UNIVERSITY OF CALIFORNIA SAN DIEGO

Characterizing the Neural Basis of Individual Differences in Behaviors Using Large-scale, Population-based Neuroimaging Studies

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

in

Cognitive Science

by

Weiqi Zhao

Committee in charge:

Professor Anders Dale, Co-Chair
Professor Terry Jernigan, Co-Chair
Professor Andrea Chiba
Professor Deanna Greene
Professor Wesley Thompson

2022

The Dissertation of Weiqi Zhao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I would like to dedicate this dissertation to those who trained me. Thank you, Terry, for being a role model and for showing me the importance of work ethics and dedication as a scientist, and thank you, Anders, for showing me the importance of passion and fun in doing science and for leading me into the world of programming and MATLAB. I would like to thank Wes Thompson and Chun-Chieh Fan for showing me the world of statistical modeling and for being patient when teaching me the basics of statistics. Thank you to Deanna Greene for her mentorship and her valuable input on my thesis project. Additionally, I would like to thank my lab members, Clare Palmer, Rob Loughnan, Diliana Pecheva, Carolina Makowski, and Diana Smith for their insightful discussions and comradery.

I am grateful for having joined the Cog Sci department where I received a lot of help and support. I would like to say thank you to Andrea Chiba, Marta Kutas, and Doug Nitz for their encouragement and support when I had doubts, and to Stephan Kaufhold and Zoe Tzu-Han Cheng for their companionship and the scheduled naps in the office.

During my Ph.D. years, I am lucky to have spent time with Shuai Tang, Tricia Ngoon, Kenny Kuhn, Qi Cheng, Yuesong Shi, Bonnie Wu, Tom Yang, Navika Chan. Thank you for your friendship and for being a part of this journey.

I would like to thank my parents and my partner, Kelvin Xu, for their unwavering love and support.

Last but not least, I would like to dedicate this body of work to the doubts and struggles, and the growth and progress I made during my Ph.D.

EPIGRAPH

"Brains don't exempt you from hitting the wall, because on the scale of science itself, we are all idiots. My analogy is that humans learning science is akin to pigeons learning Calculus – it really doesn't matter if you're pigeon Einstein, the scale of the problem is negligible. Unless you've got a notebook full of solved millennium problems, a cure for cancer, and an explanation for dark matter… shut your brag hole and get back to work, pigeon."

– wise man from Quora

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in Cerebral Cortex 2021. Weiqi Zhao, Clare E. Palmer, Wesley K. Thompson, Bader Chaarani, Hugh P. Garavan, ,B.J.Casey, Terry L. Jernigan, Anders M. Dale, & Chun Chieh Fan, Cerebral Cortex, 2021. The dissertation author was the primary investigator and co-first author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Weiqi Zhao, Donald J. Hagler, Hugh P. Garavan, Deanna J. Greene, Terry L. Jernigan & Anders M. Dale. The dissertation author was the primary researcher and author of this material.

VITA

2013 Bachelor of Science in Psychology, University of Illinois Urbana-Champaign

2016 Master of Arts in Psychology, San Diego State University

2022 Doctor of Philosophy in Cognitive Science, University of California San Diego

GRANTS

KAVLI Innovative Research Grant 2021 - 2022 ($50,000), Primary Investigator: Linking Individual Differences to the Brain's Temporal Dynamics Fingerprint

PUBLICATIONS

Zhao, W., Hagler, D. J., Garavan, H. P., Greene, D. J., Jernigan, T. J., & Dale, A. M. (2022). Task fMRI can improve the detection of behaviorally relevant functional connectivity patterns. (Manuscript in preparation).

Palmer, C. E., ***Zhao, W.,*** Loughnan, R. J., Fan, C. C., Thompson, W. K., Dale, A. M., & Jernigan, T. L. (2021). Distinct Regionalization Patterns of Cortical Morphology are Associated with Cognitive Performance Across Different Domains. Cerebral Cortex. bhab054.

Loughnan, R. J., Shadrin, A. A., Frei, O., van der Mer, D., ***Zhao, W.,*** Palmer, C. E., Thompson, W. K., Makowski, C., Jernigan, T. L., Andreassen, O. A., Fan, C. C., Dale, A. D. (2021) Generalization of Cortical Multivariate Genome-Wide Associations Within and Across Samples. BioRxiv

Zhao, W., Palmer, C. E., Thompson, W. K., Chaarani, B., Garavan, H. P., Casey, B. J., ... & Fan, C. C. (2020). Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. Cerebral Cortex. bhaa290.

Palmer, C. E., ***Zhao, W.,*** Loughnan, R. J., Fan, C. C., Thompson, W. K., Dale, A. M., & Jernigan, T. L. (2021). Determining the association between regionalisation of cortical morphology and cognition in 10,145 children. BioRxiv, 816025

Keehn, R. J. J., Sanchez, S. S., Stewart, C. R., ***Zhao, W.***, Grenesko-Stevens, E. L., Keehn, B., Muller, R-A. (2017). Impaired downregulation of visual cortex during auditory processing is associated with autism symptomatology in children and adolescents with autism spectrum disorder. Autism Research, 10(1), 130-143.

ABSTRACT OF THE DISSERTATION

Characterizing the Neural Basis of Individual Differences in Behaviors Using Large-scale, Population-based Neuroimaging Studies

by

Weiqi Zhao

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2022

Professor Anders Dale, Co-Chair
Professor Terry Jernigan, Co-Chair

Much of our understanding of the neural basis of behavioral differences are attributable to studies conducted with magnetic resonance imaging (MRI). Despite its critical role in cognitive neuroscience, MRI findings lack generalizability for translational uses due to the small and homogeneous samples of traditional MRI studies. To improve on the generalizability issue, large-scale, population-based neuroimaging studies are conducted where MRI data are collected in thousands of participants that are systematically sampled to represent the general population. Now, we have a unique opportunity to harness the statistical power afforded by population-based

neuroimaging studies to characterize and quantify the behavioral relevance of MRI measures at the population level.

The goal of this work is to take advance of a population-based neurodevelopmental study, the Adolescent Brain Cognitive Development (ABCD) Study, to shed light on the optimal fMRI design and analysis pipelines for the detection of behaviorally relevant brain signals. In Chapter 2, I challenged the traditional statistical mapping approach of MRI analysis which assumed that the behavioral differences are explained by sparse, localized brain regions. I demonstrated that at a large, population level, the effects of the association between brain activations and behavioral differences are not sparse but distributed across the cortex (Chapter 2). Aggregating the small effect sizes across the whole cortex can greatly increase the magnitude of the behavioral associations detected by fMRI tasks. This finding is consistent with the observation that behavioral differences are associated with individual differences in distributed, functional brain networks whose activities are measured by functional connectivity (FC), the pairwise correlation of activation across brain regions. In Chapter 3, I carried out a systematic investigation of the optimal fMRI paradigms for the detection of behaviorally relevant FC patterns by quantifying the behavioral prediction performance of FC patterns derived from resting-state fMRI and task fMRI. Results showed that behaviorally relevant functional brain signals are better captured by task fMRI paradigms where participants are engaged in cognitive tasks that assess similar mental constructs as the behavior of interest. These results suggest that carefully designed fMRI tasks and advanced statistical methods that capture the distributed effect sizes of the brain are more useful for the study of brain-behavior relationships at the population level

Chapter 1 Overview

1.1 Introduction

Individual differences in complex behavioral traits have important implications for life outcomes such as career options, social status, and mental health outcomes. Much scientific effort and resources have been dedicated to understanding the neurological bases of complex behavioral traits. Due to its noninvasiveness and its sensitivity to different tissue types of the brain, human magnetic resonance imaging (MRI) is well-suited to understand the neurobiological processes that support cognitive processes and offers promising avenues for discovering and identifying biomarkers that are important for mental health and cognitive outcomes (Woo et al., 2017; Gratton et al., 2020; Sui et al., 2020). However, despite its prevalence in the cognitive and clinical neuroscience literature, the translational utility of MRI is still limited. No MRI biomarkers have been developed that reliably detect individual differences in cognitive and mental health outcomes, let alone risks for psychiatric disorders.

One factor that hinders the translational utility of MRI tools is the lack of replicability (Button et al., 2013) in MRI studies. A useful biomarker needs to consistently and accurately capture individual differences in cognitive process or risk for a psychiatric disorder across individuals. Findings from MRI studies, however, often fail to replicate in independent samples. The low power of small samples (Cremer, Wager, & Yarkoni, 2017) and publication bias (Button et al., 2013) may have contributed to such limited replication across MRI studies. Traditional MRI studies are conducted with small sample sizes, resulting in large standard errors and imprecise effect size estimates (Varoquaux, 2018). Because most studies are underpowered to detect the true effects, effect sizes in MRI studies that are publishable are likely to be inflated (by chance) and do not reflect the population level effects (Yarkoni & Braver, 2010). Such inflated effect sizes (Reddan, Lindquist & Wager, 2017) and publication bias have exacerbated

the replicability issue of MRI findings. As a result, despite the rapid development of MRI tools and applications, the lack of unbiased, generalizable effect size estimates of brain-behavior associations still hinders the development of MRI biomarkers.

The focus on group comparisons instead of individual differences may also have limited the translational utility of MRI. The translational utility of MRI tools relies on the precise measurement and prediction at the individual level, but MRI studies have traditionally focused on identifying and reporting group differences and not individual differences. While quantifying group differences between patients and controls is critical to the understanding of the average effect of disease status on brain structures and functions, group comparisons are not informative at identifying and predicting individual differences. The variance in individuals is up to four times larger than in groups (Fisher et al., 2018) and greater statistical power is required to detect any individual differences (Dubois & Adolphs, 2016). Therefore, significant brain regions may be informative at differentiating between case and control but would provide limited information in identifying and predicting individuals. In order to improve the translational utility of MRI findings, MRI studies need to shift from group comparisons to the identification of structural and functional brain patterns that reliably capture individual differences in cognitive and mental health outcomes. Advocating the importance of understanding the neural correlates of trait variability, Yarkoni and Braver (2010) argued for the use of (1) large sample sizes, (2) normative samples, and (3) predictive modeling techniques.

Recognizing the limitations of small sample sizes and the group comparison approach, more MRI researchers are embracing the "Big Data" era of neuroimaging by contributing to the development of large, multi-site neuroimaging studies. Large, population-based neuroimaging studies, such as the UK Biobank and the Adolescent Brain Cognitive Development (ABCD)

Study, collect longitudinal imaging and non-imaging data in thousands of participants from multiple local communities and offer unprecedented statistical power to detect MRI features that are informative of individual differences in cognitive and mental health outcomes. The availability of large MRI databases is enabling neuroimaging researchers to reliably estimate the effect size of brain-behavior associations, addressing the inflated effect size issue of traditional MRI studies.

Large population-based neuroimaging studies have advanced the field of cognitive neuroscience by revolutionizing the methodology and the characterization of brain-behavior associations. Among the many scientific advances inspired by population-based neuroimaging studies, this thesis focuses on two topics: (1) the predictive modeling framework and (2) and the characterization of the relationship between behaviors and the distributed brain functions.

The large sample size and the statistical power afforded by population-based neuroimaging studies have allowed researchers to adopt new analysis frameworks for characterizing brain-behavior relationships. Predictive modeling framework (Gabrieli, Ghosh, Whitfield-Gabrieli, 2015; Finn & Rosenberg, 2021) is one of the exciting new analysis frameworks that aims to reconcile the lack of replicability of neuroimaging findings (Varoquaux & Poldrack, 2019). Instead of mapping behavioral functions in the brain, large neuroimaging studies allow researchers to characterize the neural basis of behavioral differences at the population level and use it to predict cognitive and mental health outcomes in new, unseen individuals. Not only can the predictive modeling framework be applied to establish the generalizability of any research findings by testing and validating these findings in new dataset and participants (Rosenberg & Finn, 2022), it also allows researchers to quantify the translational utility of neuroimaging findings, which is critical for the advance of neuroimaging biomarkers

for risks for cognitive and psychiatric diseases. Fueled by population-based neuroimaging studies, researchers have able to use individual differences in brain structure and functions to predict behavioral differences such as age (Dosenbach et al., 2010), cognitive abilities (Sripada et al., 2019, Moutoussis et al., 2021, Zhang et al., 2021), and mental health outcomes (Challis et al., 2015, Kim et al., 2016, Thomas et al., 2020).

Not only do population-based neuroimaging studies enable methodological advances in neuroimaging, new scientific discoveries and observations are also being made. One observation is that behavioral functions, especially higher-order cognitive processes, are supported by distributed brain networks. In contrast to being localized within pre-defined regions of interest (ROIs) as previously reported in traditional MRI studies, behaviorally relevant brain regions and networks identified in large sample MRI studies appear to be distributed across the whole brain (Baum et al., 2020; Sripada et al., 2020) with each brain region exerting a small effect on individual variability in cognitive and mental health outcomes (Poldrack et al. 2017). In UK Biobank, individual brain regions and networks that are highly statistically significant only explain 1% of trait variability (Smith & Nichols, 2017). Distributed brain patterns that span multiple pre-defined regions and networks are better at predicting individual variability in complex behavioral traits compared to single regions of interest (Chang et al., 2015; Sripada et al., 2020). These new observations offer new opportunities and challenges to refine and further our understanding on the contribution of brain structures and functions on the individual variability in cognitive and mental health outcomes.

1.2 Aims

This thesis characterizes the utility of different fMRI measures in understanding individual differences in cognitive outcomes by taking advantage of the predictive modeling

framework, the small effect size observation evident in population-based neuroimaging studies, and the large and diverse sample of the ABCD. With the goal of estimating the generalizable association between fMRI measures and individual differences in cognitive outcomes, I focused on quantifying the predictive performance of an MRI measure on individual differences in cognitive traits and behaviors, which is operationalized as the out-of-sample variance explained by an MRI phenotype on a cognitive outcome variable estimated using cross-validations.

This thesis was divided into two steps: (1) investigating the distribution of the generalizable association between fMRI measures and behaviors and (2) quantifying the functionally relevant information in different fMRI measures. In Chapter 1 (Zhao et al., 2021), I quantified the generalizable association between MRI measures and trait variability using a novel multivariate statistical method, the Bayesian polyvertex score (PVSB), that aggregates the distributed, sub-threshold effect sizes across the whole cortex for out of sample behavioral prediction. Using the PVSB, we were able to double the out-of-sample variance explained by task-based fMRI activations on individual variability in cognitive performance in ABCD. Comparing the predictive performance of the PVSB and the traditional MRI analysis method, I was able to show that individual variability in complex behavioral traits is associated with distributed patterns of brain activation and can be better captured by multivariate methods. This study is published in Cerebral Cortex.

In Chapter 2, I focused on quantifying the generalizable behavioral association of FC measures. Different from task-evoked activations where statistical thresholding often yields sparse, localized maps of behaviorally relevant activation patterns, the FC analysis framework adopts a distributed, network approach where the connections between brain areas are explicitly modeled and investigated for their behavioral relevance. As interests grow in understanding the

behavioral associations of FC patterns, more studies have focused on characterizing the optimal fMRI paradigms to derive functionally relevant FC patterns - specifically, whether resting-state fMRI or task fMRI provides better estimates of FC signals that are relevant for behavioral differences. In Chapter 2, I conducted a systematic examination of the behavioral relevance of resting-state and task-derived FC measures with four sets of analysis. First, I compared the prediction performance of resting-state FC and FC measures derived from three fMRI tasks on two behavioral outcome measures to assess which fMRI paradigm captures more behaviorally relevant FC patterns. Then, I decomposed the FC signal of task fMRI data to understand whether the behavioral prediction effect of task-derived FC measures is driven by the task-elicited FC changes in response to cognitive demands or the task-invariant FC component. To contextualize the behavioral prediction effect of task-derived FC measures to the behavioral prediction performance of task-evoked activation patterns examined in Chapter 1, I compared and quantified the shared and unique behavioral variance explained by the task-derived FC and task-evoked activation measures. Lastly, I examined how the amount of usable data and the adjustment of sociodemographic differences across individuals might impact the magnitude of the observed behavioral prediction accuracy of FC and task-evoked activation measures. These two factors have been previously hypothesized to moderate the behavioral prediction performance of fMRI measures.

1.3 Reference

1. Baum, G.L., Cui, Z., Roalf, D.R., Ciric, R., Betzel, R.F., Larsen, B., Cieslak, M., Cook, P.A., Xia, C.H., Moore, T.M., Ruparel, K., Oathes, D.J., Alexander-Bloch, A.F., Shinohara, R.T., Raznahan, A., Gur, R.E., Gur, R.C., Bassett, D.S., Satterthwaite, T.D., 2020. Development of structure-function coupling in human brain networks during youth. Proc. Natl. Acad. Sci. U. S. A. 117, 771–778.

2. Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376.

3. Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M., 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. Neuroimage 112, 232–243.

4. Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., Wager, T.D., 2015. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. PLoS Biol. 13, e1002180.

5. Cremers, H.R., Wager, T.D., Yarkoni, T., 2017. The relation between statistical power and inference in fMRI. PLoS One 12, e0184923.

6. Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Jr., Barch, D.M., Petersen, S.E., Schlaggar, B.L., 2010. Prediction of Individual Brain Maturity Using fMRI. Science 329.

7. Dubois, J., Adolphs, R., 2016. Building a Science of Individual Differences from fMRI. Trends Cogn. Sci. 20, 425–443.

8. Finn, E.S., Rosenberg, M.D., 2021. Beyond fingerprinting: Choosing predictive connectomes over reliable connectomes. Neuroimage 239, 118254.

9. Fisher, A.J., Medaglia, J.D., Jeronimus, B.F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. Proc. Natl. Acad. Sci. U. S. A. 115, E6106–E6115.

10. Gabrieli, J.D.E., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. Neuron 85, 11–26.

11. Gratton, C., Kraus, B.T., Greene, D.J., Gordon, E.M., Laumann, T.O., Nelson, S.M., Dosenbach, N.U.F., Petersen, S.E., 2020. Defining Individual-Specific Functional Neuroanatomy for Precision Psychiatry. Biol. Psychiatry 88, 28–39.

12. Kim, J., Calhoun, V.D., Shim, E., Lee, J.-H., 2016. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. Neuroimage 124, 127–146.

13. Moutoussis, M., Garzón, B., Neufeld, S., Bach, D.R., Rigoli, F., Goodyer, I., Bullmore, E., NSPN Consortium, Guitart-Masip, M., Dolan, R.J., 2021. Decision-making ability, psychopathology, and brain connectivity. Neuron 109, 2025–2040.e7.

14. Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. 18, 115–126.

15. Reddan, M.C., Lindquist, M.A., Wager, T.D., 2017. Effect Size Estimation in Neuroimaging. JAMA Psychiatry 74, 207–208.

16. Rosenberg, M.D., Finn, E.S., 2022. How to establish robust brain-behavior relationships without thousands of individuals. Nat. Neurosci. 25, 835–837.

17. Smith, S.M., Nichols, T.E., 2018. Statistical Challenges in "Big Data" Human Neuroimaging. Neuron 97, 263–268.

18. Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., Shedden, K., 2020. Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. Hum. Brain Mapp. 41, 3186–3197.

19. Sripada, C., Rutherford, S., Angstadt, M., Thompson, W.K., Luciana, M., Weigard, A., Hyde, L.H., Heitzeg, M., 2019. Prediction of neurocognition in youth from resting state fMRI. Mol. Psychiatry.

20. Sui, J., Jiang, R., Bustillo, J., Calhoun, V., 2020. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. Biol. Psychiatry.

21. Thomas, R.M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., van Wingen, G., 2020. Classifying Autism Spectrum Disorder Using the Temporal Statistics of Resting-State Functional MRI Data With 3D Convolutional Neural Networks. Front. Psychiatry 11, 440.

22. Varoquaux, G., 2018. Cross-validation failure: Small sample sizes lead to large error bars. Neuroimage 180, 68–77.

23. Varoquaux, G., Poldrack, R.A., 2019. Predictive models avoid excessive reductionism in cognitive neuroimaging. Curr. Opin. Neurobiol. 55, 1–6.

24. Woo, C.-W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20, 365–377.

25. 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.

26. Zhang, M., Nathaniel, U., Savill, N., Smallwood, J., Jefferies, E., 2021. Intrinsic connectivity of left ventrolateral prefrontal cortex predicts individual differences in controlled semantic retrieval. Neuroimage 246, 118760.

27. Zhao, W., Palmer, C.E., Thompson, W.K., Chaarani, B., Garavan, H.P., Casey, B.J., Jernigan, T.L., Dale, A.M., Fan, C.C., 2021. Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. Cereb. Cortex 31, 1478–1488.

Chapter 2 Individual differences in cognitive performance are better predicted by global rather than localized BOLD activity patterns across the cortex

2.1 Abstract

Despite its central role in revealing the neurobiological mechanisms of behavior, neuroimaging research faces the challenge of producing reliable biomarkers for cognitive processes and clinical outcomes. Statistically significant brain regions, identified by mass univariate statistical models commonly used in neuroimaging studies, explain minimal phenotypic variation, limiting the translational utility of neuroimaging phenotypes. This is potentially due to the observation that behavioral traits are influenced by variations in neuroimaging phenotypes that are globally distributed across the cortex and are therefore not captured by thresholded, statistical parametric maps commonly reported in neuroimaging studies. Here, we developed a novel multivariate prediction method, the Bayesian polyvertex score (PVSB), that turns a unthresholded statistical parametric map into a summary score that aggregates the many but small effects across the cortex for prediction. By explicitly assuming a globally distributed effect size pattern and operating on the mass univariate summary statistics, it was able to achieve higher out-of-sample variance explained than mass univariate and popular multivariate methods while still preserving the interpretability of a generative model. Our findings suggest that the neural basis of complex behaviors may rest in the global patterning of effect size variation of neuroimaging phenotypes, rather than in localized, candidate brain regions and networks.

2.2 Introduction

Neuroimaging is central to the search for neurobiological mechanisms of cognitive processes and psychopathology. However, by far, neuroimaging studies that aim to reveal the biological correlates of phenotypic variations have limited success in the identification of reliable

biomarkers for clinical diagnoses or cognitive functions. One of many reasons for this challenge is the reliance on statistical thresholds and mass univariate statistical models for many neuroimaging studies. In traditional neuroimaging analyses, the association between an imaging phenotype and phenotypic variation is assessed with mass univariate statistical models where the associative effect is estimated independently at each measured unit of the brain data, e.g. vertex, voxel, or region of interest. The magnitude of the generalizable brain-behavior association is usually inferred from the effect size estimates of *only the most significant* vertices/ROIs using the mass univariate estimators. Such an approach assumes that the underlying true association is sparse and localized in the cortex, and hence that clusters of vertices/ROIs with minimum p-values (min-p) form the basis of generalizable signals. Although study designs and covariates are controlled for in mass univariate statistical models used to detect brain-behavior associations, the resulting regions only explain minimal variation in behavior (Stanfield et al. 2008; Poldrack et al. 2017). With a sample size of more than 14,000 participants, Smith and Nichols (Smith and Nichols 2017) demonstrated that a statistically significant imaging composite measure, surviving Bonferroni correction of 14 million tests, explained less than 1% of the variance in behavior.

In reaction to the difficulty in finding reproducible, localized brain-behavior associations using mass univariate models, neuroimaging researchers have turned to multivariate machine learning methods that utilize all available imaging features for behavioral prediction, where the most predictive imaging features are interpreted *post hoc* (Dosenbach et al., 2010; Lebedev et al. 2014; Niu et al. 2019). This multivariate machine learning approach has shown success at capturing generalizable brain-behavior associations (Kragel et al. 2018; Sui et al. 2020), including applications in understanding individual variability in brain maturation (Brown et al. 2012), intelligence (Finn et al. 2015), emotional processing (Chang et al. 2015), and symptoms

of psychiatric disorder (Rosenberg et al. 2018), to name a few. However, many machine learning models rely on the raw imaging phenotype (Smith et al. 2015; Sripada et al. 2019; Hong et al. 2020), which prevents cross-study applications of mass univariate statistics. Some multivariate statistical methods also lack one important benefit of the statistical parametric brain mapping approach, i.e. the unbiased estimation and interpretation of a brain-behavior association with proper control for confounds. The concern of black-box applications of multivariate prediction methods has arisen (Davatzikos 2019; Efron 2020) due to the data-driven approach and the lack of hypothesis-driven, generative models, and cautious interpretation of predictive models is needed since the predictive features can be ephemeral rather than important and generalizable (Scheinost et al. 2019; Efron 2020).

Inspired by the success of the field of genetics in implementing generative models to identify generalizable genotype-phenotype mapping (Visscher et al. 2017; Efron 2020), we introduce a novel multivariate prediction method, the polyvertex score (PVS), that captures the many generalizable effect sizes across all vertices. It has several innovations. First, the PVS is a generative model that explicitly takes a global prior such that all vertices contribute to the observed brain-behavior association. No statistical threshold or dimension reduction of the imaging phenotype is necessary for the calculation of the PVS. Second, the PVS can be thresholded to reflect the user's hypothesis on the underlying signal sparsity of a brain-behavior association of interest. Comparing the predictive performance of PVSs of varying statistical thresholds yields empirical insights into the true sparseness of a brain-behavior association. Last but not least, the PVS can be applied directly to statistical parametric maps derived from mass univariate analysis, setting it apart from existing multivariate statistical methods. It can be

deployed in smaller sample studies to boost predictive power when the mass univariate summary statistics can be obtained from large neuroimaging consortiums.

For comparison purpose, two versions of the PVS were developed. The mass univariate PVS (PVS_U) is a summary measure of all the mass univariate effect sizes across the cortex, which are readily available for most neuroimaging analyses. The Bayesian PVS, on the other hand, is a multivariate extension of the PVS_U that accounts for correlation across vertices as well as the non-sparseness of the brain signal on behavior. Previous research have shown that ignoring the correlation structure among vertices results in biased estimation of the parameter of interest (Thompson et al. 2015), limiting the ability of mass-univariate approaches to localize effects and to make accurate predictions. The PVS_B (Figure 1) incorporates the covariance structure of the imaging phenotype during parameter estimation. Leveraging the unprecedented large sample size of the Adolescent Brain Cognitive Development (ABCD) Study, we demonstrated the utility of the PVS_B as a reliable multivariate method with great out-of-sample prediction performance, by comparing it to the PVS_U, the min-p, and popular multivariate methods, including lasso regression, random forest, and support vector regression with linear kernel. In addition to its good predictive performance, the property of the PVS_B enable us to demonstrate that the generalizable brain-behavior association is distributed in the global patterning of effect sizes across the cortex.

2.3 Methods

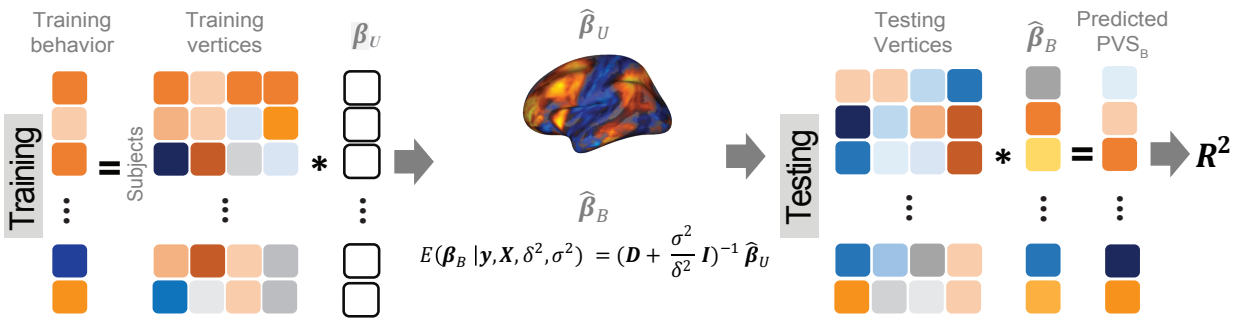


Figure 2.1 Overview of the PVS_B and the PVS_U algorithms. Ten-fold cross validation was performed to obtain a PVS_B for each individual. For each fold, mass univariate summary statistics, $\hat{\beta}_U$, were obtained from the training set which contained 90% of the complete sample. Posterior mean effect sizes at each vertex, $\widehat{\beta_B}$, were approximated by multiplying the mass univariate beta estimates, $\hat{\beta}_U$, by the inverse of the correlation structure of the brain, D, and a shrinkage factor that accounts for the number of vertices, V, and the total signal of the brain-behavior association, S. The PVS_B were subsequently calculated for the test set participants by multiplying their imaging phenotype with the $\widehat{\beta_B}$. Simulations were conducted at three levels of total explainable signal, six levels of study sample size, and four levels of proportion of non-null vertices, yielding 60 instantiations of simulation conditions with 100 iterations per condition.

We assume the relationship between a behavioral phenotype of interest and the imaging phenotype is captured by a general linear model. Specifically, let N denote the number of participants and let V denote the number of vertices. Then

$$\boldsymbol{y} = \sum_{v=1}^{V} \boldsymbol{x}_v \beta_v + \boldsymbol{\varepsilon} \quad (1)$$

where $\boldsymbol{y}$ is a standardized $N\ x\ 1$ vector of behavioral phenotypes, $\boldsymbol{x_v}$ is standardized $N\ x\ 1$ vector of imaging data, and β_v is the association parameter for the vth voxel, $v = 1, \ldots V$. More generally, the regression model (1) will also include covariates of no interest, which we omit here for simplicity of exposition.

It is often not possible to estimate model (1) directly, as there are a greater number of vertices than participants, $V > N$, and the desired associations with the behavioral phenotype is instead usually estimated using a mass-univariate regression approach across individual voxels. The mass-univariate regression model is given by

$$\boldsymbol{y} = \boldsymbol{x_v}\beta_v + \boldsymbol{\varepsilon_v},\ v = 1, \ldots, V. \quad (2)$$

Let $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_V})$ denote the $N\ x\ V$ matrix of standardized imaging phenotypes and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_V)'$ the $V\ x\ 1$ vector of association parameters of interest. The mass-univariate brain mapping model (2) omits information contained in the correlation across columns of $\boldsymbol{X}$ when estimating $\boldsymbol{\beta}$. In a least-squares framework, this is equivalent to assuming that the sample correlation matrix of the brain phenotype, $\boldsymbol{X'X} = \boldsymbol{I}$, the $V\ x\ V$ identity matrix. The least-squares estimates of $\boldsymbol{\beta}$ based on mass-univariate model (2) thus take the form

$$\widehat{\boldsymbol{\beta}}_U = \boldsymbol{X'y}. \quad (3)$$

Independent estimation of the parameter estimate at each vertex allows for estimation when $V > N$ and otherwise reduces the computational demand and produces more stable estimates when the voxels are highly correlated with each other when $N > V$.

Recent debates on reproducibility and small effect sizes in neuroimaging research are based on such mass-univariate estimates from the brain mapping framework. However, ignoring the correlation structure among vertices results in biased estimation of $\boldsymbol{\beta}$ as described in Thompson et al. (2015), limiting the ability of mass-univariate approaches to localize effects and to make accurate predictions. Moreover, the magnitude of the generalizable brain-behavior association is usually inferred from the effect size estimates of *only the most significant* vertices/ROIs using the mass univariate estimators (3). Such an approach assumes that the underlying true association is sparse and localized in the cortex, and hence that clusters of vertices/ROIs with minimum p-values (min-p) form the basis of generalizable signals. However, emerging evidence from large consortia such as ABCD indicates that the explanatory power of the brain on behavior is non-sparse, and thus cannot be captured solely by the most significant vertices/ROIs. In order to generalize the effect sizes of the whole brain phenotype, we need a

prediction framework that accounts for correlation across voxels as well as the non-sparseness of the brain signal on behavior. Rooted in this brain mapping approach, we propose the Bayesian polyvertex score (PVS_B) estimation and prediction framework.

Empirical Bayes estimation of parameter of interest

To tackle the correlated signal of the imaging phenotype at each vertex, we developed a Bayesian approach where the correlation information across vertices is incorporated into the parameter estimation process. Similar frameworks have been proposed in the field of genetics (Vilhjálmsson et al. 2015). Let $\boldsymbol{y}$ denotes standardized $N\ x\ 1$ vector of behavioral phenotypes, $\boldsymbol{X} = (\boldsymbol{x_1}, \dots, \boldsymbol{x_V})$ denote the $N\ x\ V$ matrix of standardized imaging phenotypes and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_V)'$ the $V\ x\ 1$ vector of association parameters of interest. The intuition behind the formulation of the PVS_B is to estimate the posterior expectation of the multivariate linear regression coefficients $\boldsymbol{\beta}$ from model (1) utilizing the mass-univariate estimator $\widehat{\boldsymbol{\beta}}_U$ from (3) and a regularized estimator of its $V\ x\ V$ correlation matrix.

To do so, we assume that the residuals $\boldsymbol{\varepsilon}$ are independent and normally distributed with constant variance, $\boldsymbol{\varepsilon} \sim N_N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, and give an independent normal prior with constant variance for the regression coefficients, $\boldsymbol{\beta} \sim N_V(\boldsymbol{0}, \delta^2\boldsymbol{I})$. It is easy to show that the posterior distribution of $\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}, \delta^2, \sigma^2$ is again multivariate normal with expectation

$$E(\boldsymbol{\beta}_B \mid \boldsymbol{y}, \boldsymbol{X}, \delta^2, \sigma^2) = \left(\boldsymbol{D} + \frac{\sigma^2}{\delta^2}\ \boldsymbol{I} \right)^{-1} \widehat{\boldsymbol{\beta}}_U \quad (4)$$

where $\boldsymbol{D} = \boldsymbol{X}'\boldsymbol{X}$ is the $V\ x\ V$ correlation matrix of $\widehat{\boldsymbol{\beta}}_U$. We can thus express the vertexwise posterior mean effect sizes of the brain phenotype under model (1) by weighting the mass univariate beta estimates $\widehat{\boldsymbol{\beta}}_U$ with a factor that accounts for the observed correlation structure of

the cortex $\boldsymbol{D}$ and the per-vertex variance explained δ^2. Since we do not know δ^2 and σ^2 *a priori*, we use plug-in estimators based on the summary statistics from the mass univariate model. We accomplish this using a method-of-moments estimator of the variance explained per vertex (Schwartzman et al. 2017):

$$\widehat{S^2} = \hat{m}_{eff}\left(\overline{z^2} - 1\right) \quad (6)$$

where $\overline{z^2}$ is the mean of the squared z-statistics of the mass univariate regressions across vertices, and the $\hat{m}_{eff}$ is the estimated effective number of vertices: $\hat{m}_{eff}$is calculated as the number of vertices, V divided by the second spectral moment of the correlation matrix $\boldsymbol{D}$. Then the estimated per-vertex variance explained is given by $\widehat{\delta^2} = \widehat{S^2}/V$ and the estimated residual variance is given by $\widehat{\sigma^2} = (1 - \widehat{S^2})$. Thus, our Empirical Bayes estimator for $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_B = (\boldsymbol{D} + \frac{\widehat{\sigma^2}}{\widehat{\delta^2}}\boldsymbol{I})^{-1}\widehat{\boldsymbol{\beta}}_U \quad (7)$$

The benefits of implementing this Empirical Bayes parameter estimation are twofold: 1) the procedure takes into account the correlation of the brain phenotype across vertices, and 2) the total estimated signal of the brain-behavior relationship is incorporated as a data-driven regularization parameter.

Behavioral prediction

Polyvertex scores

Motivated by the success of PRS, a polyvertex score (PVS) can be calculated from neuroimaging data by aggregating the predictive power of all vertices on a given behavioral phenotype. We implemented two types of PVS that utilize the mass univariate and Empirical

Bayes parameter estimates respectively. A mass univariate PVS (PVS_U), based on the mass univariate parameter estimates, was computed as the brain phenotype at each vertex for an individual multiplied by the mass univariate parameter estimates acquired from an independent sample:

$$\hat{y}_{PVS_U} = \sum_{j}^{V} X_j \hat{\beta}_{U,j}$$

where y is a standardized $N\ x\ 1$ vector of behavioral phenotypes, $\boldsymbol{x_v}$ is standardized $N\ x\ 1$ vector of imaging data, and β_v is the association parameter for the vth voxel, $v = 1, \ldots V$. The PVS_U summarizes the effect size at all vertices on individual variability in behavior, with the assumption of independence at each vertex.

Similarly, a Bayesian PVS (PVS_B; Figure 1) was calculated using the Empirical Bayes parameter estimates:

$$\hat{y}_{PVS_B} = \sum_{j}^{V} X_j \hat{\beta}_{B,j}$$

The PVS_B is hypothesized to harness the multivariate effect of an imaging phenotype on behavior by accounting for the correlation structure and the total explainable signal of the brain phenotype and should therefore yield a superior predictive performance over the PVS_U.

Thresholding based on statistical significance

To address the possibility that the explanatory power on behaviors is sparse and localized in the brain, a canonical assumption of mass univariate statistical models, we tested whether thresholding the number of vertices based on statistical significance would improve the prediction performance of the PVS_B. The thresholding procedure was performed as follows: we

ranked the absolute effect sizes for all vertices and removed those ranked lower than a threshold proportion. Three levels of thresholding were implemented such that the top 50%, 10% and 1% of vertices were retained for the PVS_B.

To link our predictive methods with the canonical statistical inference approach where a brain and behavior relationship is established when any single vertex shows a significant association with the behavior, we compared our methods with the predictive performance of the vertex with the most significant mass univariate z-score which we have referred to as the min-p model. A total of six prediction models were examined, namely, the PVS_U, the PVS_B, PVS_B 50%, PVS_B 10%, PVS_B 1%, and min-p.

Out-of-sample variance explained R^2 was used to evaluate the predictive accuracy of each method. Simulations were performed to assess the predictive performance of the above mentioned six methods. The simulation procedure, cross validation scheme, and simulation results were shown in the Supplementary Information.

Empirical Data

We examined whether functional neuroimaging phenotypes could predict complex behaviors with greater predictive power by (1) aggregating over all unthresholded effects across the cortex, and (2) incorporating the covariance structure of the imaging phenotype. The PVS_B and its thresholded variants, the PVS_U, and min-p model were implemented to assess the predictive power of two fMRI contrasts on two different cognitive tasks using the baseline data of the ABCD Study.

Sample

The ABCD Study is a population-based longitudinal study across 21 data acquisition sites in the U.S. following 11,875 demographically diverse children starting at 9 and 10 years old

(Garavan et al. 2018). Participants were recruited through a probability sampling procedure at the school level within the defined catchment area of the study's nationally distributed set of 21 recruitment sites. The ABCD sample also included a large twin cohort and many siblings. Family relatedness was documented and controlled for in the analyses in this paper such that twins and siblings from the same family were grouped into the same training or testing set. Study inclusion criteria were detailed in Casey et al. (Casey et al. 2018) and Hagler et al. (Hagler et al. 2019). Additional data quality control was applied to the complete baseline data of the ABCD Study, yielding a final sample of over 6000 participants. Quality control criteria and the descriptive characteristics of the final sample were presented in the Supplementary Information.

With the complete baseline data of the ABCD study, we estimated the predictive performance of the vertex-wise *2 back - 0 back* contrast from the Emotional N-back fMRI task (nBack; Casey et al. 2018) and the *correct stop vs. correct go* contrast from the Stop Signal Task (SST; Logan 1994) on the *Total Composite Score* from the NIH Toolbox Cognition Battery Ages 7-17 (TC; Gershon et al. 2013) and the *Stop Signal Reaction Time* (SSRT) from the SST task respectively. NBack-SSRT and SST-TC associations were assessed to examine the specificity of prediction. Four brain-behavior associations of interest were examined: nBack predicting TC, nBack predicting SSRT, SST predicting TC, and SST predicting SSRT. To account for the potential contribution of subcortical regions-of-interests (ROIs) on behavioral variability, a whole brain imaging phenotype was created by combining the BOLD activity of the FreeSurfer subcortical ROIs (excluding ventricles; Fischl et al. 2002) with the vertexwise fMRI data. The predictive performance of these whole brain phenotypes on the above-mentioned four associations was also estimated.

Ten-fold cross validation was performed for each association. Variance explained R^2, the squared Pearson correlation between the observed and predicted behavioral phenotypes, was calculated for each association. Rigorous covariate control was applied to the data to ensure that the identified brain-behavior associations could not attributable to demographic and socioeconomical confounds known to influence brain or behavioral variation. Failing to control for these confounds may result in inflated prediction performance for multivariate methods (Scheinost et al. 2019). Both brain and behavioral phenotypes were pre-residualized for age, sex, race, ethnicity, household income, parental education, household marital status, and scanner ID independently within each training and testing set.

FMRI tasks and processing steps

The nBack task incorporated facial and emotional processing to the traditional N-back task to assess memory and emotional regulation processes. The nBack task consisted of two runs. Within each run, participants were shown a series of stimuli and were instructed to indicate if a stimulus was the same as or different from the stimulus they saw N items earlier for each stimulus. There were two conditions for the nBack task: a 2-back vs fixation condition and a 0-back vs fixation condition which served as baseline. The *2back – 0back* contrast was used in this analysis.

The SST was used to assess the BOLD activity during inhibitory control. Participants were instructed to indicate the direction of a left or right arrow as quickly and accurately as possible, but were instructed not to respond when a left or right arrow was followed by an upward arrow. The full details of the fMRI tasks used in the ABCD Study were documented in Casey et al. (Casey et al. 2018).

Structural and task-based functional MRI data acquisition were conducted with 3T scanners, with multiband echo planar imaging with fast integrated distortion correction, and were harmonized across scanner vendors. Preprocessing steps included head motion correction, B_0 distortion correction, resampling with cubic interpolation, between-scan motion correction, and automated registration. General linear models implemented in AFNI's 3dDeconvolve were used to estimate task-related activation strength at the individual subject level, with hemodynamic response functions modeled using a gamma variate basis function plus its temporal derivative. Averaged beta coefficients per participant across two runs were calculated by weighting each run with the nominal degrees of freedom of that run, and were used in this analysis. Detailed imaging processing and analysis pipelines were described in Hagler et al. (Hagler et al. 2019).

Behavioral measures

The NIH Toolbox Cognition Battery Ages 7-17 (NTCB) is a comprehensive suite of neuro-behavioral measurements. NTCB consists of seven subtests: the Flanker Inhibitory Control and Attention Test, the Picture Sequence Memory Test, the List Sorting Working Memory Test, Picture Vocabulary Test, Oral Reading Recognition Test, Dimensional Change Card Sorting Test, and Pattern Comparison Processing Speed Test. The Total Composite score (TC), the average of all 7 subtests, is a composite index of general cognitive ability and was used in this study.

The SSRT was derived from the behavioral performance measures acquired during the SST. It was computed by subtracting the median stop signal delay of all successful stop trials from the n^{th} percentile GO reaction time, where n represents the percentage of successful inhibitions.

Multivariate method comparisons

To assess the predictive performance of the PVS_B relative existing multivariate models, we submitted the above-mentioned empirical data to three additional multivariate methods: LASSO, random forest, and support vector regression (SVR) with linear kernel, and compared their predictive performance with that of the PVS_B. These methods were chosen to complement PVS_B's statistical emphasis. While LASSO is a parametric regression-based method similar to the PVS_B, its sparsity assumption sets it aside from the PVS_B which assumes a global prior. SVR and random forest, on the other hand, are popular nonparametric methods that capture nonlinear effects for better prediction. Matlab implementation of these multivariate methods was described in the Supplementary Information. To make our results compatible with previous multivariate analyses of differing covariates treatments (Finn et al. 2015; Sripada et al. 2019; Cui et al. 2020), we repeated these analyses with two additional covariate control schemes. Out-of-sample variance explained for all multivariate methods after controlling for (1) age, sex, and scanner ID, and (2) age, sex, scanner ID, race and ethnicity, was estimated.

2.4 Results

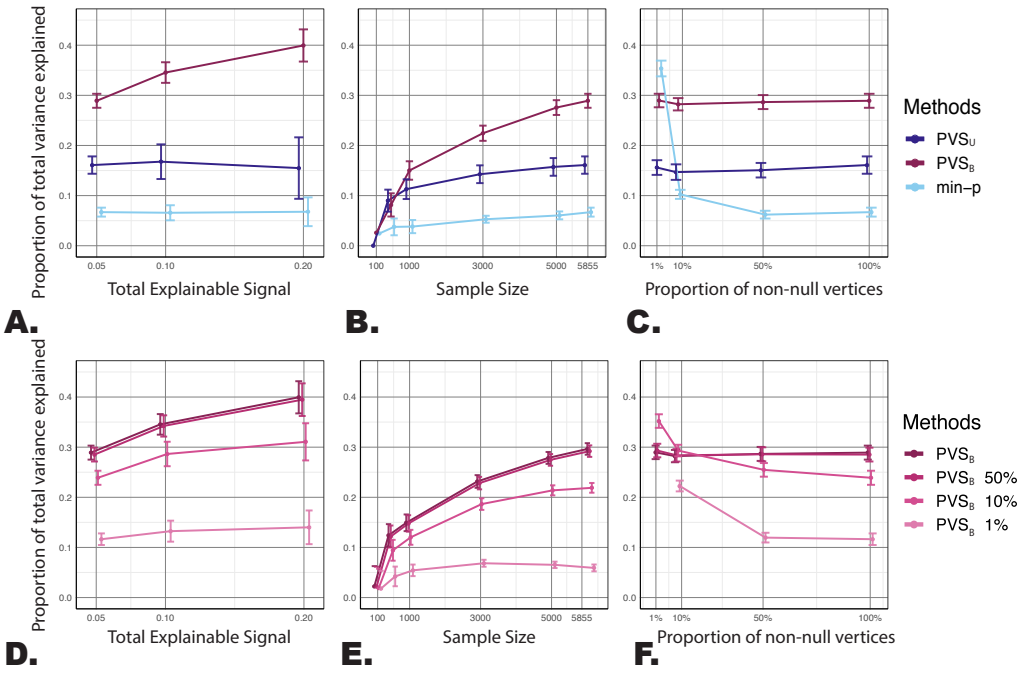


Figure 2.2 The PVS_B demonstrated superior predictive performance compared to the PVS_U, and min-p (A-C) as well as its thresholded variants (D-F) across various simulated conditions. Figure 2 showed the mean and 1.96 standard deviation confidence interval of the proportion of total variance explained by each method. The advantage of the PVS_B over the PVS_U and the min-p was most prominent at higher levels of total explainable signal (A), larger sample size (B), and with increased proportion of non-null vertices (C). When the true signal structure was nonsparse, (10%, 50% and 100% of non-null vertices; C), the advantage in prediction accuracy of the PVS_B was prominent, manifested by its superior, prediction performance compared to the PVS_U and min-p. Superior predictive performance was established for the PVS_B compared to its thresholded variants (D-E). When the true signal structure was global, the PVSB outperformed its thresholded variants across levels of total explainable signals (D) and across sample sizes (E). Sensitivity to the underlying signal structure of the PVS_B was estimated with varying simulated levels of signal sparsity (F): when the true signal structure was sparse, i.e. the proportion of non-null vertices was small, the thresholded PVS_B at the corresponding level of statistical threshold outperformed the unthresholded PVS_B, highlighting the sensitivity of the PVS_B to the underlying signal structure of the simulated brain-behavior association. The complete simulation results were reported in Supplementary Materials.

Simulation results: the PVS_B demonstrated superiority at capturing global, distributed brain-behavior association patterns

The reliability of the predictive performance of the PVS framework was established across a suite of simulated conditions (Figure 2 and Supplementary Information). The PVS_B better captured the variance explained in behavior than the PVS_U and the min-p (most significant

vertex) across varying levels of magnitude (total explainable signal; Figure 2.A), sample size (Figure 2.B), and non-sparseness of the true signal across the cortex (proportion of non-null vertices; Figure 2.C). This shows the benefit of accounting for the correlation structure of the brain for behavioral prediction. The PVS_B also outperformed its thresholded variants when the signal structure was global (Figure 2.D, Figure 2.E). When the true signal structure was sparse, thresholding PVS_B at the corresponding threshold yielded better performance (Figure 2.F), demonstrating the sensitivity of the PVS_B to the underlying true signal structure.

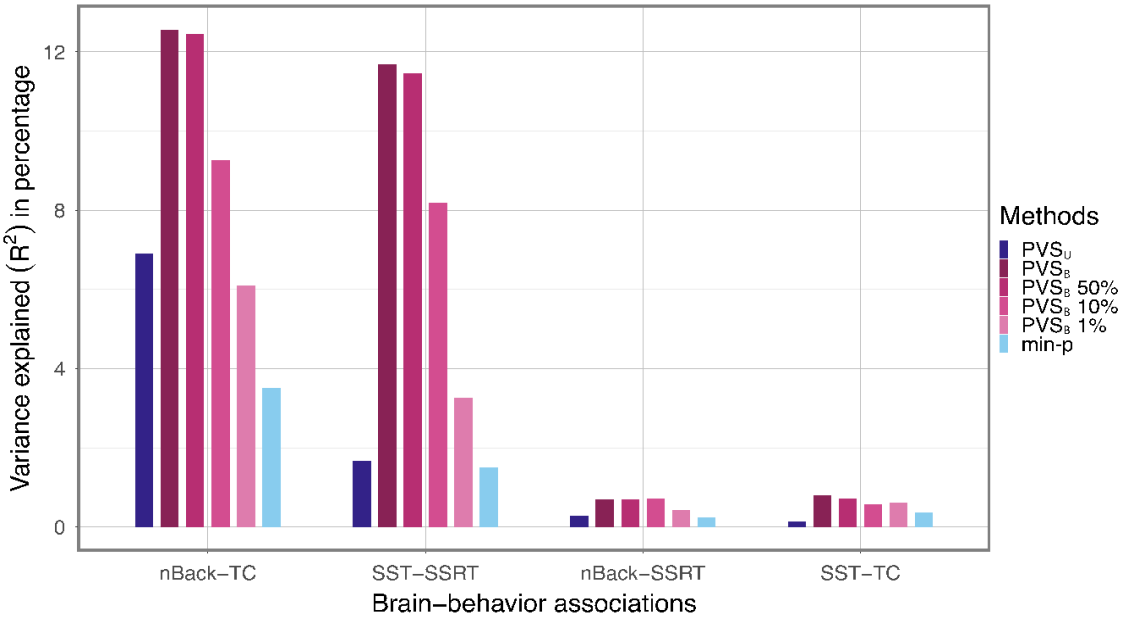


Figure 2.3 Decreased predictive performance was associated with more stringently thresholded models for the empirical brain-behavior associations. Variance explained, R^2, for four brain-behavior associations (nBack predicting TC, SST predicting SSRT, nBack predicting SSRT, SST predicting TC) were examined using the PVS_U, PVS_B, thresholded variants of the PVS_B, and min-p. Significant associations were identified for the nBack-TC and SST-SSRT, but not for the nBack-SSRT and SST-TC associations. For the two significant associations, the best prediction performance was achieved by the PVS_B where all vertices where included in the model. Predictive performance decreased as more stringent thresholds were applied. All brain and behavioral variables were pre-residualized for age and categorical variables including sex, parent marital status, highest level of parental education, household income, self-reported race and ethnicity, and MRI scanner ID.

Empirical results: behavioral variability was better predicted by the unthresholded task activation pattern captured by the PVS_B

After establishing the efficacy of the PVS_B at capturing global brain-behavior associations using simulation, we explored whether individual differences in cognitive processes can be better predicted by whole brain rather than thresholded task activation patterns using the baseline data from the ABCD Study (Figure 3). For the nBack-TC association, the PVS_B outperformed the PVS_U and min-p, capturing 12.6% compared to 6.9% and 3.5% of the variance in the n-Back-TC association. Similar improvement in prediction accuracy was also observed for the SST-SSRT association. The PVS_B was able to explain 11.7% of the variance in SSRT using the vertex-wise BOLD variation of the correct stop vs. correct go contrast from the SST, compared to 1.7% and 1.5% for the PVS_U and the min-p respectively. The increased predictive performance of the PVS_B highlights the importance of accounting for the correlation structure of the imaging phenotype when measuring the generalizable signal between brain and behavior. Interestingly, the imaging contrasts only showed associations with the behaviors that used similar underlying cognitive constructs to the fMRI tasks: no associations were found between the nBack contrast and SSRT and between the SST contrast and TC. Such specificity has been found in other fMRI studies (Rosenberg et al., 2019) and further highlights the benefit of the PVS_B at capturing effective association patterns without overfitting.

Thresholding the PVS_B at varying statistical thresholds, on the other hand, did not confer any advantage for prediction accuracy. For both significant associations, the predictive performance of the PVS_B decreased as more stringent statistical thresholds were applied. Specifically, for the nBack-TC association, decreased predictive performance was found for the PVS_B 50% (R^2=12.4%) and PVS_B 10% (R^2=9.2%) compared to the PVS_B (R^2=12.6%). A similar

drop in predictive accuracy was found for the SST-SSRT association (PVS_B: R^2=11.7%, PVS_B 50%: R^2=11.5%; PVS_B 10%: R^2=8.2%). Thresholding based on the vertex-wise p-values resulted in decreased prediction accuracy, suggesting that vertices with subthresholded p-values were still informative for behavioral prediction. Visualization of the distributed pattern of mass univariate statistics (Figure 4A, C) and the posterior mean effect sizes (Figure 4B, D) across the cortex, further corroborated our hypothesis that the predictive effect of the brain on complex behavior was indeed global and distributed across the cortex. No improvement was observed by including subcortical ROIs for prediction, as shown by the comparable predictive performance of the cortical imaging phenotype relative to that of the whole brain imaging phenotypes (Table S2).

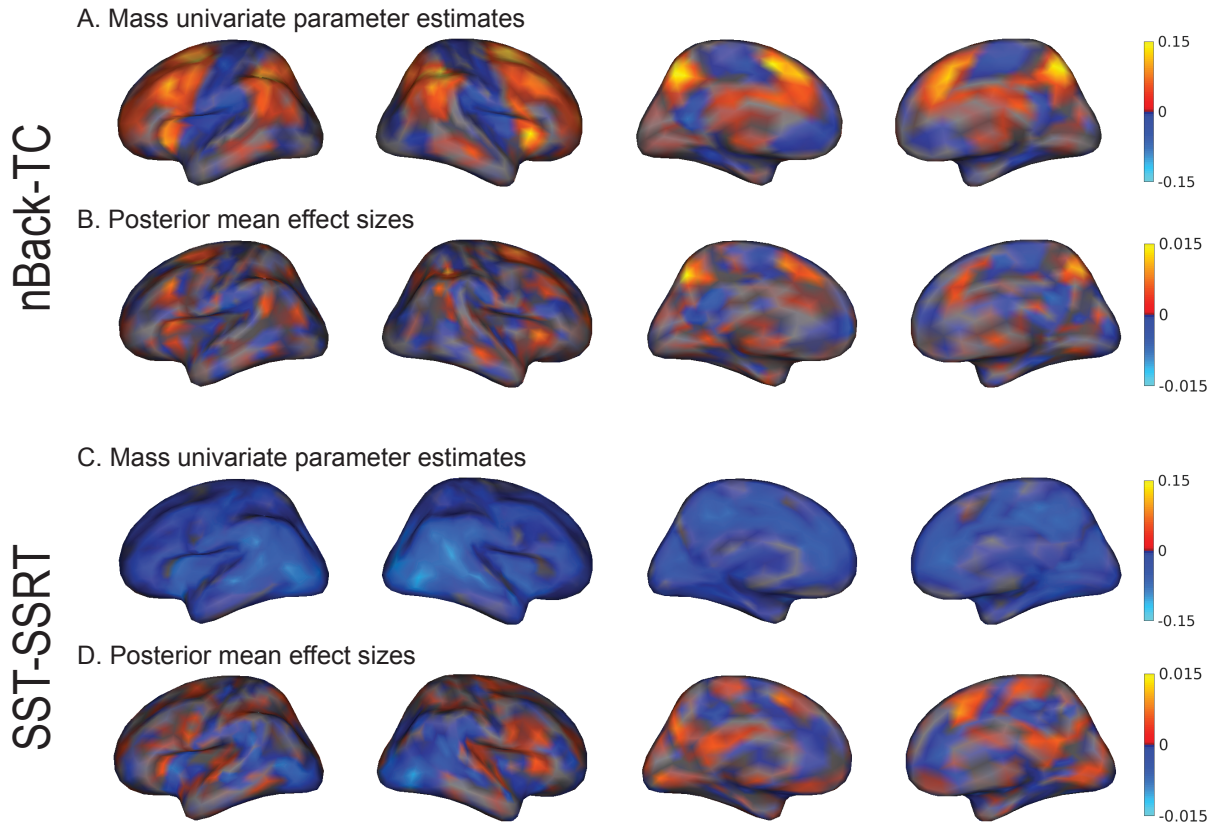


Figure 2.4 Distributed pattern of effects across the cortex for the nBack-TC and SST-SSRT associations. Unthresholded vertexwise mass univariate standardized parameter estimates and posterior mean effect sizes were displayed for the nBack-TC and SST-SSRT associations**.** The unthresholded mass univariate parameter estimates were used to calculate the PVS_U and the posterior mean effect sizes were used to calculate the PVS_B for each association. For the nBack-TC association, both mass univariate (A) and posterior mean effect size maps (B) showed distributed patterns of association across the cortex, suggesting that the association between imaging and behavioral phenotypes were global, spanning the whole cortex. Similar distributed patterns were found for the SST-SSRT association. Compared to the mass univariate statistical map (C), the posterior mean effect size map (D) of the SST-SSRT association showed greater variation in the relative weighting of brain regions on behavioral prediction, potentially contributing to the greater prediction accuracy for the PVS_B compared to the PVS_U for this association. Interpretation of the posterior mean effect size maps should be attempted with caution as the validity of these effect sizes rests upon the assumption of the global prior such that all vertices contribute to brain-behavior associations.

Multivariate method comparisons

The PVS_B demonstrated comparable if not superior predictive performance compared to other multivariate methods (Table 1). All methods except the SVR with linear kernel explained at least 10% of the variance explained of the nBack-TC association after controlling for

demographic variables. For the SST-SSRT association, the PVS_B explained more than 10% of the out-of-sample variance in behavior, while random forest and LASSO achieved only 8% and 8.7% variance explained respectively. Such difference in prediction accuracy may be due to overfitting of the random forest and LASSO to the noisier SST imaging phenotype, highlighting the importance of having high quality imaging data in multivariate statistical analysis. More stringent covariate control, on the other hand, reduced the predictive performance of all methods on the nBack-TC but not the SST-SSRT association (Table S2), suggesting that insufficient covariates control would inflate the magnitude of association detected by multivariate statistical methods and that such inflation maybe specific to the association of interest.

Table 2.1 The PVSB demonstrated comparable if not superior generalization performance relative to other multivariate methods. The out-of-sample variance explained, R^2, of each multivariate method was shown for each empirical brain-behavior association. While the PVS_B, random forest, and LASSO showed comparable predictive performance for the nBack-TC association, the PVS_B outperformed other multivariate methods for the SST-SSRT association. Minimal association was again found for the nBack-SSRT and SST-TC contrasts.

Associations	PVSB	Random Forest	LASSO	SVR
nBack-TC	11.58	10.88	10.42	2.25
nBack-SSRT	0.76	0.12	0.13	0.07
SST-TC	0.46	0.38	0.00	0.08
SST-SSRT	10.64	8.03	8.70	3.45

2.5 Discussion

In this study, we presented the empirical utility of a new multivariate prediction method, the Bayesian polyvertex score (PVS_B), and showed that greater out-of-sample behavioral prediction of imaging phenotypes could be achieved by explicitly modeling the globally distributed brain-behavioral associations across the cortex. The PVS_B captures the unthresholded,

generalizable predictive effect of an imaging phenotype on behavior by calibrating the mass univariate summary statistics with the estimated covariance structure of the imaging phenotype and a global prior of effect sizes. Using the Emotional N-back and the SST fMRI contrasts from the ABCD Study, we demonstrated that greatest predictive performance of fMRI phenotypes on complex behaviors can be achieved using the unthresholded multivariate effect size pattern captured by the PVS_B. Our findings suggested that the predictive power of imaging phenotypes on complex behaviors was distributed rather than localized across the cortex, and such global effect needs to be explicitly modeled in the statistical methods used by neuroimaging studies in order to holistically understand the neural bases of psychiatric disorders and cognitive functions.

Traditional methods used to analyze brain-behavior relationships aim to detect individual brain regions or localized clusters significantly associated with phenotypic variation. Rooted in signal detection theory, this brain mapping approach has been fruitful in characterizing the explanatory effect of brain regions on behavior but has demonstrated suboptimal replicability (Ihnen et al. 2009; King et al. 2019) and therefore limited translational utility for psychopathology. When predicting individual differences in phenotypic variation, sparse and focal association patterns, captured by statistical thresholding based on p-values, did not confer any advantage over the unthresholded association pattern captured by the PVS_B. Along with other neuroimaging studies that have reported similar distributed association patterns (Gonzalez-Castillo et al. 2012; Poldrack et al. 2017; Dubois et al. 2018; Bruin et al. 2018; Sripada et al. 2019), our findings suggest that the power for predicting individual variability in complex cognitive behaviors is globally distributed in an imaging phenotype, above and beyond a localized and sparse region or network. While lower-level visual motor processes and specialized cognitive behaviors may be more accurately captured by localized association patterns (Sereno et

al. 1995; Tsao et al. 2003), complex cognitive processes and risk for psychiatric disorders may be more accurately predicted by distributed, global patterns of BOLD activation across the cortex.

Our results further demonstrate the importance of moving beyond mass univariate statistical models in neuroimaging research (Kragel et al. 2018; Reddan et al. 2017; Smith and Nichols 2017). Common fMRI practice assumes that BOLD activity has a localized correlation structure. However, long range correlations in BOLD activity across the cortex have been well-documented by resting-state fMRI research and have demonstrated important behavioral implications. Here we found that including the covariance structure of the imaging phenotype during parameter estimation greatly improved the predictive performance of functional imaging phenotypes. Similar improvement of predictive accuracy of regional cortical morphology on cognitive outcomes was shown in children (Palmer et al., in prep). Predictive performance is of great importance to the utility of biomarkers on clinical disorders or cognitive processes. As a result, neuroimaging studies that aim to generate potential biomarkers using functional neuroimaging phenotypes should adopt and develop new statistical methods that estimate the multivariate, distributed associative effects between brain and behavior.

Our results echoed the solution of the small effect size issue adopted by the field of genetics. Similar to traditional neuroimaging studies, mass univariate statistical models and statistical thresholding are used in genome-wide association studies (GWAS) to localize genetic loci that are significantly associated with psychiatric disorders and cognitive processes. With thousands of participants and unprecedented statistical power, GWAS-based significant genetic loci only account for a fraction of the variance in complex human phenotypes. To resolve this issue, polygenic risk scores (PRS; Purcell et al. 2009) were subsequently developed from GWAS

to aggregate the small effect sizes across the whole genome, including those non-significant loci (Yang et al. 2010; Davies et al. 2011; Le Hellard et al. 2014; Torkamani et al. 2018). By pulling together the effects of many informative but not necessarily statistically significant genetic variants, the PRS greatly improved the predictive performance of genetic data on stratifying psychiatric risk based on the polygenic burden of common variants (Purcell et al. 2009; Dudbridge 2013), fueling the discovery that complex behaviors are polygenic (Visscher et al. 2017; Gibson 2018). Given the similarity of observed small effect sizes of neuroimaging and genetics research, individual variability in complex behavior may be attributable to the structural and functional differences across the whole brain. Indeed, as our results indicated, complex behaviors are polyvertex, with each vertex contributing only minimally to the variance explained in behavior and thus not surviving statistical thresholding. To capture the distributed, small predictive effects of the brain on behavior, multivariate methods (Chang et al. 2015; Finn et al. 2015; Bruin et al. 2019) are essential, and a multivariate method that captures the subthreshold effect sizes of the imaging data is needed.

The PVS_B is one of many multivariate statistical methods available for neuroimaging analyses. While the predictive advantage of various multivariate models is dependent on the sample size, imaging features, the magnitude of the effect sizes and other sample characteristics (Jollans et al. 2019), our results showed that the predictive accuracy was comparable across multivariate methods, with slightly greater performance of the PVS_B on the SST-SSRT contrast. In addition to its good prediction accuracy, the PVS_B confer other advantages. First, following the brain mapping approach, the PVS_B maps the effect sizes at every vertex without reducing the dimensionality of the imaging phenotype, providing the scientific interpretability unavailable to nonparametric multivariate methods, making it a useful tool for the neuroimaging community.

Second, the PVS_B offers empirical insights into the underlying signal structure of a brain-behavior association of interest. Last, mass univariate summary statistics from large neuroimaging dataset can be supplied to the PVS_B for prediction in a small independent sample, which makes the PVS_B a useful tool to harvest the statistical power of large neuroimaging consortiums for smaller imaging studies.

In summary, results from this study suggest that in order for neuroimaging studies to identify possible biomarkers for cognitive and clinical outcomes, greater predictive power of the functional neuroimaging phenotypes needs to establish, which can be achieved through the statistical modeling of global, distributed effects using multivariate statistical methods, one of which being the PVS_B. Using a large sample from the ABCD study, we have demonstrated the utility of employing multivariate parameter estimation and aggregating the effect across a functional neuroimaging phenotype for greater predictive power for behavior. With an increasing interest in the predictive utility of imaging phenotypes as biomarkers for health and disease, this novel work will pave the way for improving our ability to reach this goal.

2.6 Acknowledgment

Chapter 2, in full, is a reprint of the material as it appears in Cerebral Cortex 2021. Weiqi Zhao, Clare E. Palmer, Wesley K. Thompson, Bader Chaarani, Hugh P. Garavan, ,B.J.Casey, Terry L. Jernigan, Anders M. Dale, & Chun Chieh Fan, Cerebral Cortex, 2021. The dissertation author was the primary investigator and the first author of this paper.

2.7 Reference

1. Brown, T.T., Kuperman, J.M., Chung, Y., Erhart, M., McCabe, C., Hagler, D.J., Jr, Venkatraman, V.K., Akshoomoff, N., Amaral, D.G., Bloss, C.S., Casey, B.J., Chang, L., Ernst, T.M., Frazier, J.A., Gruen, J.R., Kaufmann, W.E., Kenet, T., Kennedy, D.N., Murray, S.S., Sowell, E.R., Jernigan, T.L., Dale, A.M., 2012. Neuroanatomical assessment of biological maturity. Curr. Biol. 22, 1693–1698.

2. Bruin, W., Denys, D., van Wingen, G., 2019. Diagnostic neuroimaging markers of obsessive-compulsive disorder: Initial evidence from structural and functional MRI studies. Prog. Neuropsychopharmacol. Biol. Psychiatry 91, 49–59.

3. Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Jr, Daniela Cornejo, M., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., ABCD Imaging Acquisition Workgroup, 2018. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.

4. Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., Wager, T.D., 2015. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. PLoS Biol. 13, e1002180.

5. Cui, Z., Li, H., Xia, C.H., Larsen, B., Adebimpe, A., Baum, G.L., Cieslak, M., Gur, R.E., Gur, R.C., Moore, T.M., Oathes, D.J., Alexander-Bloch, A.F., Raznahan, A., Roalf, D.R., Shinohara, R.T., Wolf, D.H., Davatzikos, C., Bassett, D.S., Fair, D.A., Fan, Y., Satterthwaite, T.D., 2020. Individual Variation in Functional Topography of Association Networks in Youth. Neuron 106, 340–353.e8.

6. Davatzikos, C., 2019. Machine learning in neuroimaging: Progress and challenges. Neuroimage 197, 652–656.

7. Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A.J., Corley, J., Redmond, P., Fox, H.C., Haggarty, P., Whalley, L.J., McNeill, G., Goddard, M.E., Espeseth, T., Lundervold, A.J., Reinvang, I., Pickles, A., Steen, V.M., Ollier, W., Porteous, D.J., Horan, M., Starr, J.M., Pendleton, N., Visscher, P.M., Deary, I.J., 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol. Psychiatry 16, 996–1005.

8. Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Jr., Barch, D.M., Petersen, S.E., Schlaggar, B.L., 2010. Prediction of Individual Brain Maturity Using fMRI. Science 329.

9. Dubois, J., Galdi, P., Paul, L.K., Adolphs, R., 2018. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373.

10. Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9, e1003348.

11. Efron, B., 2020. Prediction, Estimation, and Attribution. J. Am. Stat. Assoc. 115, 636–655.

12. Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18, 1664–1671.

13. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole Brain Segmentation: Neurotechnique Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 33, 341–355.

14. Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R.Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., Zahs, D., 2018. Recruiting the ABCD sample: Design considerations and procedures. Dev. Cogn. Neurosci. 32, 16–22.

15. Gershon, R.C., Wagster, M.V., Hendrie, H.C., Fox, N.A., Cook, K.F., Nowinski, C.J., 2013. NIH toolbox for assessment of neurological and behavioral function. Neurology 80, S2–6.

16. Gibson, G., 2018. Population genetics and GWAS: A primer. PLoS Biol.

17. Gonzalez-Castillo, J., Saad, Z.S., Handwerker, D.A., Inati, S.J., Brenowitz, N., Bandettini, P.A., 2012. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. Proc. Natl. Acad. Sci. U. S. A. 109, 5487–5492.

18. Hagler, D.J., Hatton, S.N., Makowski, C., Daniela Cornejo, M., Fair, D.A., Dick, A.S., Sutherland, M.T., Casey, B.J., Barch, D.M., Harms, M.P., Watts, R., Bjork, J.M., Garavan, H.P., Hilmer, L., Pung, C.J., Sicat, C.S., Kuperman, J., Bartsch, H., Xue, F., Heitzeg, M.M., Laird, A.R., Trinh, T.T., Gonzalez, R., Tapert, S.F., Riedel, M.C., Squeglia, L.M., Hyde, L.W., Rosenberg, M.D., Earl, E.A., Howlett, K.D., Baker, F.C., Soules, M., Diaz, J., de Leon, O.R., Thompson, W.K., Neale, M.C., Herting, M., Sowell, E.R., Alvarez, R.P., Hawes, S.W., Sanchez, M., Bodurka, J., Breslin, F.J., Morris, A.S., Paulus, M.P., Kyle Simmons, W., Polimeni, J.R., van der Kouwe, A., Nencka, A.S., Gray, K.M., Pierpaoli, C., Matochik, J.A., Noronha, A., Aklin, W.M., Conway, K., Glantz, M., Hoffman, E., Little, R., Lopez, M., Pariyadath, V., Weiss, S.R.B., Wolff-Hughes, D.L., DelCarmen-Wiggins, R., Feldstein Ewing, S.W., Miranda-Dominguez, O., Nagel, B.J., Perrone, A.J., Sturgeon, D.T., Goldstone, A., Pfefferbaum, A., Pohl, K.M., Prouty, D., Uban, K., Bookheimer, S.Y., Dapretto, M., Galvan, A., Bagot, K., Giedd, J., Alejandra Infante, M., Jacobus, J., Patrick, K., Shilling, P.D., Desikan, R., Li, Y., Sugrue, L., Banich, M.T., Friedman, N., Hewitt, J.K., Hopfer, C., Sakai, J., Tanabe, J., Cottler, L.B., Nixon, S.J., Chang, L., Cloak, C., Ernst, T., Reeves, G., Kennedy, D.N., Heeringa, S., Peltier, S., Schulenberg, J., Sripada, C., Zucker, R.A., Iacono, W.G., Luciana, M., Calabro, F.J., Clark, D.B., Lewis, D.A., Luna, B., Schirda, C., Brima, T.,

Foxe, J.J., Freedman, E.G., Mruzek, D.W., Mason, M.J., Huber, R., McGlade, E., Prescot, A., Renshaw, P.F., Yurgelun-Todd, D.A., Allgaier, N.A., Dumas, J.A., Ivanova, M., Potter, A., Florsheim, P., Larson, C., Lisdahl, K., Charness, M.E., Fuemmeler, B., Hettema, J.M., Steinberg, J., Anokhin, A.P., Glaser, P., Heath, A.C., Madden, P.A., Baskin-Sommers, A., Todd Constable, R., Grant, S.J., Dowling, G.J., Brown, S.A., Jernigan, T.L., Dale, A.M., 2018. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. bioRxiv.

19. Hong, S.-J., Sisk, L., Caballero, C., Mekhanik, A., Roy, A.K., Milham, M.P., Gee, D.G., 2020. Decomposing complex links between the childhood environment and brain structure in school-aged youth. bioRxiv.

20. Ihnen, S.K.Z., Church, J.A., Petersen, S.E., Schlaggar, B.L., 2009. Lack of generalizability of sex differences in the fMRI BOLD activity associated with language processing in adults. Neuroimage 45, 1020–1032.

21. International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752.

22. Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, S., Grigis, A., Martinot, J.-L., Paus, T., Smolka, M.N., Walter, H., Schumann, G., Garavan, H., Whelan, R., 2019. Quantifying performance of machine learning methods for neuroimaging data. Neuroimage.

23. King, J.B., Prigge, M.B.D., King, C.K., Morgan, J., Weathersby, F., Fox, J.C., Dean, D.C., 3rd, Freeman, A., Villaruz, J.A.M., Kane, K.L., Bigler, E.D., Alexander, A.L., Lange, N., Zielinski, B., Lainhart, J.E., Anderson, J.S., 2019. Generalizability and reproducibility of functional connectivity in autism. Mol. Autism 10, 27.

24. Kragel, P.A., Koban, L., Barrett, L.F., Wager, T.D., 2018. Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. Neuron 99, 257–273.

25. Le Hellard, S., Steen, V.M., 2014. Genetic architecture of cognitive traits. Scand. J. Psychol. 55, 255–262.

26. Lebedev, A.V., Westman, E., Van Westen, G.J.P., Kramberger, M.G., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., Simmons, A., Alzheimer's Disease Neuroimaging Initiative and the AddNeuroMed consortium, 2014. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. Neuroimage Clin 6, 115–125.

27. Logan, G.D., 1994. On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In: Dagenbach, D. (Ed.), Inhibitory Processes in Attention, Memory, and Language , (pp. Academic Press, xiv, San Diego, CA, US, pp. 189–239.

28. Niu, X., Zhang, F., Kounios, J., Liang, H., 2020. Improved prediction of brain age using multimodal neuroimaging data. Hum. Brain Mapp. 41, 1626–1643.

29. Palmer, C.E., Zhao, W., Loughnan, R., Zou, J., Fan, C.C., Thompson, W.K., Jernigan, T.L., Dale, A.M., 2020. Determining the association between regionalisation of cortical morphology and cognition in 10,145 children. bioRxiv.

30. Reddan, M.C., Lindquist, M.A., Wager, T.D., 2017. Effect Size Estimation in Neuroimaging. JAMA Psychiatry 74, 207–208.

31. Rosenberg, M.D., Finn, E.S., Scheinost, D., Papademetris, X., Shen, X., Constable, R.T., Chun, M.M., 2016. A neuromarker of sustained attention from whole-brain functional connectivity. Nat. Neurosci. 19, 165–171.

32. Scheinost, D., Noble, S., Horien, C., Greene, A.S., Lake, E.M., Salehi, M., Gao, S., Shen, X., O'Connor, D., Barron, D.S., Yip, S.W., Rosenberg, M.D., Constable, R.T., 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. Neuroimage 193, 35–45.

33. Schwartzman, A., Schork, A.J., Zablocki, R., Thompson, W.K., 2017. A simple, consistent estimator of heritability for genome-wide association studies. bioRxiv.

34. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893.

35. Smith, S.M., Nichols, T.E., 2018. Statistical Challenges in "Big Data" Human Neuroimaging. Neuron 97, 263–268.

36. Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E.J., Glasser, M.F., Ugurbil, K., Barch, D.M., Van Essen, D.C., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat. Neurosci. 18, 1565–1567.

37. Sripada, C., Rutherford, S., Angstadt, M., Thompson, W.K., Luciana, M., Weigard, A., Hyde, L.H., Heitzeg, M., 2019. Prediction of neurocognition in youth from resting state fMRI. Mol. Psychiatry.

38. Stanfield, A.C., McIntosh, A.M., Spencer, M.D., Philip, R., Gaur, S., Lawrie, S.M., 2008. Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. Eur. Psychiatry 23, 289–299.

39. Sui, J., Jiang, R., Bustillo, J., Calhoun, V., 2020. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. Biol. Psychiatry.

40. Thompson, W.K., Wang, Y., Schork, A.J., Witoelar, A., Zuber, V., Xu, S., Werge, T., Holland, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Andreassen, O.A., Dale, A.M., 2015. An Empirical Bayes Mixture Model for Effect Size Distributions in Genome-Wide Association Studies. PLoS Genet. 11, e1005717.

41. Torkamani, A., Wineinger, N.E., Topol, E.J., 2018. The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. 19, 581–590.

42. Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., Tootell, R.B.H., 2003. Faces and objects in macaque cerebral cortex. Nat. Neurosci. 6, 989–995.

43. Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., Hayeck, T., Won, H.-H., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study, Kathiresan, S., Pato, M., Pato, C., Tamimi, R., Stahl, E., Zaitlen, N., Pasaniuc, B., Belbin, G., Kenny, E.E., Schierup, M.H., De Jager, P., Patsopoulos, N.A., McCarroll, S., Daly, M., Purcell, S., Chasman, D., Neale, B., Goddard, M., Visscher, P.M., Kraft, P., Patterson, N., Price, A.L., 2015. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am. J. Hum. Genet. 97, 576–592.

44. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22.

45. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569.

2.8 Supplementary Methods

Simulation Procedure

We used simulations to assess our novel approach and determine whether different types of PVSs perform as expected in certain contexts. In particular, we assessed how varying levels of total explainable signal of the brain phenotype, sparseness of the true signal in the brain, and sample size influenced the predictive accuracy of the above mentioned PVS methods.

For each iteration of the simulations, the predictive effect of the brain on behavior at each vertex, the true beta coefficient, was simulated as:

$$\beta \sim \begin{cases} N\left(0, \frac{\widehat{\delta^2}}{\pi}\right) \text{ with probability } \pi \\ 0 \text{ with probability } 1 - \pi \end{cases}$$

The β's were simulated by sampling independently from a standard normal distribution. A subset of these true beta coefficients was then set to zero as determined by 1 minus the proportion of vertices containing true signal, π. For example, for an instantiation of 10% vertices scenario, 10% of the vertices were randomly assigned to have non-null effects which can account for S of outcome variations in total, whereas the beta coefficients of the other 90% were set to zero.

The signal sparsity level, was simulated at the levels of 100%, 50%, 10%, 1%, and 0.1% vertices. The 0.1% vertices level corresponds to the Min-p assumption where the significant effect lies in a single vertex. The total explainable signal of the brain phenotype was simulated at the levels of 0.01, 0.05, 0.1, and 0.2. Each combination was simulated independently 100 times, giving 2000 iterations in total.

Then, the simulated behavioral phenotype was calculated as a combination of the effect of an empirically collected brain phenotype, X, and independent noise weighted by the square

root of the total explainable signal of that iteration. The independent noise was sampled at the participant level from a standard normal distribution. To make the simulation more realistic, we used the empirical brain phenotype data as the independent variables, X , which was a $5855\ x\ 1284$ brain matrix of the weighted average of the two runs of the 2-back - 0-back contrast of the nBack fMRI task of the baseline data of ABCD (ABCD Data Release 2.0.1; NDAR DOI: 10.15154/1504041). The weighting of the two runs were determined by the nominal degrees of freedom of each participant. X was smoothed at around FWHM 5mm, pre-residualized by age and categorical variables including sex, parent marital status, highest level of parental education, household income, self-reported race and ethnicity, and MRI scanner ID.

Within each iteration, two independent samples were randomly drawn at the sample size of 100, 500, 1000, 3000, 5000, and at the full sample size, 5855, to estimate the sample size dependency of prediction accuracy. Predicted behavioral phenotypes based on nine different models were calculated with 10-fold CV. R^2 served as a metric for predictive performance. The simulation results comparing the predictive performance of the PVS_B, the PVS_U, and min-p were shown in Figure S1, and the simulation results comparing the effect of thresholding based on statistical significance were shown in Figure S2.

Evaluating the performance of the PVS_B and PVS_U

We used 10-fold cross validation to evaluate the generalization performance of the PVS_U and the PVS_B. The same training-testing schema was applied to each fold to ensure the independency between estimation and prediction. For the PVS_U, at each fold, mass univariate beta estimates obtained from the training set, containing 90% of the full sample, were multiplied by the imaging phenotype of each test set participant to obtain a PVS_U score (the predicted behavioral phenotype) for each participant in the test set. This procedure was repeated 10 times

yielding a predicted behavioral score for each participant in the full sample. For the PVS_B, the posterior effect sizes were calculated with the estimated total signal of the brain-behavior association, the correlation structure of the imaging phenotype, and the mass univariate beta estimates from the training data, and multiplied by the imaging phenotype of each test set participant to obtain a predicted PVS_B score for each participant. Variance explained, R^2 (the squared Pearson correlation between the observed and predicted behavior phenotypes) was used as a metric for prediction accuracy.

Empirical data quality control

Manual quality control was performed to exclude extreme outliers from the baseline data of the ABCD Study. Specifically, participants were included in the analysis if they had (1) fMRI data not collected by Philips scanners due to incorrect post-processing in the ABCD 2.0.1 Release, (2) two fMRI runs for the nBack and the SST task, (3) vertexwise cortical data available at the time of analysis, (4) hemispheric mean beta-weights within two standard deviation of the whole sample mean for each task, (5) at least 200 degrees of freedom over the two runs, (6) met task-specific performance criteria (participants with very poor performance accuracy or slow reaction times were excluded), and (7) have complete information of the following demographic variables: age, sex, parental marital status, highest level of parental education, household income, self-reported race and ethnicity, and MRI scanner ID. All imaging and behavioral phenotypes were pre-residualized by age and categorical demographic variables including sex, parent marital status, highest level of parental education, household income, self-reported race and ethnicity, and MRI scanner ID. There were a large number of participants with missing income data (N = 1018). For these participants, imputed income was calculated by taking the median income level across participants from the same data collection site. Descriptive

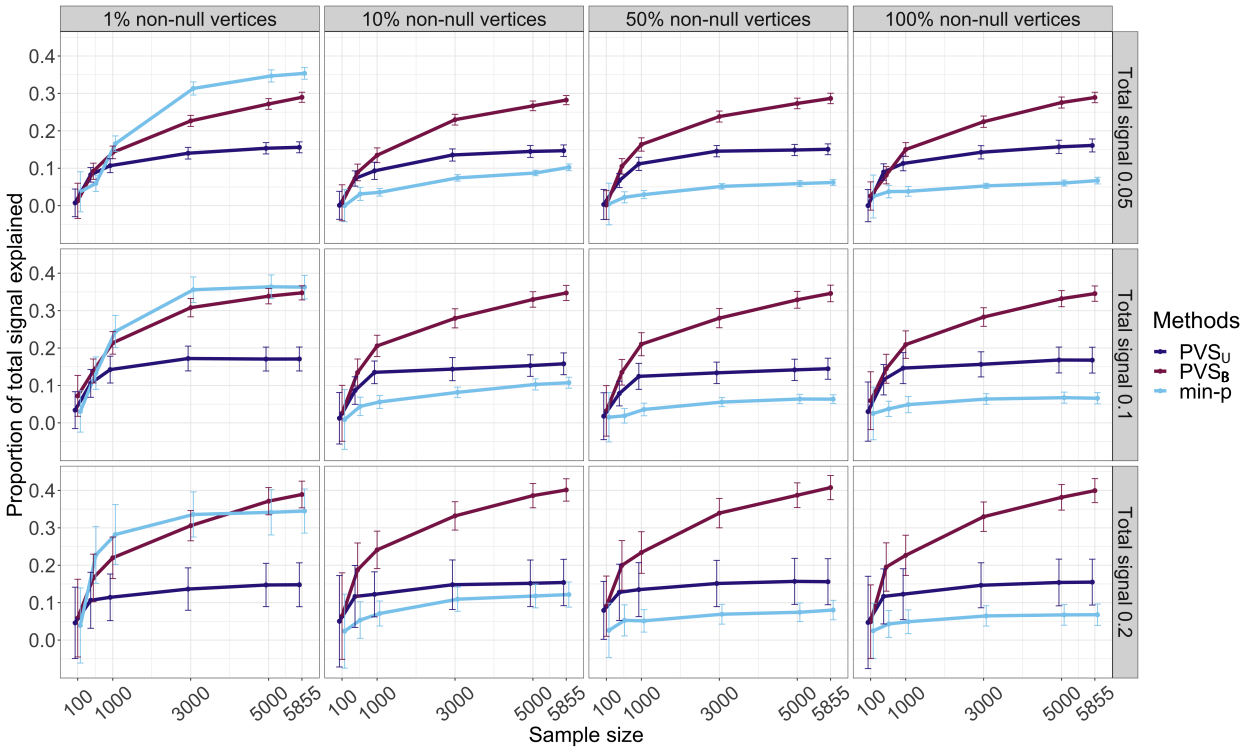
characteristics of the participants included in the empirical data analyses were reported in Supplementary Table 1.

Multivariate statistical methods implementation

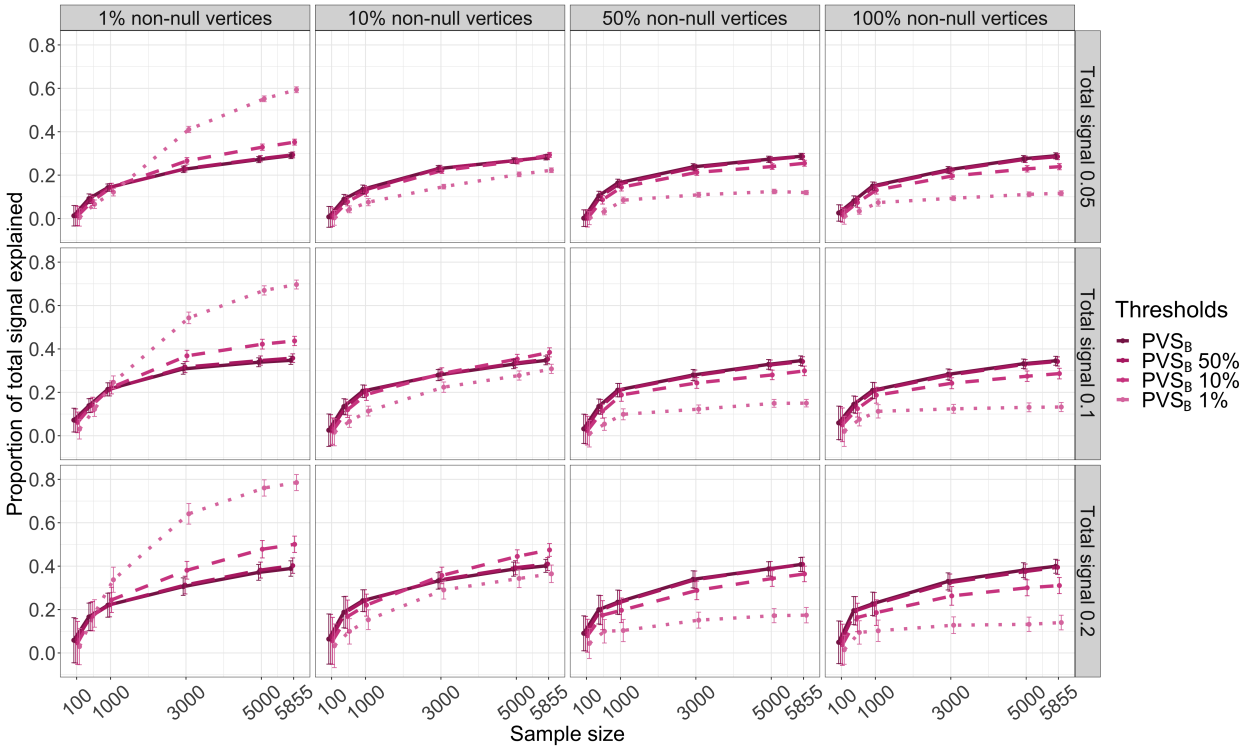
The out-of-sample predictive performance of LASSO, random forest, and support vector regression (SVR) on the four empirical brain-behavior associations was estimated using 10-fold cross validations in MATLAB (2017a). Cross-validation schemes for each brain-behavior association was pre-computed to ensure that participants of the same household were grouped together in the same training or testing set. All training and testing sets were pre-residualized and standardized independently for age, sex, race, ethnicity, household income, parental education, household marital status, and scanner ID.

LASSO was implemented with the *lasso* command with *L1* shrinkage. The regularization hyperparameter, lambda, was generated using the default setting in MATLAB (a geometric sequence of 100 values) and was selected using 10-fold cross validation within the training set. The largest lambda value that yielded the mean squared error (MSE) one standard error of the minimum MSE was selected, and was subsequently applied to the whole training set for parameter estimation. The parameter estimates were subsequently applied to the test set for behavioral prediction. Random forest was implemented with *TreeBagger*. The ensemble of one hundred regression trees trained on the training set were applied to the test set for prediction. For the SVR implementation, the training model was estimated using *fitrsvm* with linear kernel and applied to the test set for prediction.

2.9 Supplementary Figures & Tables



Supplementary Figure 2.1 The PVS_B demonstrated superior predictive performance than the PVS_U and the min-p across most simulated scenarios. The proportion of total explainable signal recovered by the PVS_U, the PVS_B and min-p, with 95% confidence interval, was plotted against the simulated sample size at four levels of proportion of non-null vertices, shown in columns, and three levels of total explainable signal shown in rows. The PVS_B showed comparable to superior predictive performance compared to the PVS_U .and min-p across at most simulated scenarios. The advantage of the PVS_B became more prominent with increased sample sizes and greater total explainable signals. However, the PVS_B suffered from including all vertices for prediction when the predictive signal in the brain was extremely sparse, i.e. 1% non-null vertices.



Supplementary Figure 2.2 The signal sparsity of the simulated brain-behavior associations can be captured by thresholding the PVS_B at varying statistical thresholds. The proportion of total signal explained by the PVS_B, the PVS_B 50% (PVS_B thresholded at 50% most significant vertices), the PVS_B 10%, and the PVS_B 1%, with 95% confidence interval, was plotted against the simulated sample size at four levels of proportion of non-null vertices, shown in columns, and three levels of total explainable signal shown in rows. The relative advantage of the PVS_B and the thresholded PVS_{BS} varied as a function of the signal sparsity level of the simulated brain-behavior association. When the true signal was not sparse, corresponded to the 50% and 100% non-null vertices conditions, the PVS_B showed comparable to superior predictive performance compared to the thresholded PVS_{BS}. As the true signal of the brain became more sparse, superior performance was achieved by thresholding the PVS_B at a statistical threshold corresponding to the signal sparsity level.

Supplementary Table 2.1 Descriptive characteristics of the ABCD Study participants included for each of the four empirical brain-behavior associations. This table showed the demographic decomposition of participants included in the nBack-TC, SST-SSRT, nBack-SSRT, and SST-TC association. The mean and the standard deviation (in brackets) of the continuous variable, age, was shown. For the categorical variables (sex, household income, parental education, marital status, family relations, and MRI scanner), the mean and the percentage of participants within each level of the categorical variable were shown.

	nBack-TC	**SST-SSRT**	**nBack-SSRT**	**SST-TC**
N	5855	6152	5850	6048
Age (mean (SD))	126.56 (9.63)	126.50 (9.66)	126.56 (9.61)	126.54 (9.67)
Sex = M (%)	2977 (50.8)	3113 (50.6)	2970 (50.8)	3068 (50.7)
Household income (%)				
[<50K]	438 (31.5)	467 (32.3)	430 (31.0)	458 (32.3)
[>=100K]	567 (40.8)	600 (41.5)	576 (41.6)	584 (41.2)
[>=50K & <100K]	385 (27.7)	380 (26.3)	379 (27.4)	376 (26.5)
Parental education (%)				
< HS Diploma	78 (5.1)	76 (4.8)	77 (5.1)	75 (4.9)
Bachelor	346 (22.8)	360 (22.8)	349 (23.1)	353 (22.8)
HS Diploma/GED	168 (11.1)	199 (12.6)	163 (10.8)	197 (12.7)
Post Graduate Degree	501 (33.1)	512 (32.4)	500 (33.1)	501 (32.4)
Some College	422 (27.9)	432 (27.4)	422 (27.9)	420 (27.2)
Marital status = yes (%)	994 (65.6)	1031 (65.3)	994 (65.8)	1013 (65.5)
Self-declared race/ethnicity (%)				
Asian	122 (2.1)	123 (2.0)	120 (2.1)	123 (2.0)
Black	654 (11.2)	731 (11.9)	652 (11.1)	714 (11.8)
Hispanic	1140 (19.5)	1188 (19.3)	1131 (19.3)	1166 (19.3)
Other	564 (9.6)	590 (9.6)	573 (9.8)	575 (9.5)
White	3375 (57.6)	3520 (57.2)	3374 (57.7)	3470 (57.4)
Family relation (%)				
Singleton	4587 (88)	4772 (87.5)	4593 (88.1)	4695 (87.5)
2 sublings	613 (11.8)	669 (12.3)	609 (11.7)	657 (12.2)
3 sublings	14 (0.2)	14 (0.2)	13 (0.2)	13 (0.3)
Scanner ID (%)				
HASH03db707f	66 (4.4)	70 (4.4)	66 (4.4)	70 (4.5)

Supplementary Table 2.1 Descriptive characteristics of the ABCD Study participants included for each of the four empirical brain-behavior associations. Continued.

HASH11ad4ed5	31 (2.0)	27 (1.7)	31 (2.1)	26 (1.7)
HASH1314a204	100 (6.6)	112 (7.1)	100 (6.6)	110 (7.1)
HASH311170b9	53 (3.5)	55 (3.5)	53 (3.5)	53 (3.4)
HASH31ce566d	4 (0.3)	2 (0.1)	4 (0.3)	2 (0.1)
HASH3935c89e	205 (13.5)	210 (13.3)	204 (13.5)	209 (13.5)
HASH48f7cbc3	2 (0.1)	3 (0.2)	2 (0.1)	3 (0.2)
HASH4b0b8b05	31 (2.0)	38 (2.4)	31 (2.1)	38 (2.5)
HASH4d1ed7b1	65 (4.3)	69 (4.4)	69 (4.6)	66 (4.3)
HASH5ac2b20b	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HASH5b0cf1bb	150 (9.9)	147 (9.3)	147 (9.7)	147 (9.5)
HASH5b2fcf80	42 (2.8)	44 (2.8)	43 (2.8)	43 (2.8)
HASH65b39280	101 (6.7)	100 (6.3)	101 (6.7)	100 (6.5)
HASH69f406fa	38 (2.5)	40 (2.5)	38 (2.5)	40 (2.6)
HASH6b4422a7	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HASH7911780b	48 (3.2)	58 (3.7)	50 (3.3)	53 (3.4)
HASH7f91147d	7 (0.5)	7 (0.4)	7 (0.5)	7 (0.5)
HASH96a0c182	63 (4.2)	75 (4.7)	66 (4.4)	68 (4.4)
HASHa3e45734	50 (3.3)	52 (3.3)	51 (3.4)	49 (3.2)
HASHb640a1b8	100 (6.6)	111 (7.0)	101 (6.7)	106 (6.9)
HASHc3bf3d9c	53 (3.5)	47 (3.0)	51 (3.4)	47 (3.0)
HASHc9398971	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HASHd422be27	79 (5.2)	79 (5.0)	77 (5.1)	78 (5.0)
HASHd7cb4c6d	60 (4.0)	49 (3.1)	59 (3.9)	49 (3.2)
HASHdb2589d4	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HASHe3ce02d3	19 (1.3)	27 (1.7)	18 (1.2)	27 (1.7)
HASHe4f6957a	140 (9.2)	152 (9.6)	134 (8.9)	150 (9.7)
HASHe76e6d72	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
HASHfeb7e81a	8 (0.5)	5 (0.3)	8 (0.5)	5 (0.3)

Supplementary Table 2.2 The out-of-sample variance explained of the cortical and the whole brain phenotypes, estimated with all multivariate statistical methods, was reduced when more rigorous covariate treatment was applied. For each multivariate model, the out-of-sample predictive performance of the cortical and the whole brain imaging phenotypes was shown. While the whole brain imaging phenotypes (vertexwise cortical data and subcortical ROIs) failed to improve the predictive performance of brain-behavior associations, the prediction accuracy of all multivariate methods was inflated when demographics and socioeconomic variables were not properly controlled for, demonstrating the importance of covariate control in multivariate prediction.

Covariates	Imaging Phenotype	Behavioral Phenotype	PVSB		RandomForest		LASSO		SVR	
			Cortex	Whole Brain	Cortex	Whole Brain	Cortex	Whole Brain	Cortex	Whole Brain
Model 1: Age + Gender	nBack	TC	18.16	18.27	17.64	18.22	17.50	17.15	5.14	4.84
	nBack	SSRT	0.60	0.59	0.32	0.44	0.08	0.04	0.13	0.06
	SST	TC	1.27	1.23	1.68	1.63	0.15	0.25	0.38	0.55
	SST	SSRT	11.57	11.64	7.89	7.91	8.97	8.85	3.60	3.47
Model 2: Model 1 + Race + Ethnicity	nBack	TC	14.22	14.25	13.53	13.80	13.20	12.89	3.53	3.75
	nBack	SSRT	0.88	0.87	0.37	0.31	0.03	0.00	0.19	0.19
	SST	TC	0.81	0.82	1.10	0.97	0.23	0.12	0.24	0.29
	SST	SSRT	11.23	11.33	7.46	8.03	8.39	8.56	3.52	3.68
Model 3: Model 2 + household income + parental education	nBack	TC	11.58	11.60	10.88	10.93	10.42	10.48	2.25	2.78
	nBack	SSRT	0.76	0.75	0.12	0.21	0.13	0.10	0.07	0.03
	SST	TC	0.46	0.47	0.38	0.58	0.00	0.00	0.08	0.22
	SST	SSRT	10.64	10.75	8.03	7.81	8.70	9.01	3.45	3.63

Chapter 3 Task fMRI paradigms may capture more behaviorally relevant information than resting-state functional connectivity

3.1 Introduction

An important aim of cognitive neuroscience is to understand how individual differences in behavioral attributes are associated with brain structure and function. With the availability of large neuroimaging datasets, recent work has pivoted towards building models that predict current or future behavior based on neuroimaging measures (Gabrieli, Ghosh, Whitfield-Gabrieli, 2015; Varoquaux & Poldrack, 2019; Finn & Rosenberg, 2021). Such predictive modeling approaches allow us to estimate better the degree to which behavioral differences are associated with individual differences in brain structure or function.

Trait differences can be predicted by individual differences in functional connectivity (FC), which measures the correlation of the BOLD response across regions of interests (ROIs) across brain regions by calculating the pairwise correlations of fMRI time series (Speer et al., 2021; Zhang et al., 2021). FC patterns are unique to an individual (Finn et al., 2015; Gratton et al., 2018), relatively stable across different mental states (Cole et al., 2014; Finn et al., 2015; Gratton et al., 2018), and sensitive to phenotypic differences including age (Dosenbach et al., 2010; Nielsen et al., 2019), cognitive abilities (Sripada et al., 2019, Moutoussis et al., 2021, Zhang et al., 2021; Chen et al., 2022), and mental health outcomes (Challis et al., 2015, Kim et al., 2016, Thomas et al., 2020; Chen et al., 2022).

FC is often estimated during resting-state fMRI acquisitions where participants are not engaged in a particular task but are simply instructed to either close their eyes or fixate on a crosshair and stay still. While resting-state fMRI has become the most common paradigm used for correlating FC patterns with behavioral traits or conditions, there is increasing evidence that

rest may not always be the optimal condition to elicit FC patterns that are most relevant to differences in behavioral phenotypes in a particular domain (Rosenberg et al., 2016; Greene et al., 2018; Jiang et al., 2019; Finn, 2021). Naturalistic tasks or traditional fMRI tasks may have more utility for the prediction of trait or state differences as they can elicit cognitive states that are directly relevant to the behavioral domain of interest (Finn et al., 2017).

Direct comparisons between resting-state FC and task-fMRI FC suggest that the latter is better at predicting both fMRI attention task performance and trait measures of attention function (Rosenberg et al., 2016), measures of general cognitive ability (Greene et al., 2018; Elliot et al., 2019) and reading comprehension (Jiang et al., 2020). A similar advantage has been shown for more passive task fMRI with naturalistic paradigms such that FC during movie-watching paradigms outperformed resting-state FC in predicting individual differences in cognitive task performance and emotional health (Finn & Bandettini, 2021).

Why might FC patterns derived from task and naturalistic paradigms be more predictive of trait differences than FC patterns derived from rest? Finn and colleagues (Finn et al., 2017; Finn & Bandettini, 2021) proposed that task fMRI and naturalistic paradigms are better candidates than resting-state for the study of behavioral differences because tasks are tailored to engage a particular behavioral domain. Like a cardiac stress test where the heart's ability to respond to external stress is measured by inducing stress in a controlled environment, fMRI tasks and naturalistic paradigms can introduce cognitive and emotional challenges to simulate brain activity. It follows that an fMRI paradigm that engages the behavioral or cognitive processes involved in the behavioral phenotype of interest is likely to amplify brain-behavior relationships (Greene et al., 2020; Greene et al., 2018). In other words, the greater behavioral relevance of the FC derived from task fMRI paradigms may be attributable to the task effects.

Previous studies examining FC during fMRI tasks differ in whether they retain the task effects in the fMRI time course for FC estimation (Table 3.1). With an explicit task design, the observed time course during a fMRI task can be decomposed into the part that is explained by the task, estimated by the task model fit, and the residual. If the task effect is retained (Vanderwal et al., 2017; Greene et al., 2018; Gao et al., 2019), the FC estimates, which we label **task-based FC**, capture the FC of the original and complete task fMRI time series. If the task effect, estimated by the task model fit, is removed (Arfanakis et al., 2000; Fair et al., 2007), the FC measures, which we label the **task-model residual-FC**, capture the FC of the component of the task fMRI time series that is not explained by the task design. The task-model-residual FC has also been called “pseudo resting-state connectivity" (Jurkiewicz et al., 2018), “task FC” (Cole et al., 2014), “task-based FC” (Cole et al., 2019), and "background connectivity" (Al-Aidroos et al., 2012) in the literature. Because of its task-invariant nature, task- model-residual FC patterns have indeed been shown to resemble resting-state FC patterns (Jurkiewicz et al., 2018; Cole et al., 2019) and are predictive of behavioral differences across individuals (O'Halloran et al., 2018; Varangis, Habeck & Stern, 2020). If the task-model-residual FC does capture the same functional brain organization as resting-state FC as previous studies suggest (Jurkiewicz et al., 2018; Cole et al., 2019), the reported superior behavioral prediction of the task-state FC over resting-state FC may be attributable to the task effects that were removed from the task-model-residual FC. We can manipulate data from fMRI tasks to study the behavioral sensitivity of the FC pattern of the task effects. By estimating the FC of the task model fit, which we call the **task-model-fit FC**, we can directly assess and compare its behavioral relevance against the task-model-residual FC, the task-based FC, and the resting-state FC. These comparisons generate new hypotheses on the source of behavioral relevance in the

task fMRI data and can provide additional information to guide the optimization of fMRI paradigms for the investigation of behavioral phenotypes.

Table 3.1 Glossary for fMRI measures used in this study.

FMRI paradigms	**Names**	**Definition**	**Names in other studies**
Resting-state fMRI	Resting-state FC	Pairwise correlation of fMRI activities at rest	
Task-related FC measures	Task-state FC	Pairwise correlation of the complete preprocessed task fMRI time series.	"Task-based FC" (Greene et al., 2018; Gao et al., 2019)
	Task-model-fit FC	Pairwise correlation of task-model-fit time series which is the task fMRI time series component explained by the task design. The task-model-fit time series is derived by multiplying the task design matrix by the beta estimates of the task condition regressors and their temporal derivative.	
	Task-model-residual FC	Pairwise correlation of the task-model-residual time series which is the task fMRI time series component that can't be explained by the task design. The task-model-residual time series is derived by subtracting the task-model-fit time series from the preprocessed task fMRI time series.	"Pseudo resting-state connectivity" (Jurkiewicz et al., 2018) "Task FC" (Cole et al., 2014) "Task-state FC" (Cole et al., 2019) "Background connectivity" (Al-Aidroos et al., 2012)
	Task-predicted FC	The task fMRI time series component that is predicted by the task design. Derived from the preprocessed fMRI time series by multiplying the task contrast betas estimates to the preprocessed fMRI time series.	--

In this study, we leveraged the large sample of the Adolescent Brain Cognitive Development (ABCD) Study ® and compared the behavioral prediction performance of resting-state FC to the task-model-fit FC, task-model-residual FC, and task-based FC derived from the Emotional N-back (nBack) task, the Stop Signal Task (SST), and the Monetary Incentive Delay (MID) task. We evaluated the out-of-sample prediction performance of each FC measure on two behavioral measures. The trait-like behavioral measure of interest was a measure of general

cognitive performance, the total composite cognitive score of the NIH Cognition Toolbox. As an example of a more proximal, state-sensitive, behavioral measure we chose a behavioral inhibition measure derived from the SST fMRI task, the stop-signal reaction time (SSRT). The behavioral prediction of the task model parameters, the beta estimates of the task condition regressors, was also estimated and contrasted with all task-derived FC measures. We also quantified how the prediction performance of FC measures changed with the amount of usable data and across sociodemographic variables, which are known to be associated with individual variability in cognitive (Korous et al., 2020) and brain outcomes (Farah, 2018; Taylor et al., 2020).

3.2 Methods

Participants

The ABCD Study is a longitudinal neuroimaging study that tracks brain and behavioral development of 11,880 children starting at 9 and 10 years old. The ABCD study used school-based recruitment strategies to create a demographically and ethnically diverse cohort (Garavan et al., 2018) with an embedded twin cohort and many siblings. Informed consent was obtained from parents/caretakers and ass was obtained from the children. Extensive descriptions of the recruitment, collection, and processing of the fMRI and the behavioral data of the ABCD study can be found in prior publications (Gavaran et al., 2018; Casey et al., 2018; Hagler et al., 2019). Participants with complete data across all the behavioral measures and covariates of interest were included in the analyses. To ensure accurate characterization of the FC matrices, participants were required to have at least 50% of usable data for each of the two runs of each fMRI task, and for each of the four resting state fMRI runs. The nBack task had the least number of participants that met the inclusion criteria (n = 3034). In order to match the number of participants across

fMRI acquisitions for the behavioral prediction analysis, we randomly selected 3034 participants from each of the other fMRI acquisitions. Around 50% of the participants are shared between the final sample of each acquisition. The additional inclusion criteria and their effect on sample size is shown in Supplementary Table 3.1.

Behavioral measures

Here, we describe the behavioral measures used in the present study. The full neurocognition battery for the ABCD Study is detailed elsewhere (Luciana et al., 2018). The NIH Toolbox Cognition Battery measures a range of cognitive domains that show substantial development during childhood and adolescence. It consists of seven subtests, including measures of vocabulary size (Picture Vocabulary Task), single word reading ability (Oral Reading Task), rapid visual processing (Pattern Comparison Processing Speed Test), working memory capacity (List Sorting Working Memory Test), episodic memory (Picture Sequence Memory Test), attention and inhibitory control (Flanker Task), and cognitive flexibility (Dimensional Change Card Sort Task). The composite measure of the NIH Toolbox Cognitive Battery, the Total Composite Score is an arithmetic average of the 7 subtests summarizing the cognitive performance of an individual across the different cognitive domains. The age-uncorrected score of the composite measure, the total composite cognition score, was used as a primary behavioral outcome of this study. In the ABCD Study, participants perform the Stop Signal Task (SST) during fMRI scans. In this task, participants are instructed to inhibit a prepotent motor response to a Go Stimulus in response to a stop signal. A tracking algorithm varies the interval between the onset of the Go stimulus and the onset of the Stop stimulus (the Stop Signal Delay) based on individual performance. The Stop Signal Reaction Time (SSRT) quantifies the speed of the inhibitory process during the SST task, such that lower SSRT reflects more efficient response

inhibition. The SSRT was calculated by subtracting participants' mean stop signal delay (SSD) from their mean reaction time during the SST fMRI task. We chose the SSRT as a second behavioral outcome because it measures a specific cognitive process, in contrast to the general cognitive abilities assessed by the total composite cognition score. In addition, the measure was derived directly from performance during the task fMRI session enabling us to assess links between task performance and the miscellaneous FC measures obtained during that task.

Resting-state and task fMRI paradigms

The neuroimaging paradigms and acquisition parameters are detailed elsewhere (Casey et al., 2018), so a brief overview is provided here. Four 5-minute resting-state fMRI runs were acquired during which participants were instructed to fixate on a crosshair. Three task fMRI acquisitions were completed after the resting-state fMRI, with two runs of each of the following tasks: Emotional N-back task (nBack), Stop Signal Task (SST), and Monetary Incentive Delay Task (MID). The order of the tasks was counterbalanced across participants. These tasks have been shown to elicit anticipated patterns of brain activation in the ABCD Study baseline data consistent with previous literature (Chaarani et al., 2021).

The nBack engages the neural correlates of working memory and emotional regulation processes. To engage working memory, the task includes 0-back and 2-back conditions, presented in a block design. For the 2-back condition, participants were instructed to indicate with a button press whether the current stimulus matched the stimulus presented 2 trials back. For the 0-back condition, a target stimulus was presented at the beginning of the block and participants were instructed to press the button when they saw the target. To engage emotion regulation, the task stimuli included happy faces, fearful faces, neutral faces, and places, presented serially.

The SST engages the neural correlates of impulsivity and inhibitory control. In an event-related design, participants were instructed to indicate the direction of a leftward or rightward pointing arrow as quickly as possible. In 16.67% of the trials, the arrow was followed by a stop signal represented as an upward arrow, and participants were instructed to withhold their response. A tracking algorithm that varied the onset of the stimulus and the onset of the stop stimulus (the stop signal delay, SSD) was implemented to ensure approximately 50% successful and 50% unsuccessful stop trials.

The MID probes the neural correlates of reward processing. For each trial, participants could either win money, lose money, or earn nothing. Wins and losses were further subdivided into small or large amounts. At the start of each trial, participants were prompted with an incentive cue of five possible trial types (win $0.20, Win $5, Lose $5, Lose $0.20, $0-no money at stake) followed by a jittered anticipation period, during which participants fixated on a crosshair. Next, a target appeared to which participants made their button response. The trial ended with positive or negative feedback to inform participants about their performance.

Image acquisition and processing

Task and resting-state MRI acquisition and preprocessing

The ABCD MRI data were collected across 21 research sites using GE 750, Siemens Prisma, and Philips Achieva and Ingenia 3T scanners. Scanning protocols were harmonized across sites. The full details of the ABCD imaging acquisition and preprocessing protocols were described in Casey et al. (2018) and Hagler et al. (2019). Briefly, T1w sMRI images (1mm isotropic) were acquired with a 3D T1w inversion prepared RF-spoiled gradient echo scan, and fMRI acquisitions (rest and task) were collected with multiband EPI with slice acceleration factor 6 (2.4 mm isotropic, TR = 800ms). The preprocessing steps for fMRI data included (i)

head motion correction, (ii) B0 distortion correction, (iii) gradient warping correction, (iv) within-scan motion correction, and (v) registration to T1w structural images. Initial frames (Siemens and Philips scanners: 8 TRs; GE DV25: 5 TRs; GE DV26: 16 TRs) were removed from the preprocessed task fMRI time course. Motion estimates were filtered to remove the effect of respiratory signals (Fair et al., 2018). The preprocessed time courses were normalized and sampled onto the cortical surface for each participant. Average time courses were calculated for a functionally defined parcellation scheme (Gordon et al., 2016) sampled from the atlas-space to individual subspace, and anatomically-defined subcortical ROIs (Fischl et al., 2002).

Task model parameters, task-based fMRI time series, task-model-fit, and task-model-residual time series estimation

The task effects were estimated at the participant level using a general linear model (GLM) that included the stimulus timing for each task condition (Hagler et al., 2019) and the temporal derivative to capture any task related changes in the fMRI time course that is not captured by our task model. The GLM modeled each task condition with a bivariate gamma function and its first temporal derivative along with 4 nuisance regressors for baseline shifts and cubic trends and 12 regressors for the six motion estimates and their temporal derivatives. For the GLM estimation, time points with framewise displacement (FD) greater than 0.9 mm were censored (Siegel et al. 2014 HBM). For the behavioral prediction response of the task model parameters, both the beta estimates of the task condition regressors and the temporal derivative were included as predictors.

The task-based time series was the task fMRI time series after preprocessing. The task-model-fit time series was the component of the preprocessed task fMRI time series that was explained by the task design and was calculated by multiplying task design matrix to the beta

estimates of task condition regressors and their temporal derivative. The task-model-residual time series was the component of the preprocessed task fMRI time series that was not explained by the task, calculated by subtracting the task-model-fit time series from the preprocessed task-based time series.

FC estimation

Several additional preprocessing steps were applied to the resting-state and task fMRI time series before the estimation of FC to reduce spurious signals that are unlikely to reflect neuronal activation. These steps included (1) censoring and residualization and removal of signals associated with cerebral white matter, ventricles, white matter, and head motion estimates and their squares and derivatives (Power et al., 2014; Satterthwaite et al., 2012), (2) motion regression where frames with FD over 0.3mm were excluded (Power et al., 2014), and (3) band-pass filtering (0.009 and 0.08 Hz) (Hallquist et al., 2013). Additional motion censoring was applied to exclude the following time points: time points with FD over 0.2mm, time points that were outliers with respect to the spatial variation across the brain, and time periods with less than 5 contiguous, sub-threshold time points. Average time courses were calculated for 333 cortical ROIs (Gordon et al., 2016) and 19 subcortical ROIs (Fischl et al., 2002) for each run and were concatenated. Pearson correlation was applied to calculate the pairwise correlation of these 352 ROIs. The r-to-z transformed correlation matrix formed the FC estimate of each time series.

Statistical analysis

Behavioral prediction algorithm

A nested 10-fold cross validation scheme was used to estimate the out-of-sample prediction performance of each set of fMRI measures. Within each training set, the mass univariate beta estimates between each fMRI measure and a behavior were estimated using the

Fast Efficient Mixed Effects Analysis (FEMA; Fan et al., 2021) where a general linear mixed effects model was estimated at each voxel or ROI. Compared to the traditional general linear models, FEMA explicitly adjusts for the effects of the nested family structure in the ABCD data and the covariates of no interest. The following sociodemographic and imaging acquisition variables were included in the FEMA models as covariates: age, biological sex, top 10 genetic PCs, highest parental education, household income, scanner ID (MRI device serial number) and software version. Mean framewise displacement (FD) and the number of usable time points were used as additional covariates for FC measures. A separate analysis was conducted without the inclusion of sociodemographic variables as covariates to probe the shared impact of sociodemographic variables on the imaging and the behavioral measure. For this analysis, only scanner ID and software version were used as covariates, along with mean FD and the number of usable time points for FC measures.

For behavioral prediction, the mass univariate beta estimates from FEMA were entered into a singular-value decomposition (SVD) based prediction method to predict the behavioral outcome of the unseen, test-set participants. Similar to our previous method, the Bayesian polyvertex score (PVSB, Zhao et al., 2021), the SVD-based prediction method applies shrinkage to the mass univariate beta estimates to improve out-of-sample prediction performance. The shrinkage factor was derived separately for each brain-behavior association with a 5-fold cross validation nested within each training set. Within each nested training set, SVD was applied to the imaging measure pre-residualized for sociodemographic covariates to approximate the covariance structure of the mass univariate beta estimates. From the SVD result, the top k singular vectors and their corresponding singular values were used to calculate a shrinkage factor that was used to reweight the mass univariate beta estimates from FEMA. One hundred k values

were selected at equal distances between 1 and the dimension of the predictor space. The best performing *k* value was selected as the shrinkage factor for the full training set. The reweighted mass univariate estimates were then applied to the test set imaging data to calculate the predicted behavioral score for each test set participant.

The predicted behavioral score summarizes the variability in the behavioral outcome that is attributable to individual differences of the imaging measure. Squared correlation between the predicted and the observed behavioral score was used as the metric for out-of-sample behavioral prediction performance of each imaging measure. The ninety-five percent confidence interval of the behavioral prediction performance of each fMRI measure was generated with bootstrap resampling (Elliot et al., 2019) using the ci_cor function (confintr package) in R. The predicted behavioral scores were also used in subsequent analyses to probe the shared and unique behavioral variance explained by different FC measures.

The effect of scan length on behavioral prediction performance

To investigate how the behavioral prediction performance of FC measures was affected by the amount of available data, we quantified the degree to which the behavioral prediction performance of FC measures varied as a function of scan length. For fMRI tasks, the task-based, task-model-fit, and task-model-residual FC estimates were estimated separately with one run and with two runs of task fMRI data, and the change in behavioral prediction performance from one run to two runs was quantified with a ninety-five percent confidence interval using bootstrap resampling (ci_cor function from confintr package in R). For resting-state fMRI, resting-state FC was estimated with one run, two runs, three runs and four runs of data, and the change in behavioral prediction performance of each resting-state FC estimate was quantified in the same way as the task fMRI. This allowed us to estimate the prediction performance of resting-state FC

as if 5 mins, 10 mins, 15 mins, or 20 mins of resting-state fMRI data had been acquired. To eliminate any confounds introduced by participants, the effect of scan length was assessed with the same cross validation scheme and the same set of participants for each fMRI paradigm.

Quantification of shared and unique behavioral variance explained by the task-predicted FC and the task-residualized FC

As the task-model-fit FC and task-model-residual FC were derived as complementary subcomponents of the same task fMRI time series, we examined if they contained unique information for behavioral differences by estimating their shared and unique behavioral variance explained. In this set of analysis, we used the predicted behavioral scores of each FC measure on each behavior as the predictor because they captured the prediction effects of FC measures on behaviors while reducing the predictor dimensionality to a single measure. We first estimated the out-of-sample behavioral prediction performance of the predicted behavioral scores of task-model-fit FC and of the task-model-residual FC individually with generalized additive mixed models (GAMMs) with sociodemographic factors as fixed effects covariates and family ID as random effects. These univariate models, with only one brain predictor in the model, gave us an estimate of the behavioral variance explained by each FC in isolation. Then, we estimated their total prediction effect by including the predicted behavioral scores of both FC measures as predictors in an augmented model. with sociodemographic factors as fixed effects covariates and family ID as random effects.

The unique variance explained by the task-model-fit FC (unique R^2 adjusted for task-model-residual FC) was calculated as the difference in R^2 between the univariate model with the predicted behavioral score of the task-model-residual FC as the only FC predictor and the augmented model with the predicted scores of both the task-model-fit FC and the task-model-

residual FC. The unique variance explained by the task-model-residual FC (unique R^2 adjusted for task-model-fit FC) was estimated as the differences in R^2 between the univariate model with the predicted behavioral score of the task-model-fit FC and the augmented model. The gamm4 (gamm4 package) function was used to perform GAMMs in R and the r.squaredGLMM (MuMIn package) function was used to estimate the behavioral variance explained (fixed effects pseudo-r-squared) of the fMRI predictors from GAMMs.

Quantification of shared and unique behavioral variance explained by the task-model-fit FC and the task model parameters

Both task-model-fit FC and task model parameters capture the task effects on brain activity. We assessed whether these two task effects measures explained unique behavioral variance by quantifying the shared and unique variance explained of the predicted behavioral scores of the task-model-fit FC and the task model parameters. An augmented model that included both measures was performed to estimate the total prediction effect of the task-model-fit FC and the task model parameters. The unique variance explained by the task-model-fit FC (unique R^2 adjusting for task model parameters) was estimated as the difference in R^2 between the augmented model and the univariate model with task model parameters. The unique variance explained by the task model fit (unique R^2 adjusting for task-model-fit FC) was estimated as the difference between the augmented model and the univariate model with task-model-fit FC. Family relatedness was modeled as a random effect and sociodemographic factors were used as fixed effects covariates for all the above-mentioned models.

The effect of sociodemographic factor adjustment on behavioral prediction performance

To understand how sociodemographic factor adjustment changes the behavioral prediction performance of fMRI measures, we reran the above behavioral prediction models

without the adjustment of sociodemographic factors and only including scanner ID, scanner software version, mean FD, and the number of usable timepoints as covariates in FEMA. The unadjusted mass univariate beta estimates of all FC and task model parameters were used to calculate the behavioral prediction performance of all fMRI measures without the adjustment of sociodemographic differences in our sample. The prediction performance of each fMRI measure with and without sociodemographic adjustment was compared.

Data Statement

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA).

3.3 Results

Task-predicted FC and task-state FC better predicted individual differences in behaviors than resting-state FC and task-residualized FC.

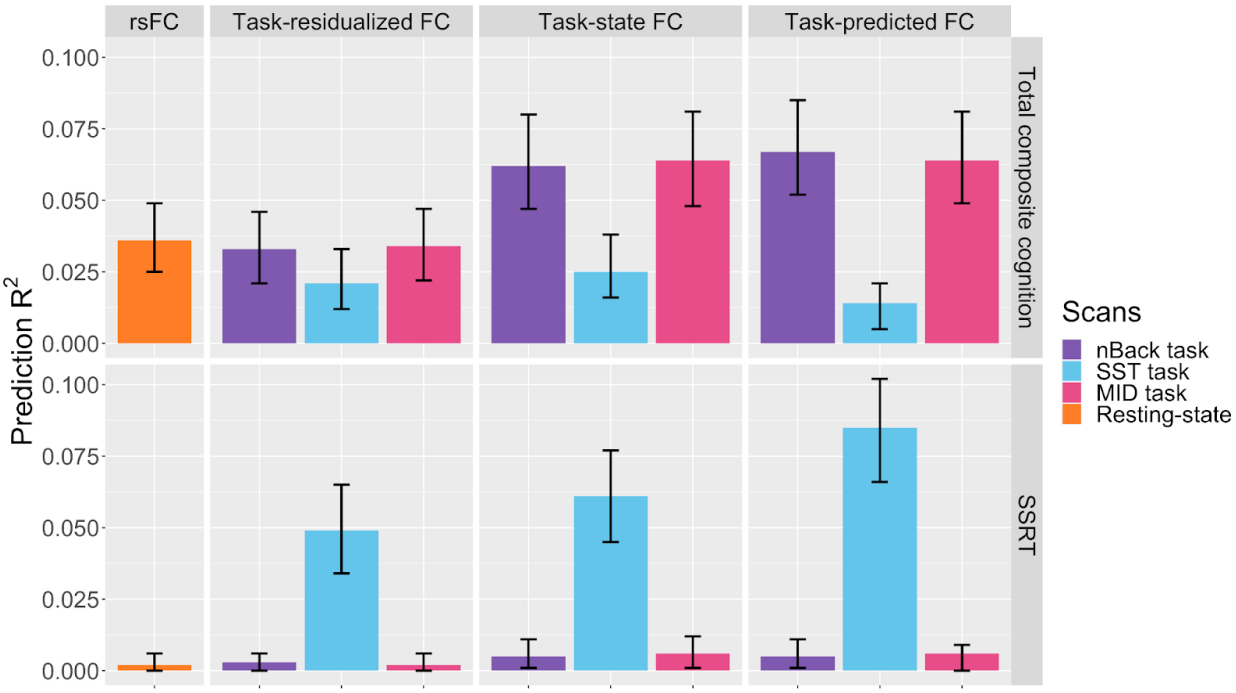


Figure 3.1 The task-state FC and task-predicted FC outperformed resting-state FC and task-residualized FC at predicting individual variabilities in total composite cognition and SSRT. For total composite cognition (top row), resting-state FC (rsFC; first column) and the task-model-residual FC measures (second column) showed similar behavioral prediction performance. Task-based FC (third column) and task-model-fit FC (fourth column) of the nBack and MID task, on the other hand, outperformed resting-state FC and task-model-residualFC explaining behavior differences in total composite cognition. For the SSRT (bottom row), only task-derived FC measures from the SST task were predictive. All SST task FC measures were predictive of the SSRT, while resting-state FC was not. Error bars show the ninety-five percent confidence intervals estimated with bootstrap resampling.

Prediction performance of resting-state FC and the three task-derived FC measures on individual differences in total composite cognition score and SSRT are shown in Figure 3.1. After adjusting for sociodemographic variables, resting-state FC explained 3.6% of the variance in total composite cognition. The nBack, SST, and MID task-model-residual FC estimates explained 3.3%, 2.1%, and 3.4% of the variance in total composite cognition, respectively.

Increased behavioral prediction was observed for the task-based FC and the task-model-fit FC derived from the nBack and the MID task. The nBack task-based FC and task-model-fit FC explained 6.2% and 6.7% of the variance in total composite cognition, and the MID task-

based FC and task-model-fit FC explained 6.4% and 6.4% of the variance in total composite cognition. We did not observe an increase in prediction for the SST task-based and the task-model-fit FC on total composite cognition (SST task-model-fit FC: R^2 = 1.4%; SST task-based FC: R^2 = 2.5%).

Only FC measures derived from the SST task were significantly predictive of the individual differences in SSRT. Among the SST task FC measures, we observed an advantage for the SST task-model-fit FC relative to the SST task-model-residual FC. The SST task-model-fit FC explained 9.5% of the variance in SSRT, while the SST task-model-residual FC explained 4.9%. We compared the FEMA z-score map of the SST task-model-residual FC on SSRT to the effect size map of the SST task-model-fit FC, the SST task-based FC, and the resting-state FC (Figure 3.2). The mass univariate beta estimates of the SST task-model-residual FC bore greater resemblance to the effect size map of the SST task-model-fit FC than to the resting-state FC, suggesting that the SST task-model-residual FC captured a similar predictive pattern as the SST task-model-fit FC.

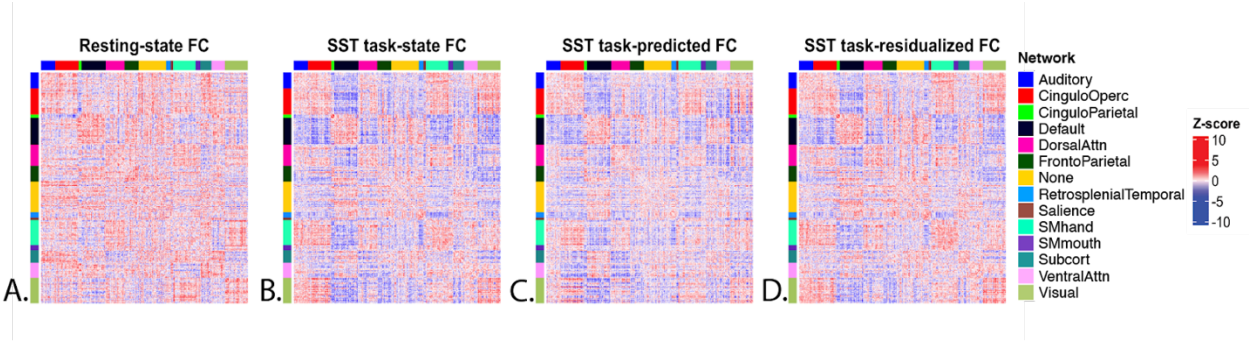


Figure 3.2 The effect size matrices of the SST task FC measures on SSRT were more similar to each other than to resting-state FC. 352 ROIs x 352 ROIs effect size matrices, organized by functional network, are shown. Each cell corresponds to the mass univariate z-score of each ROI pair on SSRT derived from FEMA analyses.

Increased scan length moderately improved behavioral prediction performance.

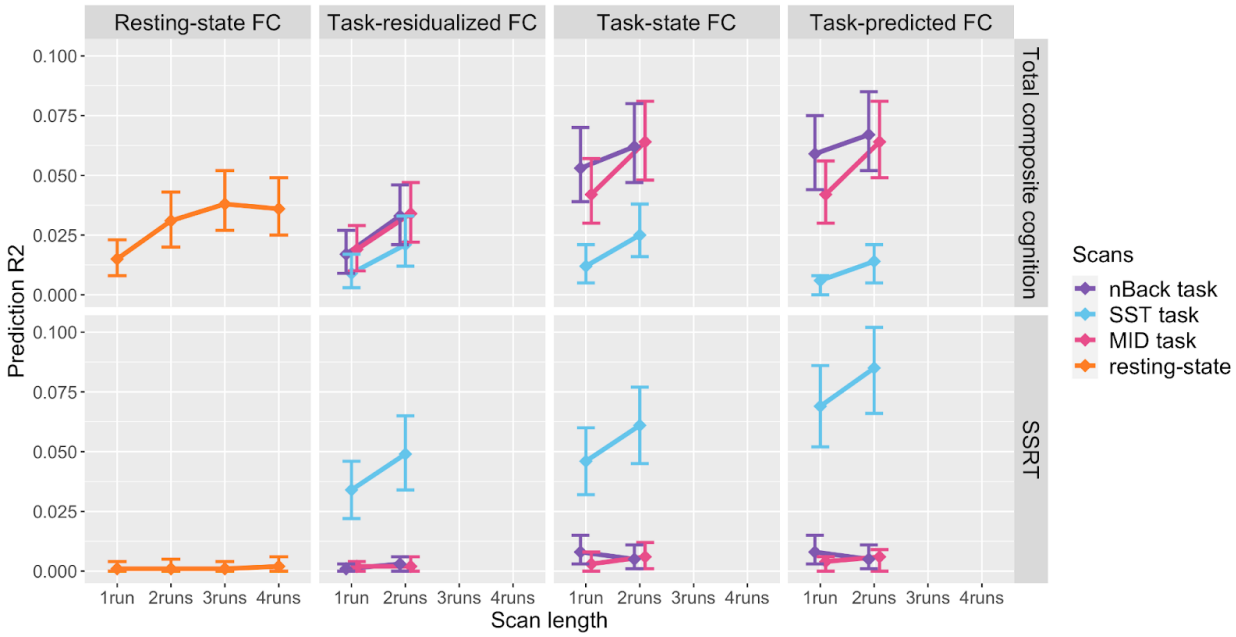


Figure 3.3 For the task-model-residual FC, task-based FC, and task-model-fit FC, a moderate increase in prediction of total composite cognition was observed when two runs of fMRI data were used to estimate the corresponding FC estimates for all three tasks (top row). The task FC measures showed an increase in prediction of the SSRT for two fMRI runs compared to one only for the SST task (bottom row). For resting-state FC, the prediction of total composite cognition increased from one run to two runs, but not beyond, and there was no change in prediction of the SSRT.

The prediction performance of all FC measures as a function of scan length is shown in Figure 3.3. For the task-model-residual FC, task-based FC, and task-model-fit FC measures, we observed a modest increase in prediction accuracy when two runs of data were used to estimate the FC matrix relative to one run. For resting-state FC, we observed a similar increase in prediction of total composite cognition (but not SSRT) when two runs of data were used relative to one, but the effect plateaued with increased number of scans.

Task-model-fit FC accounted for the behavioral variance predicted by the task-model-residual FC.

Table 3.2 The shared and unique variance explained, R^2, of the task-model-fit FC and the task-model-residual FC for each brain-behavior association.

fMRI tasks	Task-model-fit FC		Task-model-residual FC	
	R^2	**Unique R^2 adjusted for task-model-residual FC**	**R2**	**Unique R^2 adjusted for task-model-fit FC**
Behavior: Total composite cognition				
nBack	6.4%	4.1%	3.1%	0.8%
SST	1.1%	0.4%	2.0%	1.3%
MID	6.6%	4.1%	3.1%	0.6%
Behavior: SSRT				
nBack	0.4%	0.4%	0.2%	0.1%
SST	7.9%	4.5%	4.5%	1.1%
MID	0.2%	0.1%	0.3%	0.2%

Given that the task-model-fit FC and task-model-residual FC were derived from complementary subcomponents of the task fMRI time series, we examined whether these FC measures contributed unique information to behavioral prediction (Table 3.2) by quantifying the shared and unique variance explained by the predicted behavioral scores of the two FC measures. For the prediction of total composite cognition by the nBack and MID tasks, task-model-residual FC contributed minimal unique variance explained ($R^2 < 1\%$) after adjusting for task-model-fit FC. On the other hand, the nBack and MID task-model-fit FC each explained 4.1% variance in total composite cognition after adjusting for task-model-residual FC. Therefore, task-model-fit FC predicted unique behavioral variance, while task-model-residual FC did not. By contrast, SST task-model-fit FC did not contribute unique variance to predicting total composite cognition, while the SST task-model-residual FC uniquely explained 1.3% of the variance. For the SST-SSRT association, after adjusting for the shared behavioral variance explained, both the

SST task-model-fit FC and the SST task-model-residual FC predicted unique variance in SSRT (SST task-model-fit FC: unique R^2 = 4.5%; SST task-model-residual FC: unique R^2 =1.1%). We believe that the unique association between SSRT and the SST task-model-residual FC might be attributable to the insufficient modeling and removal of the SST task effect.

Task model parameters also exhibited a task-specific prediction advantage over the task-model-residual FC and explained both shared and unique behavioral variance relative to the task-model-fit FC.

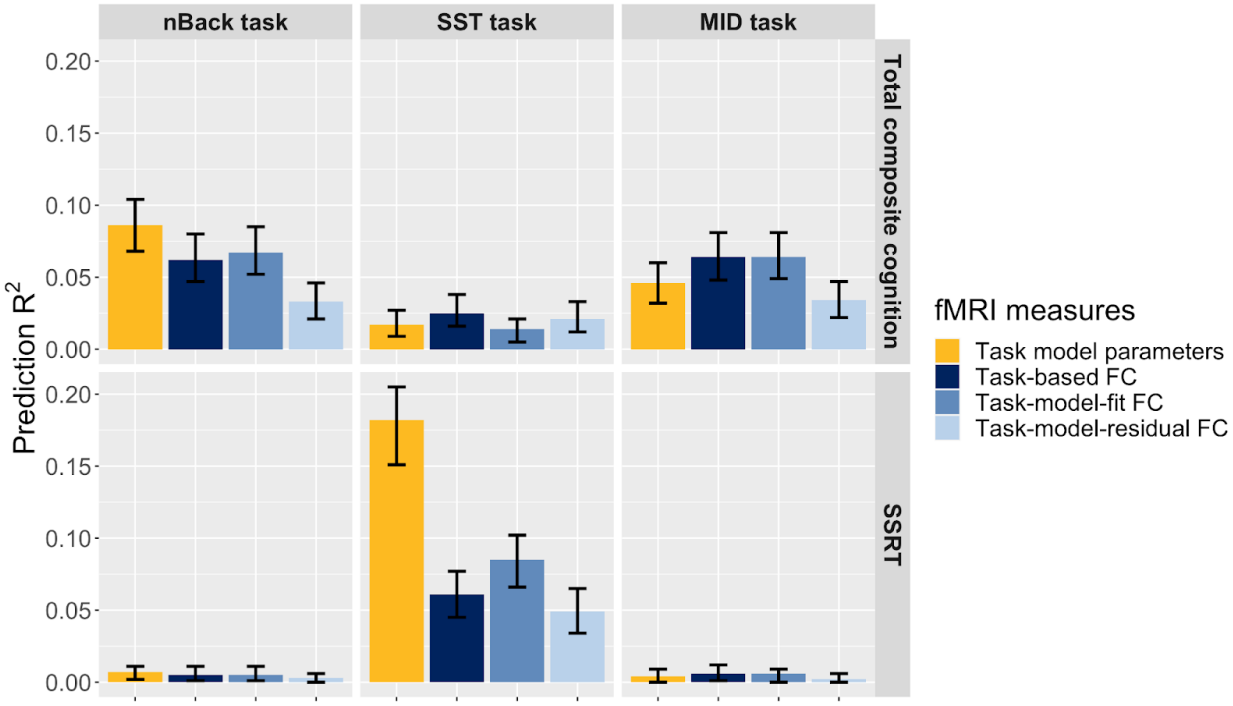


Figure 3.4 Task model parameters outperformed task-model-residual FC at predicting behavioral differences. Similar to the task-model-fit FC, task model parameters (yellow) demonstrated task-specific advantage over the task-model-residual FC in behavioral prediction. The nBack task model parameters outperformed task-model-residual FC at predicting individual differences in total composite cognition and the SST task model parameters outperformed SST task-model-residual FC at predicting variabilities in SSRT. With regard to the prediction performance of the task model parameters and the task-model-fit FC, the nBack and SST task model parameters outperformed task-model-fit FC for behavioral prediction while the MID task-model-fit FC outperformed task model parameters at predicting total composite cognition. Further investigations revealed that task model parameters and task-model-fit FC explained shared and unique information on behavioral variability.

The task model parameters were equally, if not more predictive, than the task-model-fit FC and significantly outperformed the task-model-residual FC (Figure 3.4) at predicting behavioral differences. For total composite cognition, the nBack task model parameters were more predictive than the nBack task-model-fit FC, explaining 8.4% of its variance, and the MID task model parameters explained 4.3% of the variance in total composite cognition. For SSRT, the SST task model parameters showed the best predictive performance of all fMRI measures, explaining 18.2% of the variance and doubling the prediction effect of SST task-model-fit FC.

Table 3.3 Task model parameters and task-model-fit FC explained both shared and unique variance in individual differences in behaviors. The R^2 columns display the individual variance explained for each fMRI measure, corresponding to the data shown in Figure 3.1 and Figure 3.4. The R^2 adjusted columns display the unique variance explained after adjusting for the effect of the other fMRI measure.

fMRI task	**Task model parameters**		**Task-model-fit FC**	
	R^2	**Unique R^2 adjusted for task-model-fit FC**	**R^2**	**Unique R^2 adjusted for task model parameters**
Behavior: Total composite cognition				
nBack	8.0%	3.1%	6.4%	1.5%
SST	1.5%	1.1%	1.3%	0.8%
MID	4.3%	1.7%	6.4%	3.9%
Behavior: SSRT				
nBack	0.5%	0.3%	0.4%	0.2%
SST	16.7%	11.6%	7.6%	2.5%
MID	0.3%	0.3%	0.4%	0.3%

We next examined whether the task model parameters and task-model-fit FC offered redundant functional brain information relevant for behavior by quantifying the unique behavioral variance explained by the predicted behavioral score of each brain measure after adjusting for the prediction effect of the other (Table 3.3). We observed a decrease in unique

variance explained (unique R^2) for both measures, suggesting that a proportion of the behavioral association was shared between the task model parameters and the task-model-fit FC. Though there was this decrease, both measures were uniquely associated with behavior, still explaining meaningful variance after adjusting for the effect of the other measure. For example, the nBack task model parameters explained 8% of the variance in total composite cognition. After adjusting for the effect of the task-model-fit FC, it uniquely explained 3.1% of behavioral variance. The nBack task-model-fit FC explained 6.4% of the behavioral variance in total composite cognition, and after adjusting for the effect of task model parameters, its unique R^2 dropped to 1.5%.

Adjusting for sociodemographic factors reduced the behavioral prediction performance of FC and task model parameters.

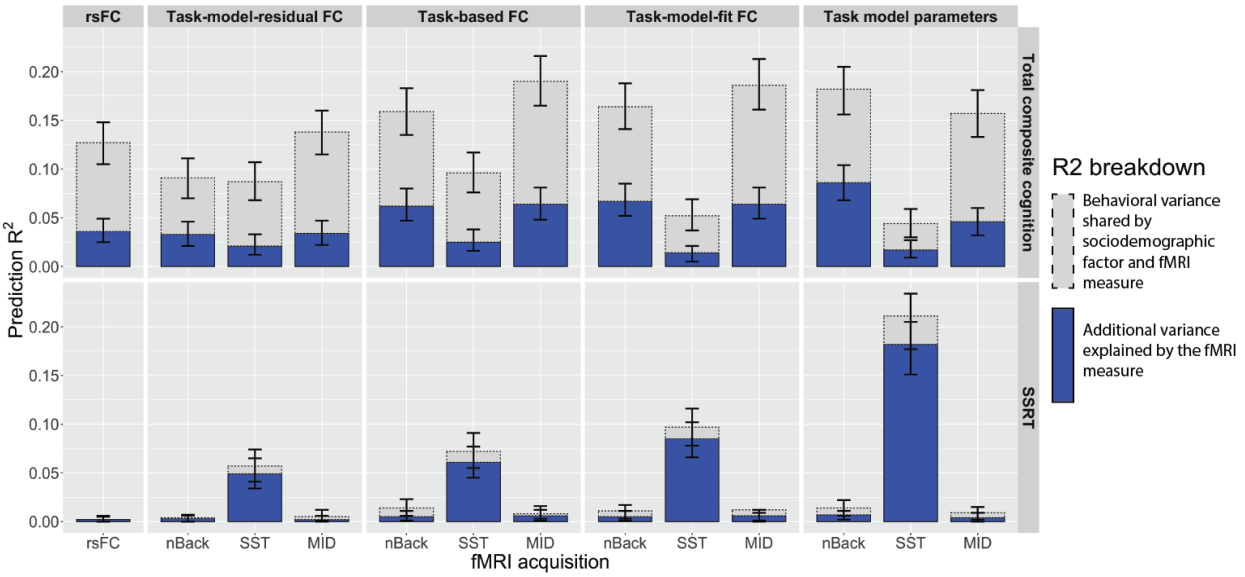


Figure 3.5 A proportion of the behavioral prediction power of task model parameters and FC measures was explained by sociodemographic variation across individuals. The unadjusted prediction R2 for total composite cognition (top row) and the SSRT (bottom row) for each fMRI measure was partitioned into two components, a variance component that was shared with sociodemogaphic factors (shown in gray) and a variance component that was additive to the effect of sociodemographic factors (shown in blue), i.e., the prediction effect after adjusting for sociodemographic covariates.

Sociodemographic variables accounted for a proportion of the unadjusted behavioral association of fMRI measures, and the effect was more prominent for the prediction of total composite cognition (Figure 3.5). When not controlling for the sociodemographic factors, resting-state FC explained 12.7% of the variance in total composite cognition. That number dropped to 3.6% after the adjustment for sociodemographic differences. A similar reduction in prediction performance was also observed for the task-derived FC measures and task model parameters. For the association between the SSRT and the SST task, we observed a more moderate effect of sociodemographic adjustment. The SST task model parameters explained 21.1% of the variability in SSRT without the adjustment of sociodemographic covariates. After covarying for sociodemographic factors, the association was 18.2%.

3.4 Discussion

Characterizing the optimal fMRI measures that capture variance in behavioral differences is a critical step to develop reliable neuroimaging biomarkers for the detection and treatment of brain and behavioral disorders. This study addressed this issue by comparing the behavioral prediction performance of resting-state and task-derived fMRI measures including resting-state FC, task-based FC, task-model-fit FC, task-model-residual FC, and task model parameters. Previous findings have suggested that task fMRI is better than resting-state fMRI at capturing behaviorally relevant FC signals (Rosenberg et al., 2016; Greene et al., 2018; Finn & Bandettini, 2021). We hypothesized that fMRI tasks better reproduce neural processes required to meet the cognitive demands that individuals experience in real life and thus elicit changes in FC patterns that are better associated with individual differences in behavioral phenotypes. We found that, when an fMRI task captured similar cognitive constructs as the behavior of interest, task-model-

fit FC and task model parameters were better than resting-state FC and the task-model-residual FC component at predicting individual differences in that behavior.

Behavioral differences are better predicted by FC patterns derived from task fMRI than resting-state fMRI.

Consistent with previous findings (Rosenberg et al., 2016; Greene et al., 2018; Finn & Bandettini, 2021), we observed an advantage for the task fMRI paradigms over resting-state fMRI at predicting individual differences in the total composite cognition score and SSRT. This finding corroborates the previous result that task manipulation accentuates the functional correlation patterns of the brain that are behaviorally relevant (Cole et al., 2021). While resting-state fMRI has been indispensable for the characterization of large-scale brain networks and provides a convenient paradigm for cross-study data aggregation, task fMRI better captures behaviorally relevant FC signals. This behavioral prediction advantage of task-derived FC measures is also task-specific, such that only fMRI tasks that evoke relevant cognitive demands and content to the behavior of interest confer this advantage (Greene et al., 2018; Finn et al., 2017). In our study, this was demonstrated by the double dissociation of the nBack and the SST task in prediction of total composite cognition and SSRT. The greater association between nBack and the total composite cognition could reflect that children with strong working memory abilities also performed better on language tasks and tasks on fluid intelligence (Rosenberg et al., 2020).

Our analysis further showed that the behavioral prediction advantage of task fMRI paradigms is driven by task-elicited FC, that is, changes in FC patterns in response to cognitive demand. While task-elicited FC fluctuations are modest compared to the individual-specific functional connectome identified at rest (Laumann et al., 2017; Gratton et al., 2018), these task-

induced modulations improve the modeling and detection of behavioral differences because they directly reflect changes in the functional brain patterns when a behavior is being performed.

Task model parameters are equally, if not more predictive, than the task-model-fit FC, and both measures confer complementary information on behavioral differences.

The task model parameters were equally, if not more predictive, than the task-model-fit FC at predicting individual differences of both behavioral measures. The SST task model parameters explained approximately 20% of the individual differences in SSRT, which is a significantly improvement relative to the SST task-model-fit FC, the best predicting FC measure from the same fMRI task. Despite the excitement of using FC measures to behavioral differences in the literature, our results suggest that fMRI task activations are at least as good, if not better, than FC measures at capturing individual differences in behaviors.

We also showed that task model parameters and task-model-fit FC contained shared and unique information for predicting behavioral differences, an observation consistent with previous reports (Larabi et al., 2018; Kowalski et al., 2019). Characterizing the behavioral relevance of both task fMRI measures allowed us to uncover unexpected behavioral association patterns with fMRI tasks. For example, we did not expect to observe an association between the MID task FC and the total composite cognition score given limited theories connecting the two measures.

However, we found that the MID task-model-fit FC was equally predictive of total composite cognition score as the nBack task, a working memory task previously associated with cognitive development (Sripada et al., 2020). This unexpected finding was backed by studies reporting similar cognitive performance prediction accuracy for FC measures derived from a working memory task and a reward processing task that captures similar cognitive constructs as the MID task, in the Human Connectome Project (HCP) (Greene et al., 2018; Jiang et al., 2020).

As both the task model parameters and task-model-fit FC measures can be readily derived from existing task fMRI data, we suggest future studies assess the behavioral relevance of both, as they might yield additive information about the neural correlates of complex behavioral phenotypes.

Sociodemographic factors treatment is crucial and yields differential implications for behavioral prediction studies of fMRI measures

Importantly, we found that adjusting for sociodemographic covariates, including age, sex at birth, ancestry, ethnicity, income, and education, significantly impacted the behavioral prediction effect of FC measures and task model parameters, and such effect was more prominent for total composite cognition than for SSRT. This is consistent with previous findings that sociodemographic factors account for substantial individual variability in fMRI phenotypes (Yaple & Yu, 2020; Rakesh, Zalesky, Whittle, 2021) and in measures of cognitive performance (Bradley & Corwyn, 2002; Korous et al., 2020), and that adjusting for sociodemographic factors reduces the effect sizes of resting-state FC measures on cognitive task performance (Marek et al., 2022).

Controlling for sociodemographic factors can substantially alter estimates of the power of brain phenotypes to predict behavioral differences. An investigator's choice to include these variables as covariates, and which to include, should be guided by the specific prediction goal of the analysis. Because sociodemographic variables are so robustly linked to both neuroimaging and behavioral phenotypes in the ABCD Study, it will probably be necessary to consider the pattern of associations across many models to begin to understand the underlying relationships. Here we have chosen to present both the model with no adjustment and the model with adjustment for all of the sociodemographic variables listed above. For our predictions of the total

composite cognition score in the general population, the results suggest robust association between this measure and functional brain phenotypes. However, the results with the full model (including covariates) suggest that when only differences among peers of the same age, sex, ancestry, ethnicity, and parental income/education are considered in the model, the associations with functional brain phenotypes are much more modest. This trend was also observed in an earlier study of ABCD participants involving structural brain phenotypes (Palmer et al. 2021). While these discrepancies in the results can sometimes lead to confusion for scientists and other stakeholders, it is important to emphasize that the different models both answer different questions about prediction, and raise new questions about the factors that reduce generalizability across groups within the population. To address this uncertainty it may be helpful for researchers to develop standards for presenting several covariate models in each publication to help readers understand better the context of their estimates of prediction from neuroimaging phenotypes (see Wagenmakers et al., 2022).

Limitations

We used a correlation-based FC estimation framework to quantify the behavioral relevance of resting-state and task fMRI data. Graph-theory derived network properties of FC measures have also been associated with behavioral outcomes (Liu et al., 2012; Khazaee, Ebrahimzadeh, Babajani-Feremi, 2015; Qian et al., 2018) and might have provided evidence for additional prediction power. The out-of-sample behavioral prediction in this study could be an underestimation of the behavioral relevance of resting-state fMRI and task fMRI data as other network-based fMRI properties might introduce additional behavioral prediction power relative to the correlation-based FC measures. This limitation, however, would not change our conclusions regarding the relative advantage of task-related FC over resting-state FC for

capturing behaviorally relevant differences, as all FC measures were processed with the same censoring and filtering criteria and were applied to the same prediction pipeline. Similarly, our choice of prediction method may also have impacted the reported out-of-sample prediction performance. Other analytical methods, such as machine learning based prediction methods, could potentially yield different estimates of the behavioral prediction performance of FC measures. Also, we limited the scope of the paper by not delving into the identification of specific brain regions/networks. Future studies should focus on characterizing the differences between the resting-state FC matrix and different task-model-fit FC matrices to understand how task manipulation changes the FC patterns of the brain and how these FC differences relate to behaviors.

3.5 Conclusion

In summary, by comparing the behavioral prediction performance of FC measures derived from task fMRI to that from resting-state FC, we provide additional evidence that fMRI tasks that evoke neural processes relevant to the behavioral phenotypes of interest are better predictors of those phenotypes than FC measures from resting-state fMRI. To maximize the ability to detect behaviorally relevant FC patterns of the brain, effort should be made to select fMRI tasks that recruit similar cognitive processes relevant to the behavioral phenotypes of interest. This work provides further support for the utility of the task activation and FC analysis frameworks for the identification of functionally relevant brain signals. It also highlights the need for consistent reporting of the results of behavioral prediction studies to examine the impact of sociodemographic covariates on the prediction, and to describe more clearly the prediction context to which the models could be expected to generalize, based on these covariates.

3.6 Acknowledgment

Chapter 3, in part, is currently being prepared for submission for publication of the material. Weiqi Zhao, Donald J. Hagler, Hugh P. Garavan, Deanna J. Greene, Terry L. Jernigan & Anders M. Dale. The dissertation author was the primary researcher and author of this material.

3.7 Reference

1. Al-Aidroos, N., Said, C.P., Turk-Browne, N.B., 2012. Top-down attention switches coupling between low-level and high-level areas of human visual cortex. Proc. Natl. Acad. Sci. U. S. A. 109, 14675–14680.

2. Bradley, R.H., Corwyn, R.F., 2002. Socioeconomic status and child development. Annu. Rev. Psychol. 53, 371–399.

3. Brown, M.R.G., Sidhu, G.S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P.H., Greenshaw, A.J., Dursun, S.M., 2012. ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. Front. Syst. Neurosci. 6, 69.

4. Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Jr, Daniela Cornejo, M., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., ABCD Imaging Acquisition Workgroup, 2018. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.

5. Castellanos, F.X., Sonuga-Barke, E.J.S., Milham, M.P., Tannock, R., 2006. Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cogn. Sci. 10, 117–123.

6. Chaarani, B., Hahn, S., Allgaier, N., Adise, S., Owens, M.M., Juliano, A.C., Yuan, D.K., Loso, H., Ivanciu, A., Albaugh, M.D., Dumas, J., Mackey, S., Laurent, J., Ivanova, M., Hagler, D.J., Cornejo, M.D., Hatton, S., Agrawal, A., Aguinaldo, L., Ahonen, L., Aklin, W., Anokhin, A.P., Arroyo, J., Avenevoli, S., Babcock, D., Bagot, K., Baker, F.C., Banich, M.T., Barch, D.M., Bartsch, H., Baskin-Sommers, A., Bjork, J.M., Blachman-Demner, D., Bloch, M., Bogdan, R., Bookheimer, S.Y., Breslin, F., Brown, S., Calabro, F.J., Calhoun, V., Casey, B.J., Chang, L., Clark, D.B., Cloak, C., Constable, R.T., Constable, K., Corley, R., Cottler, L.B., Coxe, S., Dagher, R.K., Dale, A.M., Dapretto, M., Delcarmen-Wiggins, R., Dick, A.S., Do, E.K., Dosenbach, N.U.F., Dowling, G.J., Edwards, S., Ernst, T.M., Fair, D.A., Fan, C.C., Feczko, E., Feldstein-Ewing, S.W., Florsheim, P., Foxe, J.J., Freedman, E.G., Friedman, N.P., Friedman-Hill, S., Fuemmeler, B.F., Galvan, A., Gee, D.G., Giedd, J., Glantz, M., Glaser, P., Godino, J., Gonzalez, M., Gonzalez, R., Grant, S., Gray, K.M., Haist, F., Harms, M.P., Hawes, S., Heath, A.C., Heeringa, S., Heitzeg, M.M., Hermosillo, R., Herting, M.M., Hettema, J.M., Hewitt, J.K., Heyser, C., Hoffman, E., Howlett, K., Huber, R.S., Huestis, M.A., Hyde, L.W., Iacono, W.G., Infante, M.A., Irfanoglu, O., Isaiah, A., Iyengar, S., Jacobus, J., James, R., Jean-Francois, B., Jernigan, T., Karcher, N.R., Kaufman, A., Kelley, B., Kit, B., Ksinan, A., Kuperman, J., Laird, A.R., Larson, C., LeBlanc, K., Lessov-Schlagger, C., Lever, N., Lewis, D.A., Lisdahl, K., Little, A.R., Lopez, M., Luciana, M., Luna, B., Madden, P.A., Maes, H.H., Makowski, C., Marshall, A.T., Mason, M.J., Matochik, J., McCandliss, B.D., McGlade, E., Montoya, I., Morgan, G., Morris, A., Mulford, C., Murray, P., Nagel, B.J., Neale, M.C., Neigh, G., Nencka, A., Noronha, A., Nixon, S.J.,

Palmer, C.E., Pariyadath, V., Paulus, M.P., Pelham, W.E., Pfefferbaum, D., Pierpaoli, C., Prescot, A., Prouty, D., Puttler, L.I., Rajapaske, N., Rapuano, K.M., Reeves, G., Renshaw, P.F., Riedel, M.C., Rojas, P., de la Rosa, M., Rosenberg, M.D., Ross, M.J., Sanchez, M., Schirda, C., Schloesser, D., Schulenberg, J., Sher, K.J., Sheth, C., Shilling, P.D., Simmons, W.K., Sowell, E.R., Speer, N., Spittel, M., Squeglia, L.M., Sripada, C., Steinberg, J., Striley, C., Sutherland, M.T., Tanabe, J., Tapert, S.F., Thompson, W., Tomko, R.L., Uban, K.A., Vrieze, S., Wade, N.E., Watts, R., Weiss, S., Wiens, B.A., Williams, O.D., Wilbur, A., Wing, D., Wolff-Hughes, D., Yang, R., Yurgelun-Todd, D.A., Zucker, R.A., Potter, A., Garavan, H.P., ABCD Consortium, 2021. Baseline brain function in the preadolescents of the ABCD Study. Nat. Neurosci. 24, 1176–1186.

7. Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M., 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. Neuroimage 112, 232–243.

8. Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L.Q.R., Asplund, C.L., Marek, S., Dosenbach, N.U.F., Eickhoff, S.B., Bzdok, D., Holmes, A.J., Yeo, B.T.T., 2022. Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. Nat. Commun. 13, 2217.

9. Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron 83, 238–251.

10. Cole, M.W., Ito, T., Cocuzza, C., Sanchez-Romero, R., 2021. The Functional Relevance of Task-State Functional Connectivity. J. Neurosci. 41, 2684–2702.

11. Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Jr., Barch, D.M., Petersen, S.E., Schlaggar, B.L., 2010. Prediction of Individual Brain Maturity Using fMRI. Science 329.

12. Elliott, M.L., Knodt, A.R., Cooke, M., Kim, M.J., Melzer, T.R., Keenan, R., Ireland, D., Ramrakha, S., Poulton, R., Caspi, A., Moffitt, T.E., Hariri, A.R., 2019. General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. Neuroimage 189, 516–532.

13. Fair, D.A., Miranda-Dominguez, O., Snyder, A.Z., Perrone, A., Earl, E.A., Van, A.N., Koller, J.M., Feczko, E., Klein, R.L., Mirro, A.E., Hampton, J.M., Adeyemo, B., Laumann, T.O., Gratton, C., Greene, D.J., Schlaggar, B.L., Hagler, D., Watts, R., Garavan, H., Barch, D.M., Nigg, J.T., Petersen, S.E., Dale, A., Feldstein-Ewing, S.W., Nagel, B.J., Dosenbach, N.U.F., 2018. Correction of respiratory artifacts in MRI head motion estimates. bioRxiv.

14. Fair, D.A., Schlaggar, B.L., Cohen, A.L., Miezin, F.M., Dosenbach, N.U.F., Wenger, K.K., Fox, M.D., Snyder, A.Z., Raichle, M.E., Petersen, S.E., 2007. A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. Neuroimage 35, 396–405.

15. Fan, C.C., Palmer, C.E., Iverson, J., Pecheva, D., Thompson, W.K., Hagler, D., Jernigan, T.L., Dale, A.M., 2021. FEMA: Fast and efficient mixed-effects algorithm for population-scale whole brain imaging data. bioRxiv.

16. Farah, M.J., 2018. Socioeconomic status and the brain: prospects for neuroscience-informed policy. Nat. Rev. Neurosci. 19, 428–438.

17. Finn, E.S., Bandettini, P.A., 2021. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. Neuroimage 235, 117963.

18. Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? Neuroimage 160, 140–151.

19. Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18, 1664–1671.

20. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole Brain Segmentation: Neurotechnique Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 33, 341–355.

21. Gabrieli, J.D.E., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. Neuron 85, 11–26.

22. Gao, S., Greene, A.S., Constable, R.T., Scheinost, D., 2019. Combining multiple connectomes improves predictive modeling of phenotypic measures. Neuroimage 201, 116038.

23. Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R.Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., Zahs, D., 2018. Recruiting the ABCD sample: Design considerations and procedures. Dev. Cogn. Neurosci. 32, 16–22.

24. Gauggel, S., Rieger, M., Feghoff, T.-A., 2004. Inhibition of ongoing responses in patients with Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 75, 539–544.

25. Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2016. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cereb. Cortex 26, 288–303.

26. Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., Dosenbach, N.U.F., Petersen, S.E., 2018. Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. Neuron 98, 439–452.e5.

27. Greene, A.S., Gao, S., Noble, S., Scheinost, D., Constable, R.T., 2020. How Tasks Change Whole-Brain Functional Organization to Reveal Brain-Phenotype Relationships. Cell Rep. 32, 108066.

28. Greene, A.S., Gao, S., Scheinost, D., Constable, R.T., 2018. Task-induced brain state manipulation improves prediction of individual traits. Nat. Commun. 9, 2807.

29. Hagler, D.J., Hatton, S.N., Makowski, C., Daniela Cornejo, M., Fair, D.A., Dick, A.S., Sutherland, M.T., Casey, B.J., Barch, D.M., Harms, M.P., Watts, R., Bjork, J.M., Garavan, H.P., Hilmer, L., Pung, C.J., Sicat, C.S., Kuperman, J., Bartsch, H., Xue, F., Heitzeg, M.M., Laird, A.R., Trinh, T.T., Gonzalez, R., Tapert, S.F., Riedel, M.C., Squeglia, L.M., Hyde, L.W., Rosenberg, M.D., Earl, E.A., Howlett, K.D., Baker, F.C., Soules, M., Diaz, J., de Leon, O.R., Thompson, W.K., Neale, M.C., Herting, M., Sowell, E.R., Alvarez, R.P., Hawes, S.W., Sanchez, M., Bodurka, J., Breslin, F.J., Morris, A.S., Paulus, M.P., Kyle Simmons, W., Polimeni, J.R., van der Kouwe, A., Nencka, A.S., Gray, K.M., Pierpaoli, C., Matochik, J.A., Noronha, A., Aklin, W.M., Conway, K., Glantz, M., Hoffman, E., Little, R., Lopez, M., Pariyadath, V., Weiss, S.R.B., Wolff-Hughes, D.L., DelCarmen-Wiggins, R., Feldstein Ewing, S.W., Miranda-Dominguez, O., Nagel, B.J., Perrone, A.J., Sturgeon, D.T., Goldstone, A., Pfefferbaum, A., Pohl, K.M., Prouty, D., Uban, K., Bookheimer, S.Y., Dapretto, M., Galvan, A., Bagot, K., Giedd, J., Alejandra Infante, M., Jacobus, J., Patrick, K., Shilling, P.D., Desikan, R., Li, Y., Sugrue, L., Banich, M.T., Friedman, N., Hewitt, J.K., Hopfer, C., Sakai, J., Tanabe, J., Cottler, L.B., Nixon, S.J., Chang, L., Cloak, C., Ernst, T., Reeves, G., Kennedy, D.N., Heeringa, S., Peltier, S., Schulenberg, J., Sripada, C., Zucker, R.A., Iacono, W.G., Luciana, M., Calabro, F.J., Clark, D.B., Lewis, D.A., Luna, B., Schirda, C., Brima, T., Foxe, J.J., Freedman, E.G., Mruzek, D.W., Mason, M.J., Huber, R., McGlade, E., Prescot, A., Renshaw, P.F., Yurgelun-Todd, D.A., Allgaier, N.A., Dumas, J.A., Ivanova, M., Potter, A., Florsheim, P., Larson, C., Lisdahl, K., Charness, M.E., Fuemmeler, B., Hettema, J.M., Steinberg, J., Anokhin, A.P., Glaser, P., Heath, A.C., Madden, P.A., Baskin-Sommers, A., Todd Constable, R., Grant, S.J., Dowling, G.J., Brown, S.A., Jernigan, T.L., Dale, A.M., 2018. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. bioRxiv.

30. Hallquist, M.N., Hwang, K., Luna, B., 2013. The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. Neuroimage 82, 208–225.

31. Hyatt, C.S., Owens, M.M., Crowe, M.L., Carter, N.T., Lynam, D.R., Miller, J.D., 2020. The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. Neuroimage 205, 116225.

32. Jurkiewicz, M.T., Crawley, A.P., Mikulis, D.J., 2018. Is Rest Really Rest? Resting-State Functional Connectivity During Rest and Motor Task Paradigms. Brain Connect. 8, 268–275.

33. Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A., 2016. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. Brain Imaging Behav. 10, 799–817.

34. Kim, J., Calhoun, V.D., Shim, E., Lee, J.-H., 2016. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. Neuroimage 124, 127–146.

35. Korous, K.M., Causadias, J.M., Bradley, R.H., Luthar, S.S., Levy, R., 2020. A Systematic Overview of Meta-Analyses on Socioeconomic Status, Cognitive Ability, and Achievement: The Need to Focus on Specific Pathways. Psychol. Rep. 33294120984127.

36. Kowalski, J., Wypych, M., Marchewka, A., Dragan, M., 2019. Neural Correlates of Cognitive-Attentional Syndrome: An fMRI Study on Repetitive Negative Thinking Induction and Resting State Functional Connectivity. Front. Psychol. 10, 648.

37. Larabi, D.I., van der Meer, L., Pijnenborg, G.H.M., Ćurčić-Blake, B., Aleman, A., 2018. Insight and emotion regulation in schizophrenia: A brain activation and functional connectivity study. Neuroimage Clin 20, 762–771.

38. Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.-Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U.F., Schlaggar, B.L., Mumford, J.A., Poldrack, R.A., Petersen, S.E., 2015. Functional System and Areal Organization of a Highly Sampled Individual Human Brain. Neuron 87, 657–670.

39. Liu, Z., Zhang, Y., Yan, H., Bai, L., Dai, R., Wei, W., Zhong, C., Xue, T., Wang, H., Feng, Y., You, Y., Zhang, X., Tian, J., 2012. Altered topological patterns of brain networks in mild cognitive impairment and Alzheimer's disease: a resting-state fMRI study. Psychiatry Res. 202, 118–125.

40. Luciana, M., Bjork, J.M., Nagel, B.J., Barch, D.M., Gonzalez, R., Nixon, S.J., Banich, M.T., 2018. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. Dev. Cogn. Neurosci. 32, 67–79.

41. Lynam, D.R., Hoyle, R.H., Newman, J.P., 2006. The perils of partialling: cautionary tales from aggression and psychopathy. Assessment 13, 328–341.

42. Moutoussis, M., Garzón, B., Neufeld, S., Bach, D.R., Rigoli, F., Goodyer, I., Bullmore, E., NSPN Consortium, Guitart-Masip, M., Dolan, R.J., 2021. Decision-making ability, psychopathology, and brain connectivity. Neuron 109, 2025–2040.e7.

43. Nielsen, A.N., Greene, D.J., Gratton, C., Dosenbach, N.U.F., Petersen, S.E., Schlaggar, B.L., 2019. Evaluating the Prediction of Brain Maturity From Functional Connectivity After Motion Artifact Denoising. Cereb. Cortex 29, 2455–2469.

44. Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84, 320–341.

45. Qian, X., Loo, B.R.Y., Castellanos, F.X., Liu, S., Koh, H.L., Poh, X.W.W., Krishnan, R., Fung, D., Chee, M.W., Guan, C., Lee, T.-S., Lim, C.G., Zhou, J., 2018. Brain-computer-interface-based intervention re-normalizes brain functional network topology in children with attention deficit/hyperactivity disorder. Transl. Psychiatry 8, 149.

46. Rakesh, D., Zalesky, A., Whittle, S., 2021. Similar but distinct - Effects of different socioeconomic indicators on resting state functional connectivity: Findings from the Adolescent Brain Cognitive Development (ABCD) Study®. Dev. Cogn. Neurosci. 51, 101005.

47. Rosenberg, M.D., Finn, E.S., Scheinost, D., Papademetris, X., Shen, X., Constable, R.T., Chun, M.M., 2016. A neuromarker of sustained attention from whole-brain functional connectivity. Nat. Neurosci. 19, 165–171.

48. Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632.

49. Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22, 1359–1366.

50. Speer, S.P.H., Smidts, A., Boksem, M.A.S., 2021. Individual differences in (dis)honesty are represented in the brain's functional connectivity at rest. Neuroimage 246, 118761.

51. Sripada, C., Angstadt, M., Taxali, A., Clark, D.A., Greathouse, T., Rutherford, S., Dickens, J.R., Shedden, K., Gard, A.M., Hyde, L.W., Weigard, A., Heitzeg, M., 2021. Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth. Transl. Psychiatry 11, 571.

52. Sripada, C., Rutherford, S., Angstadt, M., Thompson, W.K., Luciana, M., Weigard, A., Hyde, L.H., Heitzeg, M., 2019. Prediction of neurocognition in youth from resting state fMRI. Mol. Psychiatry.

53. Taylor, R.L., Cooper, S.R., Jackson, J.J., Barch, D.M., 2020. Assessment of Neighborhood Poverty, Cognitive Function, and Prefrontal and Hippocampal Volumes in Children. JAMA Netw Open 3, e2023774.

54. Thomas, R.M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., van Wingen, G., 2020. Classifying Autism Spectrum Disorder Using the Temporal Statistics of Resting-State Functional MRI Data With 3D Convolutional Neural Networks. Front. Psychiatry 11, 440.

55. Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157, 521–530.

56. Varoquaux, G., Poldrack, R.A., 2019. Predictive models avoid excessive reductionism in cognitive neuroimaging. Curr. Opin. Neurobiol. 55, 1–6.

57. Wagenmakers, E.-J., Sarafoglou, A., Aczel, B., 2022. One statistical analysis must not rule them all [WWW Document]. Nature Publishing Group UK. URL http://dx.doi.org/10.1038/d41586-022-01332-8 (accessed 8.10.22).

58. Yaple, Z.A., Yu, R., 2020. Functional and Structural Brain Correlates of Socioeconomic Status. Cereb. Cortex 30, 181–196.

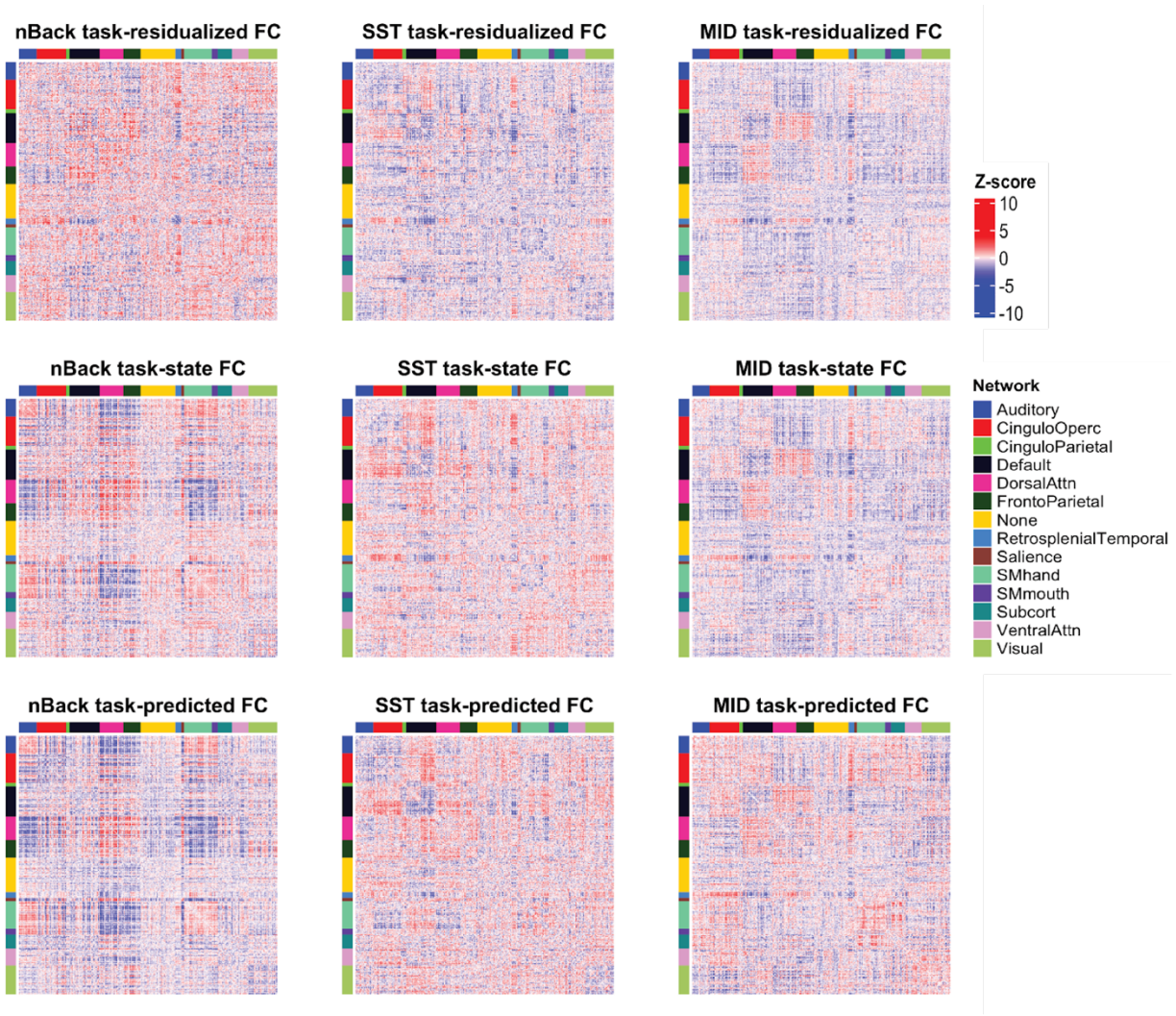
59. Zhang, M., Nathaniel, U., Savill, N., Smallwood, J., Jefferies, E., 2021. Intrinsic connectivity of left ventrolateral prefrontal cortex predicts individual differences in controlled semantic retrieval. Neuroimage 246, 118760.

60. Zhao, W., Palmer, C.E., Thompson, W.K., Chaarani, B., Garavan, H.P., Casey, B.J., Jernigan, T.L., Dale, A.M., Fan, C.C., 2021. Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. Cereb. Cortex 31, 1478–1488.

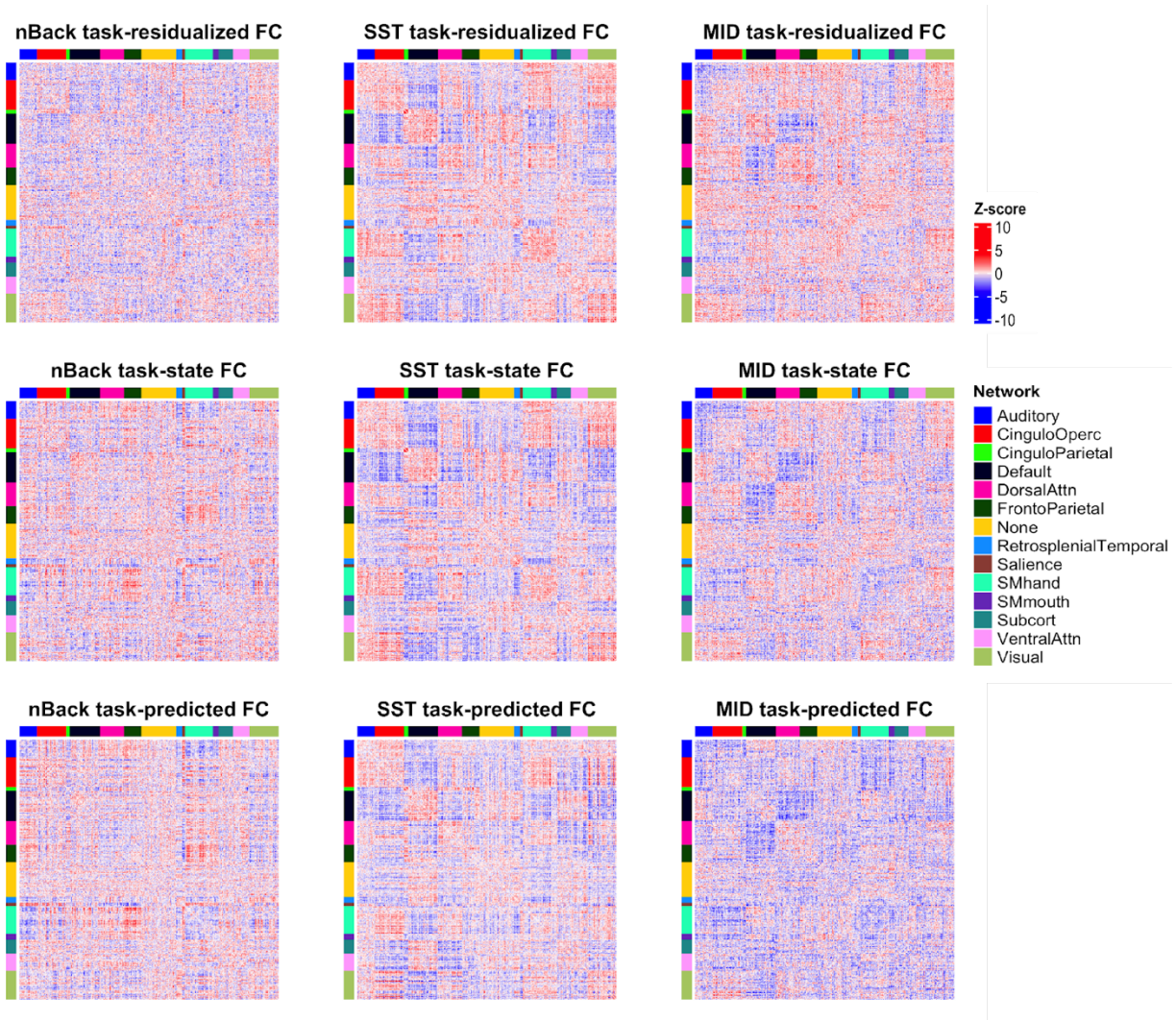
3.8 Supplementary Figures & Tables

Supplementary Table 3.1 Additional inclusion criteria and final sample size for each fMRI modality.

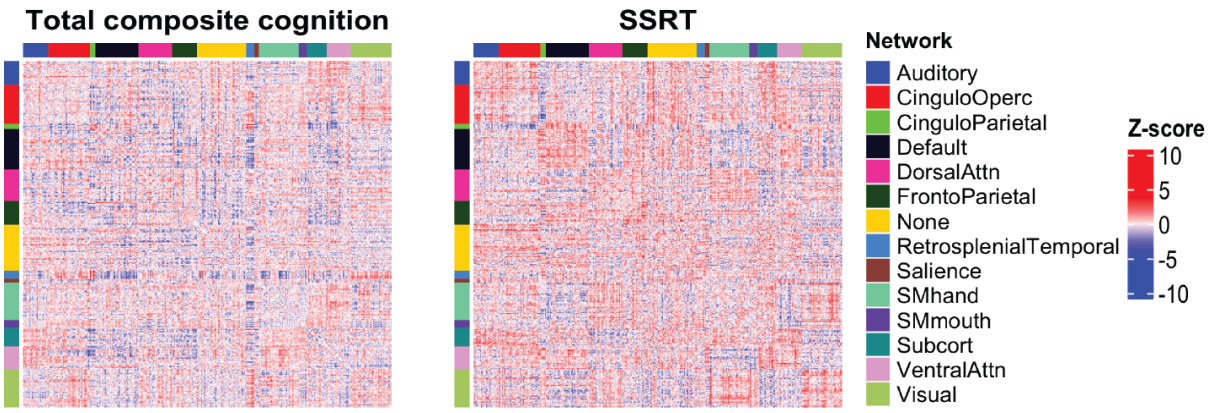
Inclusion criteria	**nBack Task**	**SST Task**	**MID Task**	**Resting-state fMRI**
Number of subjects with usable data for the task-state, task-predicted, and task-residualized FC	4880	5250	4685	4291
Number of subjects with well-defined fMRI task design matrix	4051	4332	4399	4291
Number of subjects who passed ABCD recommended inclusion criteria	3568	3931	4266	3999
Number of subjects with complete data for behavioral and sociodemographic variables	3034	3414	3575	3463
Number of subjects used in behavioral prediction analysis	3034	3034 (randomly sampled from N = 3414)	3034 (randomly sampled from N = 3575)	3034 (randomly sampled from N = 3463)



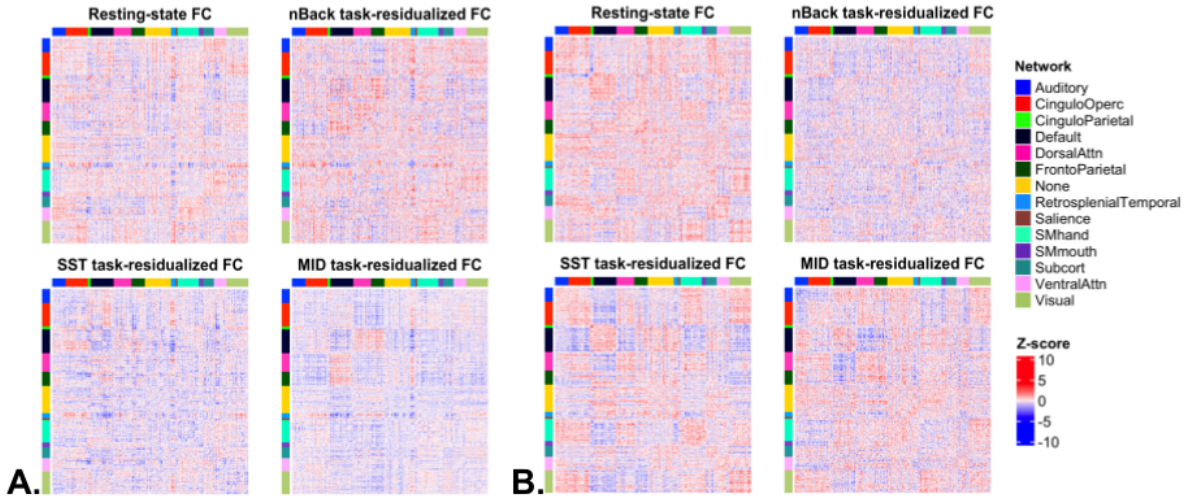
Supplementary Figure 3.1 The mass univariate effect size maps the associations between total composite cognition and each task-related FC measures.



Supplementary Figure 3.2 The mass univariate effect size maps the associations between SSRT and each task-related FC measure.



Supplementary Figure 3.3 The mass univariate effect size maps the associations between resting-state FC and each behavioral measure.



Supplementary Figure 3.4 A comparison between the effect size maps of resting-state FC and task-residualized FC measures. We presented the mass univariate z-scores of resting-state FC and task-residualized FC measures on total composite cognition (Panel A) and SSRT (Panel B).

Chapter 4 Conclusion

This thesis systematically examined how to maximize the behavioral prediction utility of task-based activations and FC measures derived from resting-state and task fMRI paradigms. Chapter 2 highlighted the importance of using multivariate methods to aggregate the distributed effect sizes of task fMRI activation maps on behaviors. Chapter 3 demonstrated the advantage of using task fMRI paradigms to elicit behaviorally relevance task activation and FC patterns. Together, both studies implied the importance of fMRI task design at eliciting behaviorally meaningful fMRI patterns during fMRI acquisitions. By demonstrating how task fMRI data can be better measured and modeled to understand behaviors, this thesis provided some evidence on how to take advantage of fMRI acquisitions to derive robust and reliable neuroimaging signatures for cognitive and mental health outcomes.