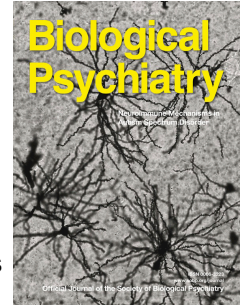# Journal Pre-proof

Brain-based predictions of psychiatric illness-linked behaviors across the sexes

Elvisha Dhamala, Leon Qi Rong Ooi, Jianzhong Chen, Jocelyn A. Ricard, Emily Berkeley, Sidhant Chopra, Yueyue Qu, Xihan Zhang, Connor Lawhead, B.T. Thomas Yeo, Avram J. Holmes

Please cite this article as: Dhamala E., Rong Ooi L.Q., Chen J., Ricard J.A., Berkeley E., Chopra S., Qu Y., Zhang X., Lawhead C., Yeo B.T.T. & Holmes A.J., Brain-based predictions of psychiatric illness-linked behaviors across the sexes, *Biological Psychiatry* (2023), doi: https://doi.org/10.1016/j.biopsych.2023.03.025.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Brain-based predictions of psychiatric illness-linked behaviors across the sexes

Running Title: Brain networks predict distinct behaviors across sexes

Authors:

Elvisha Dhamala[1-3,*], Leon Qi Rong Ooi[4-7], Jianzhong Chen[5-7], Jocelyn A. Ricard[2], Emily Berkeley[8], Sidhant Chopra[2], Yueyue Qu[2], Xihan Zhang[2], Connor Lawhead[2], B.T. Thomas Yeo[4-7,9], Avram J. Holmes[2-3,10-12,*]

Affiliations:

[1] Institute of Behavioral Sciences, Feinstein Institutes for Medical Research, Manhasset, USA

[2] Department of Psychology, Yale University, New Haven, USA

[3] Kavli Institute for Neuroscience, Yale University, New Haven, USA

[4] Centre for Sleep & Cognition & Centre for Translational Magnetic Resonance Research, Yong Loo Lin School of Medicine, Singapore, National University of Singapore, Singapore, Singapore

[5] Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

[6] N.1 Institute for Health & Institute for Digital Medicine, National University of Singapore, Singapore, Singapore

[7] Integrative Sciences and Engineering Programme (ISEP), National University of Singapore, Singapore, Singapore

[8] Connecticut College, New London, USA

[9] Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, USA

[10] Department of Psychiatry, Yale University, New Haven, USA

[11] Wu Tsai Institute, Yale University, New Haven, USA

[12] Department of Psychiatry, Brain Health Institute, Rutgers University, Piscataway, USA

[*] Corresponding authors:

Elvisha Dhamala (elvisha@gmail.com) and Avram J. Holmes (avram.holmes@yale.edu)

Abstract

Background: Individual differences in functional brain connectivity can be used to predict both the presence of psychiatric illness and variability in associated behaviors. However, despite evidence for sex differences in functional network connectivity and in the prevalence, presentation, and trajectory of psychiatric illnesses, the extent to which disorder-relevant aspects of network connectivity are shared or unique across the sexes remains to be determined.

Methods: In this work, we used predictive modeling approaches to evaluate whether shared or unique functional connectivity correlates underlie the expression of psychiatric illness-linked behaviors in males and females in data from the Adolescent Brain Cognitive Development study (n=5260; 2571 females).

Results: We demonstrate that functional connectivity profiles predict individual differences in externalizing behaviors in males and females, but only predict internalizing behaviors in females. Furthermore, models trained to predict externalizing behaviors in males generalize to predict internalizing behaviors in females, and models trained to predict internalizing behaviors in females generalize to predict externalizing behaviors in males. Finally, the neurobiological correlates of many behaviors are largely shared within and across sexes: functional connections within and between heteromodal association networks including default, limbic, control, and dorsal attention networks are associated with internalizing and externalizing behaviors.

Conclusions: Taken together, these findings suggest that shared neurobiological patterns may manifest as distinct behaviors across the sexes. Based on these results, we recommend that both clinicians and researchers carefully consider how sex may influence the presentation of psychiatric illnesses, especially those along the internalizing-externalizing spectrum.

Introduction

A primary aim of research in psychiatry is to establish the neurobiological correlates of illness-relevant behaviors, facilitating illness prediction, diagnosis, and treatment. Critical to this goal is the consideration of associated individual differences, for instance underlying sex differences. Females are more likely to be diagnosed with affective and anxiety disorders, while males are more likely to meet diagnostic criteria for antisocial and substance use disorders[1-3]. Relatedly, across cultures, females are more likely to express internalizing behaviors directed at one-self (e.g., loneliness) while males are more likely to exhibit externalizing behaviors directed at others or the environment (e.g., aggression)[3, 4]. These differences emerge across childhood, become more evident during adolescence, and persist throughout the lifespan[2]. While sex differences in the prevalence and expression of psychiatric illnesses have been extensively studied at the population-level[5], the underlying neurobiological correlates are not yet fully understood. Genetics, hormones, immunology, neurobiology, environment, and a host of psychosocial factors all likely contribute to expressed behaviors and these contributions may vary across disorders and throughout the lifespan[2]. One possibility is that these factors uniquely contribute to distinct biological underpinnings and associated behavioral expression patterns across the sexes. An alternative, but not mutually exclusive possibility, is that shared biological features may link to dissociable behaviors across the sexes. A thorough understanding of the sex differences that exist in the neurobiological correlates of psychiatric illness-relevant behaviors will facilitate the development and implementation of sex-specific and personalized preventative interventions, diagnostic procedures, and therapeutic modalities.

Functional magnetic resonance imaging is a non-invasive neuroimaging technique to estimate regional neural activation, as inferred though the detection of changes in blood oxygenation levels. Correlations between these signals can be used to quantify the functional coupling (or connectivity) between pairs of brain regions. While sex differences in the functional activation and organization of the brain have been extensively studied, there are few reliable sex/gender differences and many reproducible findings are largely driven by differences in brain size[6]. Functional connectivity profiles exhibit various sex differences throughout the lifespan[7-12]. Females have greater within-network connectivity while males have greater between-network connectivity[9]. These differences are modulated by brain size[13], genetics[14] and hormonal fluctuations[15-18], but must also reflect other biological, social, and environmental influences. Prior analyses have found that functional connections within and between heteromodal networks, and particularly the default and frontoparietal control networks, are largely driving

these differences(10, 12). Intriguingly, functional disruptions within and between the default and control networks, along with the salience network, are implicated in a range of psychiatric phenotypes(19). While sex differences in functional connectivity have been studied extensively, it is not yet known whether there are sex differences in the associations between functional connectivity and psychiatric illness-linked behaviors.

Over the last decade, data-driven predictive modeling approaches have become increasingly used to study brain-behavior relationships(20). These approaches can generate individual-level clinically informative predictions of diagnosis, symptom profile, and treatment response and identify the underlying neurobiology associated with distinct behaviors(20). Through these approaches, functional connectivity can predict individual differences in cognition, personality, and mental health(21-26). These models have been used to establish the neurobiological correlates of attention(27, 28), memory(29), anxiety(30), depression(31), psychosis(32), and substance abuse(33, 34). When developing predictive models, it is crucial to ensure that they are not only accurate within circumscribed groups but that they can also generalize across populations. Prior work indicates that predictions of cognitive and personality traits can fail to generalize across sexes(22, 35-37). To circumvent these issues–and given the known sex differences in psychiatric illnesses–the use of sex-specific prediction models may yield more accurate and generalizable predictions and provide insight into sex differences in the neurobiological correlates of psychiatric illnesses. Moreover, the examination of these brain-behavior relationships in children can reveal whether sex differences emerge prior to adolescence when many of the differences in psychiatric illness risk and presentation become more evident.

Here, we sought to identify whether shared or unique neurobiological correlates underlie the expression of distinct psychiatric illness-linked behaviors across the sexes during childhood. To do so, we quantified the functional connectivity correlates of 17 distinct psychiatric illness-relevant behaviors in children from the Adolescent Brain Cognitive Development (ABCD) dataset. By examining differences in predictive accuracy across sexes and behaviors, we demonstrate that externalizing behaviors can be accurately predicted in males and females, but internalizing behaviors can only be successfully predicted in females. Next, we determine that predictive models generalize within domains both within and between sexes, but only generalize across domains between sexes. Specifically, models trained to predict externalizing behaviors in either sex generalize to predict related behaviors in both sexes. However, models trained to

4

predict externalizing behaviors in males also generalize to predict internalizing behaviors in females, and models trained to predict internalizing behaviors in females generalize to predict externalizing behaviors in males. Investigating the network correlates of these behaviors, we reveal that functional connectivity within and between shared heteromodal association networks are associated with internalizing and externalizing behaviors, and these correlates are shared across the sexes. Collectively, these results suggest that shared aspects of neurobiology may underlie distinct behaviors across the sexes. As such, these findings may change the way clinicians consider sex when diagnosing psychiatric disorders across the internalizing-externalizing spectrum.

Methods

Dataset

We included 5260 children (2571 females, ages 9-10; Figure 1A) from the Adolescent Brain Cognitive Development (ABCD) 2.0.1 release(38). After quality control of imaging data(24, 39), we filtered participants based on availability of functional MRI scans and behavioral scores. Detailed information about our sample construction is in the supplemental methods along with demographic information (age, race/ethnicity, and income) for subjects included and excluded from these analyses (Table S1).

Behavioral Data

We included eight syndrome scales and three summary scores from the Child Behavior Checklist(40), and six DSM-5 oriented scales for a total of 17 behavioral scores (Figure 1B). Details about the scales and behavioral scores are in the supplemental methods. We used non-parametric Mann-Whitney U rank tests to evaluate sex differences in the behavioral scores. All p-values were corrected for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (q=0.05) procedure(41). We also computed non-parametric correlations between the behavioral scores for each sex to evaluate any underlying relationships that may exist between the behavioral scores and influence subsequent analyses. Finally, we computed full correlations between summary motion data (mean framewise displacement) and behavioral scores to determine whether residual motion artifacts could be driving behavioral prediction performance.

Image Acquisition and Processing

The minimally preprocessed resting-state functional MRI data were processed as previously described(24, 42). Key processing steps are summarized in the supplemental methods. Once processed, we extracted regional time series for 400 cortical(43) and 19 subcortical(44) parcels (Figure 1C). Full correlations were computed between those time series yielding a 419x419 pairwise regional functional connectivity matrix (Figure 1D).

Predictive Modeling

Linear regression models and deep learning algorithms achieve comparable accuracies for brain-based behavioral predictions(25), but linear models avoid overfitting, are more interpretable, and are less computationally expensive(20). The predictive models used here rely on a similar framework as those previously described(21, 22, 45) to perform novel analyses

addressing cross-behavioral model generalization within and across the sexes in the context of psychiatric illness-linked behaviors. We used sex-specific cross-validated linear ridge regression models to predict each behavioral score based on functional connectivity data while accounting for site (Figure 1E). Model performance was evaluated using prediction accuracy within each sex and behavior and model generalizability across sexes and behaviors. We also evaluated whether models performed better than chance using null distributions(46). To ensure model generalizability was not driven by correlations in the underlying behavioral data, we computed the full correlation between the (upper triangular) model generalizability matrix and behavioral correlation matrix for each sex. We also extracted feature weights from the models and transformed them using the Haufe transformation(47) to increase their interpretability and reliability(24, 42, 48). We then summarized pairwise regional feature weights at a network-level to support interpretability as previously described(22). Details about the modeling approach, significance testing, Haufe transformation, and summarization of feature weights to a network-level are in the supplemental methods.

Data and Code Availability

All ABCD data can be accessed via the NIMH Data Archive. All code used to generate the results are available on GitHub [https://github.com/elvisha/ABCD_sexspecific_clinicalpredictions].

7

Results

<u>Males and females exhibit overlapping behaviors</u>

Males and females exhibit largely overlapping distributions of behavioral scores (Figure S1A, Table S2), but there are significant (corrected ps<0.01) sex differences in the withdrawn/depressed, somatic complaints, externalizing, rule-breaking, aggressive, thought problems, attention problems, and total problems syndrome scales, as well as affective, somatic, oppositional defiant, conduct, and ADHD DSM-5 oriented scales. Across the scores with sex differences, males report greater scores in all behaviors apart from somatic complaints and somatic problems.

Within each sex, behavioral scores are strongly correlated within behavioral domains (Figure S1B). Correlations between internalizing scores range from 0.27 to 0.91 in males, and 0.28 to 0.92 in females. Correlations between externalizing scores range from 0.53 to 0.96 in males, and 0.51 to 0.96 in females. Meanwhile, correlations across behavioral domains are numerically weaker. Correlations between internalizing and externalizing scores range from 0.23 to 0.55 in males, and 0.25 to 0.52 in females.

Behavioral scores are not correlated with motion (Table S3).

Here, we replicate prior findings demonstrating sex differences in the prevalence of behaviors associated with an increased risk for illness onset and provide evidence suggesting that these differences may emerge prior to adolescence. These findings also suggest that predictive models may be more likely to generalize within sexes rather than across sexes. Additionally, we observe similar relationships between psychiatric illness-linked behaviors in males and females. These observed relationships suggest that models may be more likely to generalize within behavioral domains rather than across behavioral domains.

Brain-based predictive models predict psychiatric illness-linked behaviors

Linear ridge regression models were trained to predict psychiatric illness-linked behaviors in males and females based on functional connectivity. Model accuracies, defined as the correlation between true and predicted scores, were evaluated in comparison to null models (Figure 2A).

In males, models successfully predict behaviors (corrected p<0.05) within the externalizing domain (externalizing (r=0.12), rule-breaking (r=0.14), and aggressive (r=0.10) behaviors), along with attention (r=0.13), social (r=0.12), and total (r=0.11) problems, as well as behaviors related to oppositional defiant (r=0.08), conduct (r=0.13), and attention deficit/hyperactivity (ADHD; r=0.12) disorders.

In females, models successfully predict behaviors (corrected p<0.05) within the internalizing (withdrawn/depressed (r=0.09)) and externalizing domains (externalizing (r=0.11), rule-breaking (r=0.15), and aggressive (r=0.09) behaviors), along with attention (r=0.11) and social (r=0.07) problems, as well as behaviors related to conduct disorders (r=0.14) and ADHD (r=0.08).

In prior work, we observed that internalizing behaviors are more difficult to predict than externalizing behaviors(24). Here, we replicate those findings and further suggest that the predictability of specific behaviors may differ across the sexes.

Brain-based predictive models generalize across sexes and behaviors

Generalizability, defined as the prediction accuracy obtained when a model is evaluated on a population and/or behavior distinct from the population and/or behavior it was trained on, was evaluated across sexes and behaviors (Figure 2B). Model generalizability was not related to the behavioral correlations in males (r=0.07, p=0.43) or females (r=0.10, p=0.27).

Models trained in males (Figure 2B, top row) to predict externalizing syndromes, attention, social, and total problems, as well as behaviors related to oppositional defiant disorder, conduct disorder, and ADHD generalize (corrected p<0.05) across those behaviors in both sexes. These models also generalize (corrected p<0.05) to predict internalizing (specifically withdrawn/depressed) syndromes and behaviors related to affective disorders in females, but not in males (dashed boxes). Additionally, models trained to predict internalizing syndromes and affective behaviors weakly generalize (corrected p<0.05) to predict some externalizing syndromes, attention problems, and behaviors related to ADHD in both sexes.

Models trained in females (Figure 2B, bottom row) to predict externalizing syndromes, attention, and social problems, as well as behaviors related to conduct disorders and ADHD generalize (corrected p<0.05) across those behaviors in both sexes. Surprisingly, these models exhibit generally greater generalizability in males (bottom left panel) than in females (bottom right panel). Moreover, models trained to predict internalizing syndromes (specifically withdrawn/depressed) and affective behaviors generalize (corrected p<0.05) to predict externalizing syndromes, thought, attention, social, and total problems, and behaviors related to oppositional defiant disorder, conduct disorder, and ADHD in males (dashed boxes). Similar results are also observed when generalizing (corrected p<0.05) within females but to a lesser extent.

These results suggest that brain-based predictive models trained in one domain can generalize to predict other related behaviors within the same domain, and this generalizability is not driven by the underlying correlations between the behaviors. Models may also generalize to predict behaviors in unrelated domains and this generalizability may be more evident across sexes rather than within sexes. The greater predictability we observe in males reflects those previously reported for cognitive predictions(22, 49), and may represent underlying differences in the strengths of brain-behavior relationships and the signal-to-noise ratio in behavioral scores.

10

Functional correlates are shared across behaviors and sexes

Pairwise regional feature weights were extracted from the models and Haufe-transformed. Correlations between these Haufe-transformed feature weights were analyzed across both sexes and all behaviors (Figure 3).

Feature weights are strongly correlated across behaviors and sexes, and the strongest correlations are observed within behavioral domains (solid boxes). One notable exception is the features involved in the prediction of anxious/depressed behaviors and somatic complaints, as well as anxiety and somatic diagnoses in both sexes as they exhibit generally weak correlations with features for all other predictions including those within the internalizing domain, but strong positive correlations with each other (rows and columns corresponding to those behaviors).

In prior work, we demonstrated that shared features predict a smaller subset of psychiatric illness-linked behaviors(24). Here, we replicate those findings in a larger set of behaviors within the mental health domain and demonstrate that even though males and females may exhibit behavioral differences, shared neurobiological features underlie their expression.

Functional connectivity within and between heteromodal association networks predict psychiatric illness-linked behaviors

Regional pairwise feature weights were summarized within the Yeo 17-network solution(50). Positive and negative feature weights were separately averaged to yield positive and negative network-level associations. For simplicity, we show the corresponding network-level features for two behaviors (withdrawn/depressed and rule-breaking) characteristic of internalizing and externalizing domains in Figures 4-5, and for all other behaviors in the supplemental materials (Figures S3-S16). Correlations between these associations across the sexes are shown in Table S4.

Functional connectivity within and between distinct heteromodal networks (e.g., temporal parietal, default, control, limbic, and attention) networks are positively and negatively associated with withdrawn/depressed behaviors (Figure 4). These associations are largely shared across the sexes ($r_{positive}$=0.89, $r_{negative}$=0.72). Functional connections positively and negatively associated with rule-breaking behaviors (Figure 5) are largely similar to those associated with withdrawn/depressed behaviors and are also shared across the sexes ($r_{positive}$=0.90 for positive, and $r_{negative}$=0.94). Similar results are also observed for all other behavioral scores predicted.

These findings are in line with prior work demonstrating that functional connections in heteromodal association networks are largely implicated in a wide range of psychiatric illnesses(51-54). We further demonstrate that shared functional connectivity correlates underlie internalizing and externalizing behaviors across the sexes. Altogether, these findings suggest that shared neurobiological correlates are likely to be observed across internalizing and externalizing behaviors.

Discussion

Brain-based predictive modeling has provided foundational insights into the neurobiological correlates of psychiatric illness(20, 55-57). While associations between functional connectivity and distinct psychiatric illnesses and behaviors have been studied extensively, prior work has not yet addressed whether those relationships are shared across the sexes. The expression of psychiatric illnesses is known to differ across the sexes, but it is not clear whether these differences map onto functional connectivity. Here, we demonstrate in a large sample of 5260 children from the ABCD dataset that functional connectivity profiles predict externalizing behaviors in males and females, but internalizing behaviors are generally only predictable in females. Models trained to predict behaviors within a given domain generalize to predict related behaviors within and between sexes. Moreover, models trained to predict externalizing behaviors in males can predict internalizing behaviors in females. Likewise, models trained to predict internalizing behaviors in females can predict externalizing behaviors in males. Across both sexes, functional connections within and between heteromodal association networks underlie the expression of internalizing and externalizing behaviors. Taken together, these results reveal that shared disruptions in functional connectivity can manifest as distinct psychiatric illness-linked behaviors across the sexes.

Psychiatric diagnoses describe clusters of problematic behaviors that tend to overlap across diagnoses(58), lack clear discernible boundaries(58), and exhibit high rates of comorbidity(59). Consequently, it is extremely difficult to isolate disorder-specific biomarkers. To understand the neurobiological processes that underlie distinct psychiatric illnesses, several different approaches have been posited. The dimensional approach proposes that psychopathology–and an individual's vulnerability to a particular psychiatric illness–can be described along distinct dimensions of psychiatric illness(60, 61). Similarly, the internalizing-externalizing model suggests that psychiatric illnesses are manifestations of internalizing and externalizing dimensions(62), where internalizing dimensions affect an individual's internal state and externalizing dimensions affect an individual's external response to the world(63). An alternative theory, the p-factor, suggests a single factor of psychopathology makes individuals broadly vulnerable to psychiatric illness and the specific illness they develop is determined by other factors(64). Regardless of how we characterize distinct psychiatric illnesses, an understanding of their underlying associations with brain-based biomarkers is crucial for the development of personalized diagnostic approaches and treatment interventions. These present analyses

suggest behavioral prediction models may be broadly generalizable across dimensional measures and diagnosis-based scales, increasing their clinical utility. This generalizability paired with the overlap in the associations between connectivity and behavior suggests a shared set of functional connections may underlie general psychopathology, and the behavioral manifestation of that psychopathology may be influenced by other factors. By moving beyond the categorical medical model and integrating dimensional measures, we can improve our understanding of the range of psychiatric symptom profiles that may be associated with variability in brain functioning.

Psychiatric illnesses and associated behaviors are harder to predict than cognition and exhibit weaker associations with neurobiological features(23, 24). Relatedly, brain-based models of internalizing behaviors and illnesses achieve weaker prediction accuracies than those of externalizing behaviors and illnesses(24). The general lack of predictability of internalizing behaviors may be related to individual differences in the signal-to-noise ratio in the associations between functional connectivity and the behaviors themselves. Furthermore, the presence of significant predictions of internalizing behaviors in females, but not in males, may be underscored by the earlier development of functional networks, and especially the heteromodal association networks, in females during childhood(8, 65). The delayed development of association networks–which drive these behavioral predictions–could in part explain the lower observed accuracies in males.

Functional disruptions in heteromodal association networks are implicated across dimensions and disorders: affective and psychotic illnesses are related to frontoparietal control, limbic, default, and attention network connectivity(24, 51-54). Here, we find functional connections within and between those networks predict individual differences in psychiatric illness-linked behaviors across both sexes. These findings suggest the existence of transdiagnostic and disorder-specific functional signatures of psychiatric illnesses and illness-linked behaviors. Finally, shared genetic and environmental influences have been shown to underlie the covariant expression of negative affect, internalizing behaviors, and externalizing behaviors(66). Our results further suggest these traits may also share neurobiological influences, which may in part be driven by genetic and environmental influences on neurobiology itself.

Sex differences in neurobiology and behavior have been studied extensively(2, 5, 7, 8, 67-80). More recently, researchers have examined sex differences in brain-behavior relationships(22,

35, 36, 45, 49). To explain the underlying factors driving these differences, sex-based and gender-based theories have been proposed. Sex-based theories posit that sex chromosomes, brain structure, the hypothalamic-pituitary-adrenal axis, immune processes, and gonadal hormones underlie sex differences in psychiatric illnesses, while gender-based theories emphasize the contributions of parental expectations, gender socialization, gender roles, gender identities, and diagnostic biases(3). Here, we demonstrate functional correlates of psychiatric illness-linked behaviors are largely shared across the sexes. Our inability to identify sex-specific correlates suggests that sex-independent and sex-specific models of psychopathology based on functional connectivity may yield generally similar predictions, as shown in our prior work focusing on cognitive behaviors(22). Furthermore, shared functional correlates are associated with the expression of internalizing and externalizing behaviors, of which, internalizing are generally more prevalent in females and externalizing in males(1). These findings suggest that differences observed in the expression of psychiatric illness-linked behaviors across the sexes during childhood may not be dependent on sex-specific functional connectivity profiles, indicating that sex- or gender- related factors are involved. Interestingly, in our current sample, we observed greater withdrawn/depressed and affective illness-linked behaviors in males than females, which contradicts the typical patterns reported in adolescents and adults(1-3). Impulse control disorders (which may be accompanied by some affective symptoms) emerge during childhood while mood disorders emerge during or after puberty(81, 82), Thus, our focus on the presentation of these behaviors and disorders in children rather than adolescents or adults may, at least in part, account for this discrepancy.

These findings are subject to several limitations. First, these analyses relied on a large community-based sample of children between the ages of 9 and 10. As these children undergo puberty and go through adolescence, they will likely exhibit changes in their behavioral expressions and brain biology, particularly in the heteromodal association networks(83-86). As such, the underlying brain-behavior relationships are subject to change throughout the course of adolescence. Second, since the ABCD dataset does not include information about gender identity or fluidity at baseline, this study only used information about each subject's parent-reported sex at birth. Throughout the course of development, males and females are exposed to gender-differentiated experiences and enculturation. Given the lack of data pertaining to gender identity or expression in the data used here, we cannot disentangle whether the observed sex differences are driven by inherent sex differences in neurobiology and/or behavior, a manifestation of gender-related differences, or a combination of the two such that innate

biological differences are further exaggerated by sociocultural and environmental factors(87). Analyses of follow-up time points in ABCD (which include the Gender Identity Questionnaire) can address these open questions. Third, this study used a single dataset which was collected entirely in the United States. The dataset was acquired at different sites (and scanners) across the country suggesting these results are somewhat generalizable, but it does not represent the global extent of racial, ethnical, or cultural diversity. While we accounted for site effects, we did not account for effects of race/ethnicity or socioeconomic status. Thus, we cannot rule out the possibility that other factors did not bias our predictive performance(88, 89). As such, further research is needed to address whether these results are generalizable across populations(88, 89) with known differences in the expression, diagnosis, and stigmatization of psychiatric illnesses(90-92). Fourth, the data release used here is an older release. Newer releases have addressed some of the issues present here (e.g., incorrect post-processing of data acquired on Phillips scanners) and include a greater number of subjects that can be included in subsequent analyses. Fifth, although a large proportion of the participants included exhibit minimal problematic behaviors, we did not consider diagnostic and medication information. Additional analyses in participants with psychiatric diagnoses or comorbidities taking medication would inform how both elevated illness risk and/or current pathology may alter the observed these brain-behavior relationships.

An understanding of the neurobiological differences that exist in the presentation of psychiatric illnesses across the sexes is crucial for the development of accurate and reliable biomarkers for diagnosis and treatment. Here, we demonstrate that functional connectivity profiles associated with psychiatric illness-linked behaviors in children do not differ between the sexes. Based on these findings, we can speculate that males and females may default to distinct gender-specific socially acceptable presentations of the same underlying sex/gender-neutral mental distress along a spectrum of internalizing-externalizing behaviors. As such, a potential solution could be the development and implementation of diagnostic tools that can recognize that both internalizing and externalizing behaviors may represent the same underlying disease.

Financial Disclosures

All authors reported no biomedical financial interests or potential conflicts of interest.

References

1.   Eaton NR, Keyes KM, Krueger RF, Balsis S, Skodol AE, Markon KE, et al. (2012): An invariant dimensional liability model of gender differences in mental disorder prevalence: evidence from a national sample. *Journal of abnormal psychology*. 121:282.

2.   Green T, Flash S, Reiss AL (2019): Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacology*. 44:9-21.

3.   Christiansen DM, McCarthy MM, Seeman MV (2022): Where sex meets gender: How sex and gender come together to cause sex differences in mental illness. *Frontiers in Psychiatry*.1322.

4.   Lau TWI, Lim CG, Acharryya S, Lim-Ashworth N, Tan YR, Fung SSD (2021): Gender differences in externalizing and internalizing problems in Singaporean children and adolescents with attention-deficit/hyperactivity disorder. *Child and adolescent psychiatry and mental health*. 15:1-11.

5.   Riecher-Rössler A (2017): Sex and gender differences in mental disorders. *The Lancet Psychiatry*. 4:8-9.

6.   Eliot L, Ahmed A, Khan H, Patel J (2021): Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience & Biobehavioral Reviews*.

7.   Gong G, He Y, Evans AC (2011): Brain connectivity: gender makes a difference. *Neuroscientist*. 17:575-591.

8.   Gur RC, Gur RE (2017): Complementarity of sex differences in brain and behavior: From laterality to multimodal neuroimaging. *J Neurosci Res*. 95:189-199.

9.   Satterthwaite TD, Wolf DH, Roalf DR, Ruparel K, Erus G, Vandekar S, et al. (2015): Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cereb Cortex*. 25:2383-2394.

10.   Weis S, Patil KR, Hoffstaedter F, Nostro A, Yeo BTT, Eickhoff SB (2019): Sex Classification by Resting State Brain Connectivity. *Cereb Cortex*.

11.   Wheelock MD, Hect JL, Hernandez-Andrade E, Hassan SS, Romero R, Eggebrecht AT, et al. (2019): Sex differences in functional connectivity during fetal brain development. *Dev Cogn Neurosci*. 36:100632.

12.   Zhang C, Dougherty CC, Baum SA, White T, Michael AM (2018): Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity. *Hum Brain Mapp*. 39:1765-1776.

13.   Hänggi J, Fövenyi L, Liem F, Meyer M, Jäncke L (2014): The hypothesis of neuronal interconnectivity as a function of brain size—a general organization principle of the human connectome. *Frontiers in human neuroscience*. 8:915.

14.   Xie S, Yang J, Zhang Z, Zhao C, Bi Y, Zhao Q, et al. (2017): The effects of the X chromosome on intrinsic functional connectivity in the human brain: evidence from Turner syndrome patients. *Cerebral Cortex*. 27:474-484.

15.   Hjelmervik H, Hausmann M, Osnes B, Westerhausen R, Specht K (2014): Resting States Are Resting Traits - An fMRI Study of Sex Differences and Menstrual Cycle Effects in Resting State Cognitive Control Networks. *Plos One*. 9.

16.   Weis S, Hodgetts S, Hausmann M (2019): Sex differences and menstrual cycle effects in cognitive and sensory resting state networks. *Brain Cognition*. 131:66-73.

17.   Mueller JM, Pritschet L, Santander T, Taylor CM, Grafton ST, Jacobs EG, et al. (2020): Dynamic community detection reveals transient reorganization of functional brain networks across a female menstrual cycle. *Network Neuroscience*.1-28.

18.   Pritschet L, Santander T, Taylor CM, Layher E, Yu S, Miller MB, et al. (2020): Functional reorganization of brain networks across the human menstrual cycle. *NeuroImage*. 220:117091.

19.   Sha Z, Wager TD, Mechelli A, He Y (2019): Common dysfunction of large-scale neurocognitive networks across psychiatric disorders. *Biol Psychiat*. 85:379-388.
20.   Dhamala E, Yeo BT, Holmes AJ (2022): Methodological Considerations for Brain-Based Predictive Modelling in Psychiatry. *Biol Psychiat*.
21.   Dhamala E, Jamison KW, Jaywant A, Dennis S, Kuceyeski A (2021): Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. *Hum Brain Mapp*. 42:3102-3118.
22.   Dhamala E, Jamison KW, Jaywant A, Kuceyeski A (2022): Shared functional connections within and between cortical networks predict cognitive abilities in adult males and females. *Hum Brain Mapp*.
23.   Ooi LQR, Chen J, Shaoshi Z, Kong R, Tam A, Li J, et al. (2022): Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *NeuroImage*.119636.
24.   Chen J, Tam A, Kebets V, Orban C, Ooi LQR, Asplund CL, et al. (2022): Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nat Commun*. 13:2217.
25.   He T, Kong R, Holmes AJ, Nguyen M, Sabuncu MR, Eickhoff SB, et al. (2020): Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*. 206:116276.
26.   Li JW, Kong R, Liegeois R, Orban C, Tan YR, Sun NB, et al. (2019): Global signal regression strengthens association between resting-state functional connectivity and behavior. *Neuroimage*. 196:126-141.
27.   Yoo K, Rosenberg MD, Hsu W-T, Zhang S, Li C-SR, Scheinost D, et al. (2018): Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *Neuroimage*. 167:11-22.
28.   Rosenberg MD, Hsu W-T, Scheinost D, Todd Constable R, Chun MM (2018): Connectome-based models predict separable components of attention in novel individuals. *Journal of Cognitive Neuroscience*. 30:160-173.
29.   Barron DS, Gao S, Dadashkarimi J, Greene AS, Spann MN, Noble S, et al. (2021): Transdiagnostic, connectome-based prediction of memory constructs across psychiatric disorders. *Cerebral Cortex*. 31:2523-2533.
30.   Wang Z, Goerlich KS, Ai H, Aleman A, Luo Y-j, Xu P (2021): Connectome-based predictive modeling of individual anxiety. *Cerebral Cortex*. 31:3006-3020.
31.   Du Y, Fu Z, Sui J, Gao S, Xing Y, Lin D, et al. (2020): NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *NeuroImage: Clinical*. 28:102375.
32.   Collin G, Nieto-Castanon A, Shenton ME, Pasternak O, Kelly S, Keshavan MS, et al. (2020): Brain functional connectivity data enhance prediction of clinical outcome in youth at risk for psychosis. *Neuroimage Clin*. 26:102108.
33.   Lichenstein SD, Scheinost D, Potenza MN, Carroll KM, Yip SW (2021): Dissociable neural substrates of opioid and cocaine use identified via connectome-based modelling. *Molecular psychiatry*. 26:4383-4393.
34.   Yip SW, Scheinost D, Potenza MN, Carroll KM (2019): Connectome-based prediction of cocaine abstinence. *American Journal of Psychiatry*. 176:156-164.
35.   Jiang R, Calhoun VD, Cui Y, Qi S, Zhuo C, Li J, et al. (2020): Multimodal data revealed different neurobiological correlates of intelligence between males and females. *Brain imaging and behavior*. 14:1979-1993.
36.   Jiang R, Calhoun VD, Fan L, Zuo N, Jung R, Qi S, et al. (2020): Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores. *Cerebral Cortex*. 30:888-900.

37.   Nostro AD, Müller VI, Varikuti DP, Pläschke RN, Hoffstaedter F, Langner R, et al. (2018): Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function*. 223:2699-2719.

38.   Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. (2018): The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. 32:43-54.

39.   Ooi LQR, Chen J, Zhang S, Kong R, Li J, Dhamala E, et al. (2022): Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *BioRxiv*.

40.   Achenbach TM (2001): Manual for ASEBA school-age forms & profiles. *University of Vermont, Research Center for Children, Youth & Families*.

41.   Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B*. 57:289-300.

42.   Chen J, Ooi LQR, Li J, Asplund CL, Eickhoff SB, Bzdok D, et al. (2022): There is no fundamental trade-off between prediction accuracy and feature importance reliability. *bioRxiv*.

43.   Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, et al. (2018): Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex*. 28:3095-3114.

44.   Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. (2002): Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33:341-355.

45.   Dhamala E, Ooi LQR, Chen J, Kong R, Anderson KM, Chin R, et al. (2022): Proportional intracranial volume correction differentially biases behavioral predictions across neuroanatomical features and populations. *NeuroImage*.

46.   Parkes L, Moore TM, Calkins ME, Cook PA, Cieslak M, Roalf DR, et al. (2021): Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure. *Translational psychiatry*. 11:1-13.

47.   Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, et al. (2014): On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*. 87:96-110.

48.   Tian Y, Zalesky A (2021): Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? *bioRxiv*.

49.   Greene AS, Gao S, Scheinost D, Constable RT (2018): Task-induced brain state manipulation improves prediction of individual traits. *Nature communications*. 9:1-13.

50.   Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. (2011): The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*. 106:1125-1165.

51.   Menon V (2011): Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in cognitive sciences*. 15:483-506.

52.   Baker JT, Dillon DG, Patrick LM, Roffman JL, Brady RO, Pizzagalli DA, et al. (2019): Functional connectomics of affective and psychotic pathology. *Proceedings of the National Academy of Sciences*. 116:9050-9059.

53.   Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. (2015): Identification of a common neurobiological substrate for mental illness. *JAMA psychiatry*. 72:305-315.

54.   Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, et al. (2018): Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun*. 9:3003.

55.   Scheinost D, Noble S, Horien C, Greene AS, Lake EM, Salehi M, et al. (2019): Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*. 193:35-45.

56.   Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, et al. (2017): Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols*. 12:506-518.

57.   Yip SW, Kiluk B, Scheinost D (2020): Toward addiction prediction: an overview of cross-validated predictive modeling findings and considerations for future neuroimaging research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 5:748-758.

58.   Hyman SE (2010): The diagnosis of mental disorders: the problem of reification. *Annual review of clinical psychology*. 6:155-179.

59.   Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, De Jonge P, et al. (2019): Exploring comorbidity within mental disorders among a Danish national population. *JAMA psychiatry*. 76:259-270.

60.   Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. (2017): The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of abnormal psychology*. 126:454.

61.   Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. (2010): Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am Psychiatric Assoc, pp 748-751.

62.   Krueger RF, Eaton NR (2015): Transdiagnostic factors of mental disorders. *World Psychiatry*. 14:27.

63.   Marshall M (2020): The hidden links between mental disorders. *Nature*. 581:19-22.

64.   Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. (2014): The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical psychological science*. 2:119-137.

65.   Shanmugan S, Seidlitz J, Cui Z, Adebimpe A, Bassett DS, Bertolero MA, et al. (2022): Sex differences in the functional topography of association networks in youth. *Proceedings of the National Academy of Sciences*. 119:e2110416119.

66.   Mikolajewski AJ, Allan NP, Hart SA, Lonigan CJ, Taylor J (2013): Negative affect shares genetic and environmental influences with symptoms of childhood internalizing and externalizing disorders. *Journal of abnormal child psychology*. 41:411-423.

67.   Cosgrove KP, Mazure CM, Staley JK (2007): Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biol Psychiatry*. 62:847-855.

68.   De Bellis MD, Keshavan MS, Beers SR, Hall J, Frustaci K, Masalehdan A, et al. (2001): Sex differences in brain maturation during childhood and adolescence. *Cereb Cortex*. 11:552-557.

69.   De Vries GJ (2004): Minireview: Sex differences in adult and developing brains: compensation, compensation, compensation. *Endocrinology*. 145:1063-1068.

70.   Earls F (1987): Sex differences in psychiatric disorders: origins and developmental influences. *Psychiatric developments*. 5:1-23.

71.   Fairweather H (1976): Sex differences in cognition. *Cognition*. 4:231-280.

72.   Gur RE, Gur RC (2016): Sex differences in brain and behavior in adolescence: Findings from the Philadelphia Neurodevelopmental Cohort. *Neurosci Biobehav Rev*. 70:159-170.

73.   Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, et al. (2014): Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci U S A*. 111:823-828.

74.   Jäncke L (2018): Sex/gender differences in cognition, neurophysiology, and neuroanatomy. *F1000Research*. 7.

75.   Lenroot RK, Giedd JN (2010): Sex differences in the adolescent brain. *Brain Cogn*. 72:46-55.

76.   Lenroot RK, Gogtay N, Greenstein DK, Wells EM, Wallace GL, Clasen LS, et al. (2007): Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage*. 36:1065-1073.

77.   Ritchie SJ, Cox SR, Shen X, Lombardo MV, Reus LM, Alloza C, et al. (2018): Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cerebral Cortex*. 28:2959-2975.

78.   Rodriguez G, Warkentin S, Risberg J, Rosadini G (1988): Sex differences in regional cerebral blood flow. *J Cereb Blood Flow Metab*. 8:783-789.

79.   Sanchis-Segura C, Ibañez-Gual MV, Adrián-Ventura J, Aguirre N, Gómez-Cruz ÁJ, Avila C, et al. (2019): Sex differences in gray matter volume: how many and how large are they really? *Biology of sex Differences*. 10:1-19.

80.   Scheinost D, Finn ES, Tokoglu F, Shen X, Papademetris X, Hampson M, et al. (2015): Sex differences in normal age trajectories of functional brain networks. *Human brain mapping*. 36:1524-1535.

81.   Paus T, Keshavan M, Giedd JN (2008): Why do many psychiatric disorders emerge during adolescence? *Nature reviews neuroscience*. 9:947-957.

82.   Schraufnagel CD, Brumback RA, Harper CR, Weinberg WA (2001): Affective illness in children and adolescents: patterns of presentation in relation to pubertal maturation and family history. *Journal of Child Neurology*. 16:553-561.

83.   Fair DA, Cohen AL, Dosenbach NU, Church JA, Miezin FM, Barch DM, et al. (2008): The maturing architecture of the brain's default network. *Proc Natl Acad Sci U S A*. 105:4028-4032.

84.   Fair DA, Cohen AL, Power JD, Dosenbach NU, Church JA, Miezin FM, et al. (2009): Functional brain networks develop from a "local to distributed" organization. *PLoS Comput Biol*. 5:e1000381.

85.   Fair DA, Dosenbach NU, Church JA, Cohen AL, Brahmbhatt S, Miezin FM, et al. (2007): Development of distinct control networks through segregation and integration. *Proc Natl Acad Sci U S A*. 104:13507-13512.

86.   Power JD, Fair DA, Schlaggar BL, Petersen SE (2010): The development of human functional brain networks. *Neuron*. 67:735-748.

87.   Eliot L (2011): The trouble with sex differences. *Neuron*. 72:895-898.

88.   Li J, Bzdok D, Chen J, Tam A, Ooi LQR, Holmes AJ, et al. (2022): Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. *Science Advances*. 8:eabj1812.

89.   Ricard J, Parker T, Dhamala E, Kwasa J, Allsop A, Holmes A (2022): Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data. *Nature Neuroscience*.1-8.

90.   Krendl AC, Pescosolido BA (2020): Countries and cultural differences in the stigma of mental illness: the east–west divide. *Journal of Cross-Cultural Psychology*. 51:149-167.

91.   Chen JA, Stevens C, Wong SH, Liu CH (2019): Psychiatric symptoms and diagnoses among US college students: A comparison by race and ethnicity. *Psychiatric services*. 70:442-449.

92.   Bailey RK, Mokonogho J, Kumar A (2019): Racial and ethnic differences in depression: current perspectives. *Neuropsychiatric disease and treatment*. 15:603.

Figure Captions

Figure 1: Experimental Workflow.
(A) Population: We included 5260 children (9-10 years old) from the Adolescent Brain Cognitive Development (ABCD) dataset, including 2689 males (51%) and 2571 females (49%). (B) Behavioral Data: We included 17 behavioral scores from the Child Behavior Checklist which includes syndrome scales and DSM-5 oriented scales. Syndrome scales included measures of composite and individual internalizing behaviors (shown in blue), composite and individual externalizing behaviors (shown in orange), other problems (shown in green), and a summary score of total problems (red). DSM-5 Oriented Scales included scores relating to affective, anxiety, somatic, oppositional defiant, conduct, and attention deficit/hyperactivity (ADHD) disorders. (C) Parcellation: We used the Schaefer cortical parcellation of 400 regions, and each region was assigned to one of 17 large-scale cortical networks. Image reproduced under a CC BY 4.0 license: https://doi.org/10.6084/m9.figshare.10062482.v1. We also included 19 subcortical regions in our analyses, which were assigned to a subcortical network. Image reproduced under a CC BY 4.0 license: https://doi.org/10.6084/m9.figshare.10063016.v1. (D) Neuroimaging Data: For each subject, we extracted their functional MRI time series data for the 400 cortical parcels and 19 subcortical parcels. Pairwise correlation was computed for all pairs of time series to obtain the estimated functional connectivity. (E) Predictive Models: Linear ridge regression models were trained to predict individual behavioral scores based on the upper triangular functional connectivity matrix in a sex-specific manner. Data were split into training and test sets. For each training set, a separate model was optimized and trained to predict each behavior. Once optimized and trained, models were evaluated for accuracy and generalizability. Prediction accuracy is defined as the correlation between the true and predicted behavioral scores in the test set for each split. We computed average accuracy by taking the mean across the distinct train-test splits. Once models were trained and tested within sexes and behaviors, we evaluated model generalizability across both sexes and all 17 behavioral scores. Model generalizability is defined as the accuracy obtained when a given model is evaluated on a population (i.e., sex) and/or behavioral score that is unique from the population/behavioral score that the model was trained on. This is distinct from model accuracy which is defined as the prediction accuracy obtained when evaluating the model on the same populations (i.e., sex) and behavioral score (using a hold-out test set) that it was trained on.

Figure 2: Predictive models of psychiatric illness-linked behaviors are accurate and generalizable across sexes and behaviors.
(A) Model Accuracy: Model prediction accuracy (correlation coefficient between true and predicted scores) for all behaviors for males (left) and females (right). Black asterisks (*) denote that model performed significantly better than chance (corrected $p<0.05$). The shape of the violin plots indicates the entire distribution of values, dashed lines indicate the median, and dotted lines indicate the interquartile range. (B) Model Generalizability: Model generalizability across sexes and behaviors for all models. Results from models trained in males are shown at the top, and models trained in females at the bottom. Results from models tested in males are shown on the left, and models tested in females on the right. Prediction accuracy (correlation coefficient between true and predicted scores) is shown for all predictions that performed better

than chance (corrected *p*<0.05) as per the color scale. For predictions that did not perform better than chance, the corresponding space is left blank. Model accuracy is shown along the diagonal for the male-trained male-tested and female-trained female-tested models (corresponding violin plots shown in Figure 2A). Dashed black boxes highlight sex differences in generalizability across behavioral domains.

Figure 3: Shared functional connectivity features underlie distinct behaviors across the sexes.

Correlation coefficient between Haufe-transformed pairwise regional feature weights from distinct models. Models trained in males are shown at the top and on the left, models trained in females are shown at the bottom and on the right. Warmer colors indicate a positive correlation and cooler colors indicate a negative correlation. Solid black boxes highlight correlations between feature weights within behavioral domains within and between sexes.

Figure 4: Shared network-level functional connections underlie withdrawn/depressed behaviors in males and females.

Positive (top) and negative (bottom) associations between network-level functional connectivity and rule-breaking behaviors in males (left) and females (right). Regional feature weights were summarized to a network-level by assigning cortical regions to one of 17 Yeo networks, and subcortical regions to a subcortical network. Colors next to the network labels along the vertical and horizontal axes correspond to the network colors from Figure 1C. Warmer colors within the heatmap indicate a positive association and cooler colors indicate a negative association. For visualization, values within each matrix were divided by the absolute maximum value across the positive and negative matrices for each sex. Correlations between positive associations across sexes, $r_{positive}$=0.89. Correlations between negative associations across sexes, $r_{negative}$=0.72.

Figure 5: Shared network-level functional connections underlie rule-breaking behaviors in males and females.
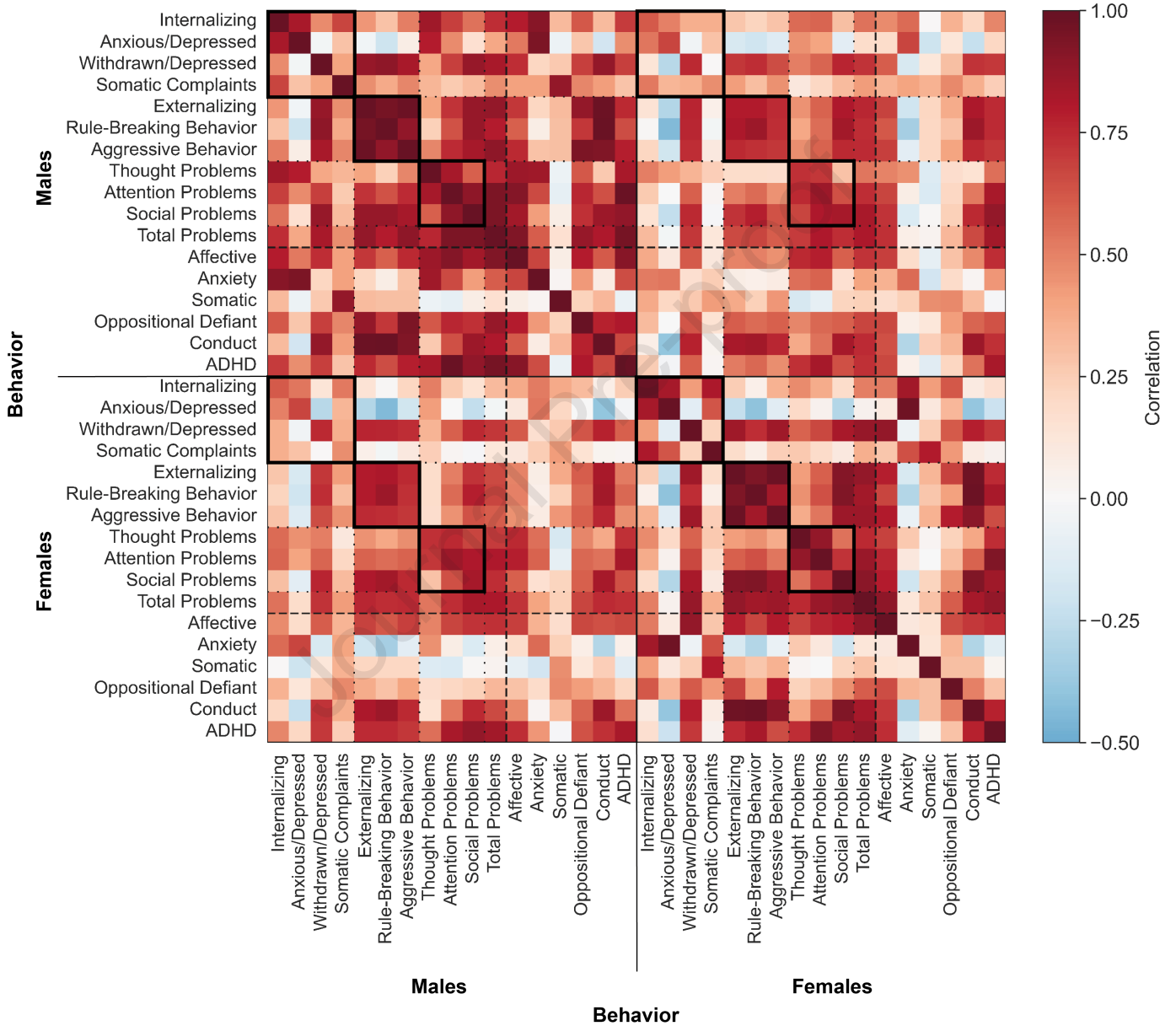
Positive (top) and negative (bottom) associations between network-level functional connectivity and rule-breaking behaviors in males (left) and females (right). Regional feature weights were summarized to a network-level by assigning cortical regions to one of 17 Yeo networks, and subcortical regions to a subcortical network. Colors next to the network labels along the vertical and horizontal axes correspond to the network colors from Figure 1C. Warmer colors within the heatmap indicate a positive association and cooler colors indicate a negative association. For visualization, values within each matrix were divided by the absolute maximum value across the positive and negative matrices for each sex. Correlations between positive associations across sexes, $r_{positive}$=0.90. Correlations between negative associations across sexes, $r_{negative}$=0.94.

Network-Level Feature Weights
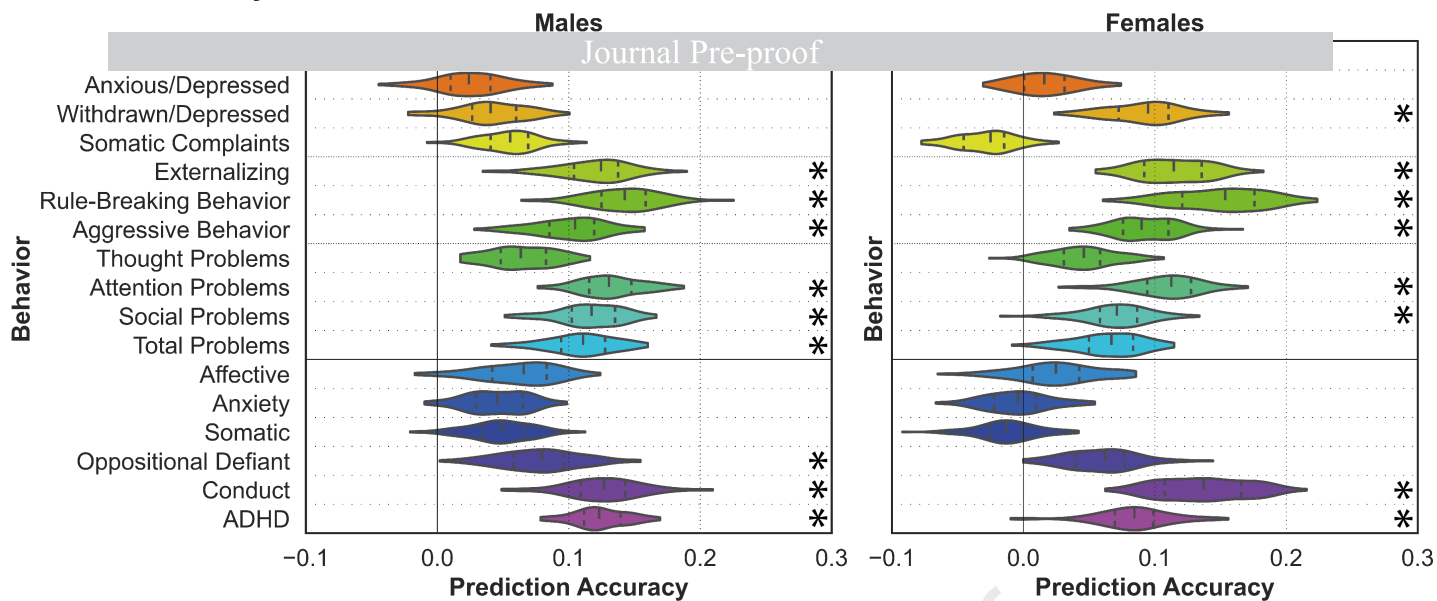
Rule-Breaking Behavior

Network-Level Feature Weights
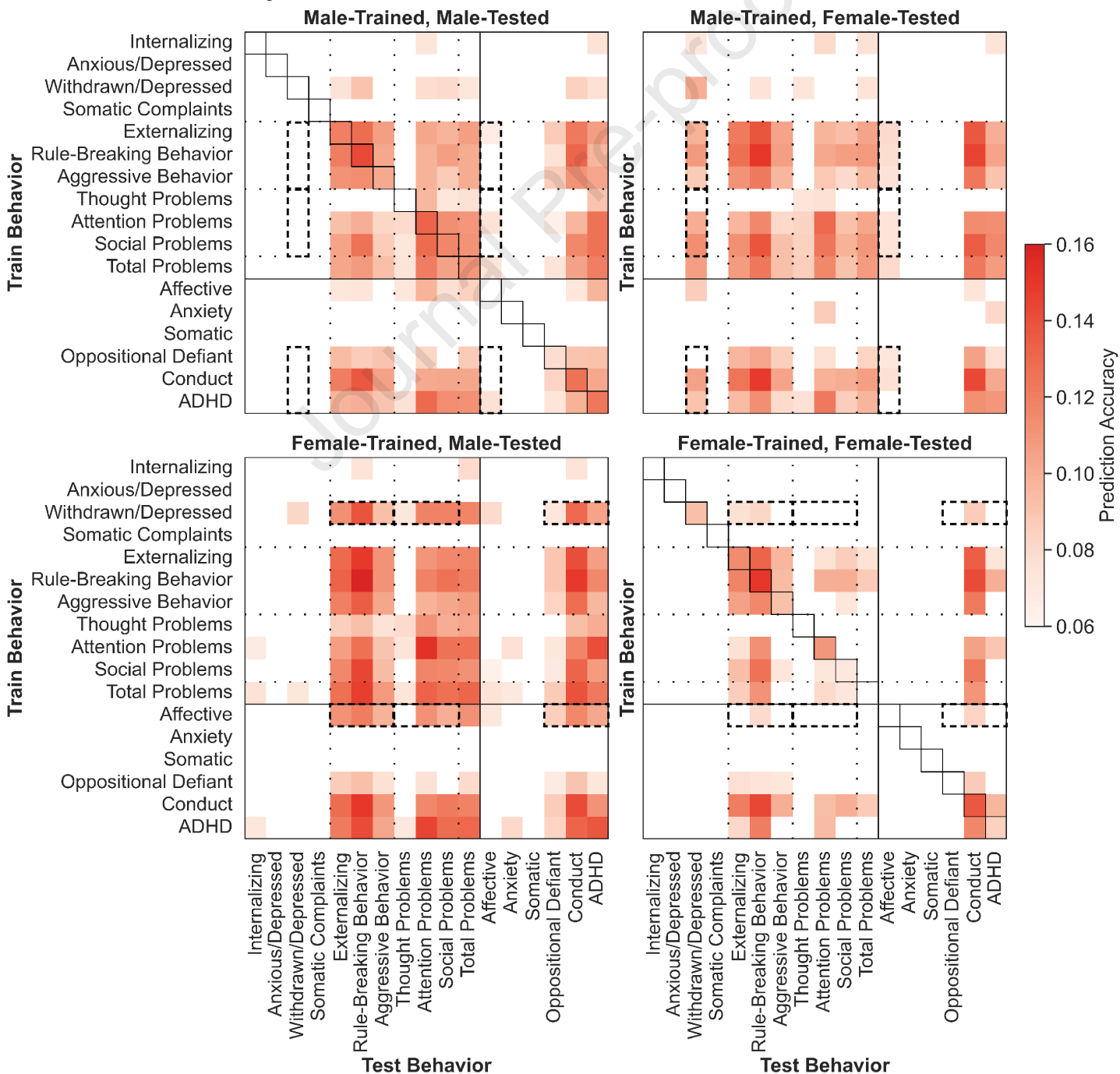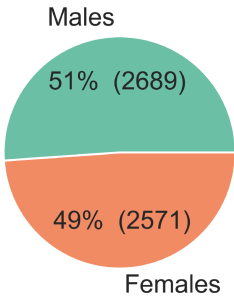
Withdrawn/Depressed
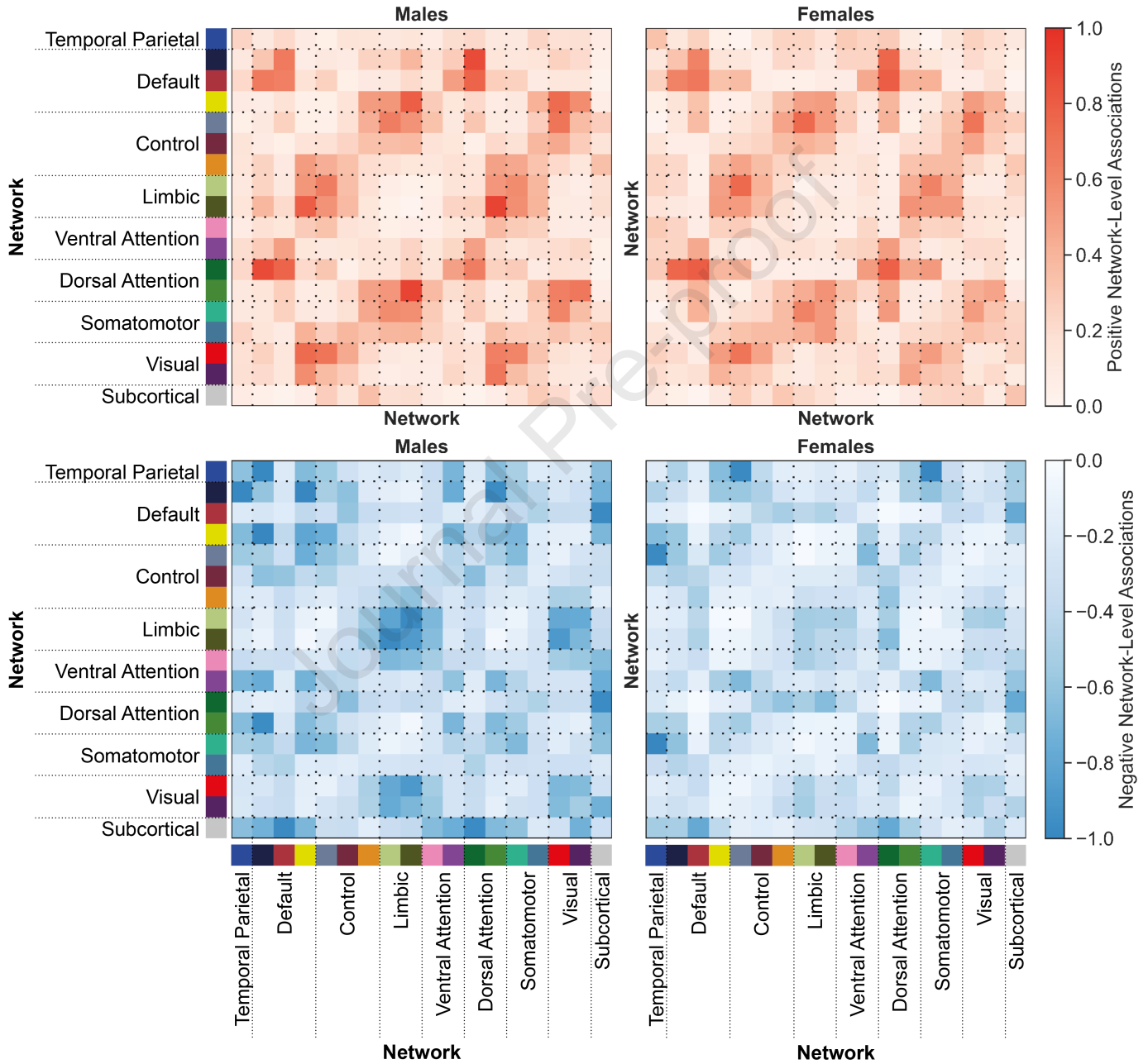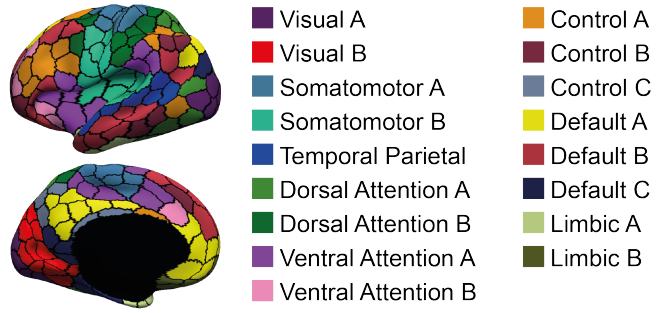
Correlation of Feature Weights

A. Model Accuracy

B. Model Generalizability

A. Population

Males

51% (2689)

49% (2571)

Females

C. Parcellation

- Visual A
- Visual B
- Somatomotor A
- Somatomotor B
- Temporal Parietal
- Dorsal Attention A
- Dorsal Attention B
- Ventral Attention A
- Ventral Attention B
- Control A
- Control B
- Control C
- Default A
- Default B
- Default C
- Limbic A
- Limbic B

- Accumbens
- Amygdala
- Brainstem
- Caudate
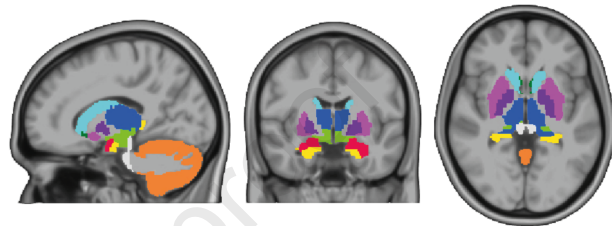- Cerebellum
- Ventral Diencephalon
- Hippocampus
- Pallidum
- Putamen
- Thalamus

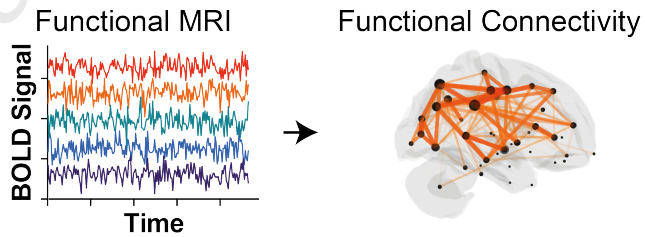## B. Behavioral Data

Syndrome Scales

Internalizing
Anxious/Depressed
Withdrawn/Depressed
Somatic Complaints
Externalizing
Rule-Breaking Behavior
Aggressive Behavior
Thought Problems
Attention Problems
Social Problems
Total Problems

DSM-5 Oriented Scales

Affective
Anxiety
Somatic
Oppositional Defiant
Conduct
ADHD

D. Neuroimaging Data

Functional MRI

BOLD Signal

Time

Functional Connectivity

E. Predictive Models: Linear Ridge Regression

Train/Test Split

ABCD (n=5260)

100x

Train
Test

100x

Train
Test

Cross Validation: Train Set

Evaluation: Test Set

For each sex, separate models were optimized and trained (using the train sets) to predict each behavior. Then, all models were evaluated (using the test sets) across sexes and across behaviors.