understanding how they relate to other behaviors. For example, we now know that at a particular age in young adulthood, how much sleep an individual gets is most strongly related to Updating-specific abilities, but atypical amounts of sleep (less than 7 hours or more than 9) predict better Shifting-specific abilities. Since it is still unclear how to best train or elevate EF abilities and whether training effects generalize and stick around, it might be easier to alter sleep patterns. Budgeting enough time for sleep and addressing any sleep issues might help to normalize EF abilities during this particularly critical time

for the development of EFs if the effect of sleep on EFs is causal. Importantly, adequate sleep seems to have implications for later EF abilities, making it even more important to stress the benefits of sufficient sleep. So while sleep duration might not be a good candidate endophenotype for EFs, and EFs might not be a practical endophenotype for psychopathology, important information about the nature of relationships and identification of potential mechanisms can still gained from these types of studies.

References Cited

Achenbach, T.M. (1991) Manual for the Child Behavior Checklist 4-18 and 1991 Profile. University of Vermont Department of Psychiatry, Burlington, VT

Amann, B., Gomar, J.J., Ortiz-Gil, J., McKenna, P., Sans-Sansa, B., Sarró, S., … Pomarol-Clotet, E. (2012) Executive dysfunction and memory impairment in schizoaffective disorder: A comparison with bipolar disorder, schizophrenia and healthy controls. *Psychological Medicine* 42:2127–2135

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders:* Fifth Edition. Arlington, VA, American Psychiatric Association.

Anderson, B., Storfer-Isser, A., Taylor, H. G., Rosen, C. L., & Redline, S. (2009). Associations of Executive Function With Sleepiness and Sleep Duration in Adolescents. *Pediatrics*, *123*(4), e701–e707. http://doi.org/10.1542/peds.2008-1182

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., … Riemann, D. (2011). Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders*, *135*(1–3), 10–19. http://doi.org/10.1016/j.jad.2011.01.011

Barclay, N. L. N., Eley, T. C. T., Buysse, D. D. J., Rijsdijk, F. V, & Gregory, A. M. (2010). Genetic and environmental influences on different components of the Pittsburgh Sleep Quality Index and their overlap. *Sleep*, *33*(5), 659–68.

Barclay, N. L., & Gregory, A. M. (2013). Quantitative genetic research on sleep: A review of normal sleep, sleep disturbances and associated emotional, behavioural, and health-related difficulties. *Sleep Medicine Reviews*, *17*(1), 29–40. http://doi.org/10.1016/j.smrv.2012.01.008

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-Related Changes in Heritability of Behavioral Phenotypes Over Adolescence and Young Adulthood: A Meta-Analysis. *Twin Res Hum Genet*, *10*(3), 423–433. http://doi.org/10.1375/twin.10.3.423

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*(3), 180–200. http://doi.org/10.1016/j.dr.2009.05.002

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). NIH Public Access. *Learning and Individual Differences*, *21*(4), 327–336.

http://doi.org/10.1016/j.lindif.2011.01.007.Relations

Bishop, S. J., Fossella, J., Croucher, C. J., & Duncan, J. (2008). COMT val158met genotype affects recruitment of neural mechanisms supporting fluid intelligence. *Cerebral Cortex*, *18*(9), 2132–2140.

Bollen, K.A. (1989) Structural equations with latent variables. New York, NY: Wiley

Boonstra, T. W., Stins, J. F., Daffertshofer, A., & Beek, P. J. (2007). Effects of sleep deprivation on neural functioning: An integrative review. *Cellular and Molecular Life Sciences*, *64*(7–8), 934–946. http://doi.org/10.1007/s00018-007-6457-8

Breslau, N., Roth, T., Rosenthal, L., & Andreski, P. (1996). Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young adults. *Biological Psychiatry*, *39*(6), 411–418. http://doi.org/10.1016/0006-3223(95)00188-3

Browning, B., Yu, Z. (2009) Simultaneous genotype calling and haplotype phase inference improves genotype accuracy and reduces false positive associations for genome-wide association studies. *The American Journal of Human Genetics* 85:847-861

Burdick, K.E., Gunawardane, N., Woodberry, K., Malhotra, A.K. (2009) The role of general intelligence as an intermediate phenotype for neuropsychiatric disorders. *Cognitive Neuropsychiatry* 12: 299-311

Camargo, A., Azuaje, F., Wang, H., Zheng, H. (2008) Permutation-based statistical tests for multiple hypotheses. *Source Code for Biology and Medicine* 3(15)

Cannon, T.D., Keller, M.C. (2006) Endophenotypes in the genetic analyses of mental disorders. *Annual Review of Clinical Psychology* 2:267–290

Carskadon, M. A., Acebo, C., & Jenni, O. G. (2004). Regulation of adolescent sleep: Implications for behavior. *Annals of the New York Academy of Sciences*, *1021*, 276–291. http://doi.org/10.1196/annals.1308.032

Chang, P. P., Ford, D. E., Mead, L. A., Cooper-Patrick, L., & Klang, M. J. (1997). Insomnia in Young Men and Subsequent Depression. *American Journal of Epidemiology*, *146*(2), 459–469. http://doi.org/10.1016/j.amjmed.2014.11.036.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*(2), 257–271.

Collette, F., Van Der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., &

Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Human Brain Mapping*, *25*(4), 409–423.

Colom, R., Privado, J., García, L. F., Estrada, E., Cuevas, L., & Shih, P. C. (2015). Fluid intelligence and working memory capacity: Is the time for working on intelligence problems relevant for explaining their large relationship? *Personality and Individual Differences*, *79*, 75–80.

Crittenden, B. M., & Duncan, J. (2014). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cerebral Cortex*, *24*(2), 532–540.

Dahl, R. E. (1996). The impact of inadequate sleep on children's daytime cognitive function. *Seminars in Pediatric Neurology*, *3*(1), 44–50. http://doi.org/10.1016/S1071-9091(96)80028-3

Davis, O.S.P., Butcher, L.M., Docherty, S.J., Meaburn, E.L., Curtis, C.J.C., Simpson, M.A., … Plomin, R. (2010) A three-stage genome-wide association study of general cognitive ability: Hunting the small effects. *Behavior Genetics* 40:759–767

Delaneau, O., Marchini, J., Zagury, J.F. (2011) A linear complexity phasing method for thousands of genomes. *Nature Methods* 9:179–181

Derringer, J., Corley, R.P., Haberstick, B.C., Young, S.E., Demmitt, B.A., Howrigan, D.P., … McQueen, M.B. (2015) Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behavior Genetics* 45:375–381

Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Medicine Reviews*, *14*(3), 179–189. http://doi.org/10.1016/j.smrv.2009.10.004

Dick, D.M., Agrawal, A., Keller, M.C., Adkins, A., Aliev, F., Monroe, S., … Sher, K.J. (2015) Candidate gene-environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science* 10:37–59

Drake, C. L., Friedman, N. P., Wright, K. P., & Roth, T. (2011). Sleep reactivity and insomnia: genetic and environmental influences. *Sleep*, *34*, 1179–1188. http://doi.org/10.5665/SLEEP.1234

Drake, C., Richardson, G., Roehrs, T., Scofield, H., & Roth, T. (2004). Vulnerability to stress-related sleep disturbance and hyperarousal. *Sleep*, *27*(2), 285–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15124724

Dudbridge, F. (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genetics* 9:e1003348

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 145–199.

Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-drob, E. M. (2016). Strong Genetic Overlap Between Executive Functions and Intelligence. *Journal of Experimental Psychology: General*, *145*(9), 1141–1159.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., & Sklar, P. (2005) Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry* 57:1313–1323

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16616–21.

Flint, J., Munafò, M.R. (2007) The endophenotype concept in psychiatric genetics. *Psychological Medicine* 37:163–180

Friedman, N.P., Haberstick, B.C., Willcutt, E.G., Miyake, A., Young, S.E., Corley, R.P., Hewitt, J.K. (2007) Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science* 18:893–900

Friedman, N. P., & Miyake, A. (2016). Unity and Diversity of Executive Functions: Individual Differences as a Window on Cognitive Structure. *Cortex*, 1–19.

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340.

Friedman, N.P., Miyake, A., Corley, R.P., Young, S.E., Defries, J.C., Hewitt, J.K. (2006) Not all executive functions are related to intelligence. *Psychological Science* 17:172–179

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

Friedman, N.P., Miyake, A., Young, S.E., Defries, J.C., Corley, R.P., Hewitt, J.K. (2008) Individual differences in executive function are almost entirely genetic in origin. *Journal of Experimental Psychology: General* 137:201–225

Glahn, D.C., Bearden, C.E., Niendam, T.A., Escamilla, M.A. (2004) The feasibility of neuropsychological endophenotypes in the search for genes associated with bipolar affective disorder. *Bipolar Disorders* 6:171–182

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

Gregory, A. M., Rijsdijk, F. V, Dahl, R. E., McGuffin, P., & Eley, T. C. (2006). Associations between sleep problems, anxiety, and depression in twins at 8 years of age. *Pediatrics*, *118*(3), 1124–1132. http://doi.org/10.1542/peds.2005-3118

Gustavson, D. E., & Miyake, A. (2015). Trait worry is associated with difficulties in working memory updating. *Cognition & Emotion*, *9931*(October), 1–15. http://doi.org/10.1080/02699931.2015.1060194

Hagenaars, S.P., Harris, S.E., Davies, G., Hill, W.D., Liewald, D.C.M., Ritchie, S.J., … Deary, I.J. (2016) Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112151) and 24 GWAS consortia. *Molecular Psychiatry* advanced online publication

Hagenauer, M. H., Perryman, J. I., Lee, T. M., & Carskadon, M. A. (2009). Adolescent changes in the homeostatic and circadian regulation of sleep. *Developmental Neuroscience*, *31*(4), 276–284. http://doi.org/10.1159/000216538

Hasler, G., Drevets, W.C., Manji, H.K., Charney, D.S. (2004) Discovering endophenotypes for major depression. *Neuropsychopharmacology* 29:1765–1781

Herd, S.A., O'Reilly, R.C., Hazy, T.E., Chatham, C.H., Brant, A.M., Friedman, N.P. (2014) A neural network model of individual differences in task switching abilities. *Neuropsychologia* 62:375-389

Howie, B.N., Donnelly, P., Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics* 5:e1000529

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. http://doi.org/10.1037//1082-989x.3.4.424

Ibrahim-Verbaas, C.A., Bressler, J., Debette, S., Schuur, M., Smith, A.V., Bis, J.C., … Mosley, T.H. (2016) GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Molecular Psychiatry* 21:189–197

Irwin, M. R., Cole, J. C., & Nicassio, P. M. (2006). Comparative meta-analysis of behavioral interventions for insomnia and their efficacy in middle-aged adults and in older adults 55+years of age. *Health Psychology*, *25*(1), 3–14. http://doi.org/10.1037/0278-6133.25.1.3

Jessor, R., & Jessor, S. L. (1977). Problem behavior and psychosocial development: A longitudinal study of youth.

Johnson, D. P., Whisman, M. A., Corley, R. P., Hewitt, J. K., & Friedman, N. P. (2014). Genetic and environmental influences on rumination and its covariation with depression. *Cognition & Emotion*, *28*(7), 1270–1286. http://doi.org/10.1080/02699931.2014.881325

Kendler, K.S., Neale, M.C. (2010) Endophentype: A conceptual analysis. *Molecular Psychiatry* 15:789–797

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., Eaves, L.J. (1992) Major depression and generalized anxiety disorder: Same genes, (partly) different environments? *Archives of General Psychiatry* 49:716–722

Knutson, K. L. (2005). Sex differences in the association between sleep and body mass index in adolescents. *Journal of Pediatrics*, *147*(6), 830–834. http://doi.org/10.1016/j.jpeds.2005.07.019

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.B., Rimfeld, K., von Stumm, S., Dale, P.S., Breen, G., O'Reilly, P.F., Plomin, R. (2015) Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry* 1-6

Lencz, T., Knowles, E., Davies, G., Guha, S., Liewald, D.C., Starr, J.M., … Malhotra, A.K. (2014) Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: A report from the Cognitive Genomics consorTium (COGENT). *Molecular Psychiatry* 19:168–174

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis. *Neuropsychology*, *24*(1), 9–34. http://doi.org/10.1037/a0017336

McIntosh, A.M., Gow, A., Luciano, M., Davies, G., Liewald, D.C., Harris, S.E., … Deary I.J. (2013) Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biological Psychiatry* 73:938–943

Meltzer, L. J., & Mindell, J. A. (2014). Systematic Review and Meta-Analysis of Behavioral Interventions for Pediatric Insomnia. *Journal of Pediatric Psychology*, *39*(8), 932–948. http://doi.org/10.1093/jpepsy/jsu161

Miles, J.H. (2011) Autism spectrum disorders: A genetics review. *Genetics in Medicine* 13:278–294

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children. *Journal of Cognition and Development*, *13*(3), 395–423. http://doi.org/10.1080/15248372.2011.585478

Miyake, A., Friedman, N.P. (2012) The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science* 21:8–14

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology* 41: 49-100

Morin, C. M., Culbert, J. P., & Schwartz, S. M. (1994). Nonpharmacological interventions for insomnia: a meta-analysis of treatment efficacy. *The American Journal of Psychiatry*, *151*(8), 1172–1180. http://doi.org/10.1176/ajp.151.8.1172

Morrison, A.C., Bare, L.A., Chambless, L.E., Ellis, S.G., Malloy, M., Kane, J.P., … Boerwinkle, E. (2007) Prediction of coronary heart disease risk using a genetic risk score: The atherosclerosis risk in communities study. *American Journal of Epidemiology* 166:28–35

Muthén, L.K. , Muthén, B.O. (1998-2012) Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

Naismith, S. L., Norrie, L., Lewis, S. J., Rogers, N. L., Scott, E. M., & Hickie, I. B. (2009). Does sleep disturbance mediate neuropsychological functioning in older people with depression? *Journal of Affective Disorders*, *116*(1–2), 139–143. http://doi.org/10.1016/j.jad.2008.11.017

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., … Nelson, S. (2010) Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 49:884–897

Nyden, A., Hagberg, B., Gousse, V., Rastam, M. (2011) A cognitive endophenotype of autism in families with multiple incidence. *Research in Autism Spectrum Disorder* 5:191–200

Park, S., Cho, M. J. E., Chang, S. M., Bae, J. N., Jeon, H. J., Cho, S. J., … Hong, J. P. (2010). Relationships of sleep duration with sociodemographic and health-related

factors, psychiatric disorders and sleep disturbances in a community sample of Korean adults. *Journal of Sleep Research*, *19*(4), 567–577. http://doi.org/10.1111/j.1365-2869.2010.00841.x

Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P., & Langinvainio, H. (1983). Genetic and environmental determination of human sleep. *Sleep*, *6*(3), 179–185.

Plomin, R., Haworth, C.M.A., Meaburn, E.L., Price, T.S., Davis O.S.P. (2013) Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science* 24:562–568

The Psychiatric GWAS Consortium Steering Committee (2009) A framework for interpreting genome-wide association studies of psychiatric disorders. *Molecular Psychiatry* 14:10–17

Purcell, S., Neale, B.M., Todd-brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., … Sham, P.C. (2007) REPORT PLINK : A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81:559–575

Purkayastha, S., Wager, T. D., & Nichols, T. E. (2008). Inferring individual differences in fMRI: Finding brain regions with significant within subject correlation. *Statistica Sinica*, *18*, 1483–1500.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Radloff, L.S. (1977) The CES-D scale : A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1:385–401

Raniti, M. B., Allen, N. B., Schwartz, O., Waloszek, J. M., Byrne, M. L., Woods, M. J., … Trinder, J. (2016). Sleep Duration and Sleep Quality: Associations with Depressive Symptoms Across Adolescence. *Behavioral Sleep Medicine*, *2002*(August), 1–18. http://doi.org/10.1080/15402002.2015.1120198

Rhea, S.A., Gross, A.A., Haberstick, B.C., Corley, R.P. (2013) Colorado Twin Registry – An Update. *Twin Research and Human Genetics* 16:351-357

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Common and Specific Aspects of Executive Function. *NeuroImage*, *1*(104), 69–78. http://doi.org/10.1016/j.neuroimage.2014.09.045

Rietveld, C.A., Esko T., Davies, G., Pers, T.H., Turley, P., Benyamin, B., … Philipp, D. (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences* 111:13790–13794

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.H., Holmans, P.A., … O'Donovan, M.C. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427

Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., ... Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry* 18:497-511

Roberts, R. E., & Duong, H. T. (2014). The prospective association between sleep deprivation and depression among adolescents. *Sleep*, *37*(2), 239–44. http://doi.org/10.5665/sleep.3388

Robins, L.N., Cottler, L.B., Bucholz, K.K., Compton, W.M., North, C.S., Rourke, K.M. (2000) Diagnostic interview schedule for the DSM-IV (DIS-IV). St Louis, MO: Washington University School of Medicine

Rose, E.J., Donohoe, G. (2013) Brain vs behavior: An effect size comparison of neuroimaging and cognitive studies of genetic risk for schizophrenia. *Schizophrenia Bulletin* 39:518–526

Rosenthal, M., Wallace, G.L., Lawson, R., Wills, M.C., Dixon, E., Yerys, B.E., Kenworthy, L. (2013) Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology* 27:13–18

Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., Schwab-Stone, M.E. (2000) NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry* 39: 28-38

Shih, R.A., Belmonte, P.L., Zandi, P.P. (2004) A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *International Review of Psychiatry* 16:260–283

Sklar, P., Ripke, S., Scott, L.J,. Andreassen, O.A., Chichon, S., Craddock, N., … Corvin A. (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics* 43:977–983

Sletten, T. L., Rajaratnam, S. M. W., Wright, M. J., Zhu, G., Naismith, S., Martin, N. G., & Hickie, I. (2013). Genetic and environmental contributions to sleep-wake behavior in 12-year-old twins. *Sleep*, *36*(11), 1715–22. http://doi.org/10.5665/sleep.3136

Smoller, J.W., Finn, C.T. (2003) Family, twin, and adoption studies of bipolar disorder.

*American Journal of Medical Genetic*s 123C:48–58

Snitz, B.E., MacDonald III, A.W., Carter, S.C. (2005) Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: A meta-analytic review of putative endophenotypes. *Schizophrenia Bulletin* 32:179–194

Snyder, H.R. (2013) Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin* 139:81–132

Snyder, H.R., Miyake, A., Hankin, B.L. (2015) Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology* 6:328

Stern, Y., Habeck, C., Moeller, J., Scarmeas, N., Anderson, K. E., Hilton, H. J., … Van Heertum, R. (2005). Brain networks associated with cognitive reserve in healthy young and old adults. *Cerebral Cortex*, *15*(4), 394–402.

Stoodley, C. J., Valera, E. M., & Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *NeuroImage*, *59*(2), 1560–1570.

Sullivan, P.F. (2010) The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 68:182-186

Sullivan, P.F., Kendler, K.S., Neale, M.C. (2003) Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* 60:1187-1192

Sullivan, P.F., Neale, M.C., Kendler, K.S. (2000) Genetic epidemiology of major depression: Review and meta-analysis. American Journal of Psychiatry 157:1552–1562

Supartini, A., Honda, T., Basri, N. A., Haeuchi, Y., Chen, S., Ichimiya, A., … Kumagai, S. (2016). The Impact of Sleep Timing, Sleep Duration, and Sleep Quality on Depressive Symptoms and Suicidal Ideation amongst Japanese Freshmen: The EQUSITE Study. *Sleep Disorders*, *2016*, 1–10. http://doi.org/10.1155/2016/8737654

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

The International HapMap Consortium (2003) The International HapMap Project. Nature 426:789–796

Thorisson, G.A., Smith, A.V., Krishnan, L., Stein, L.D. (2005) The International HapMap Project Web site. *Genome Research* 1592–1593

Tracy, R.P. (2008) 'Deep phenotyping: characterizing populations in the era of genomics and systems biology. *Current Opinion in Lipidology* 19:151-157

Valiente, C., Eisenberg, N., Spinrad, T. L., Haugen, R., Thompson, M. S., & Kupfer, a. (2013). Effortful Control and Impulsivity as Concurrent and Longitudinal Predictors of Academic Achievement. *The Journal of Early Adolescence*, *33*, 946–972. http://doi.org/10.1177/0272431613477239

Van Mill, J. G., Hoogendijk, W. J. G., Vogelzangs, N., Van Dyck, R., & Penninx, B. W. J. H. (2010). Insomnia and sleep duration in a large cohort of patients with major depressive disorder and anxiety disorders. *Journal of Clinical Psychiatry*, *71*(3), 239–246. http://doi.org/10.4088/JCP.09m05218gry

Virta, J. J., Heikkilä, K., Perola, M., Koskenvuo, M., Räihä, I., Rinne, J. O., & Kaprio, J. (2013). Midlife Sleep Characteristics Associated with Late Life Cognitive Function. *Sleep*, *36*(10), 1533–1541. http://doi.org/10.5665/sleep.3052

Wager, T. D., Sylvester, C. Y. C., Lacey, S. C., Nee, D. E., Franklin, M., & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *NeuroImage*, *27*(2), 323–340.

Walters, J.T.R., Owen, M.J. (2007) Endophenotypes in psychiatric genetics. *Molecular Psychiatry* 12:886–890

Watson, N. F., Buchwald, D., Vitiello, M. V, Noonan, C., & Goldberg, J. (2010). A twin study of sleep duration and body mass index. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, *6*(1), 11–7.

Watson, N. F., Harden, K. P., Buchwald, D., Vitiello, M. V., Pack, A. I., Weigle, D. S., & Goldberg, J. (2012). Sleep duration and body mass index in twins: A gene-environment interaction. *Sleep: Journal of Sleep and Sleep Disorders Research*, *35*(5), 597–603. http://doi.org/10.5665/sleep.1810

Wechsler, D. (1997) Wechsler adult intelligence scale (3rd edition). San Antonio, TX: Psychological Corporation

Wechsler, D. (1999) Wechsler abbreviated scale of intelligence (WASI). San Antonio, TX: Psychological Corporation

Willcutt, E.G., Doyle, A.E., Nigg, J.T., Faraone, S.V., Pennington, B.F. (2005) Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry* 57:1336–1346

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-smith, I., Torralva, T., …

Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899–14902.

Yarkoni, T., & Braver, T. S. (2010). Cognitive neuroscience approaches to individual differences in working memory and executive control: Conceptual and methodological issues. *Handbook of Individual Differences in Cognition: Attention, Memory, and Executive Control.*, 87–107.

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130. http://doi.org/10.1037/a0014657

Zhai, L., Zhang, H., & Zhang, D. (2015). Sleep duration and depression among adults: A meta-analysis of prospective studies. *Depression and Anxiety*, *32*(9), 664–670. http://doi.org/10.1002/da.22386

URLs

PGC: http://www.med.unc.edu/pgc/downloads

R: http://www.R-project.org/