

---

# Hierarchical neurocognitive model of externalizing and internalizing comorbidity

---

In the format provided by the  
authors and unedited

## Supplementary Method

**IMAGEN.** IMAGEN is a large-scale longitudinal neuroimaging–genetics cohort study (N = 2000 at age 14, N = 1300 at age 19, N = 1100 at age 23) conducted to understand the biological basis of individual variability in psychological and behavioural traits and their relationship to common psychiatric disorders. The study involves a thorough neuropsychological, behavioural, clinical and environmental assessment of each participant. Participants also undergo biological characterisation with the collection of T1-weighted structural magnetic resonance imaging, task-based functional magnetic resonance imaging and genetic data. In the current investigation, we used task and resting-state MRI data, genetic data, and behaviour data. Notably, as a population-based approach, IMAGEN has balanced sample sizes for male and female participants (based on self-reported sex).

**DAWBA and SDQ.** Behavioural symptoms of the IMAGEN participants were assessed using screening questions from the Development and Well-Being Assessment (DAWBA) and the Strengths and Difficulties Questionnaire (SDQ). DAWBA is a wide-ranging psychiatric screening questionnaire<sup>1</sup> and has previously been used to define subthreshold clinical symptoms in neuroimaging studies of subclinical psychopathology<sup>2</sup>. The SDQ was also used in the present investigation, as this questionnaire contributes to the assignment of diagnostic status in the DAWBA<sup>1</sup>. At age 14, the parent-rated externalizing symptoms contained attentional deficit/hyperactivity disorder (ADHD, 23 items), oppositional defiance disorder (ODD, 11 items), conduct disorder (CD, 10 items) and autism spectrum disorder (ASD, 7 items). The child-rated internalizing symptoms included general anxiety disorder (GAD, 7 items), depression (8 items), specific phobia (SP, 13 items) and eating disorder (ED, 5 items). The full set of psychiatric questions asked in the present investigation can be found in Supplementary Table 1. The choice of using different versions of questionnaires (i.e. parents-rated externalizing symptoms and child-rated internalizing symptoms) at age 14 was based on findings that externalizing problem scores from parents are more reliable than those from children themselves, and vice versa<sup>3</sup>. At age 19, however, as parents-rated questionnaires were unavailable, we used child-rated questionnaires for both externalizing and internalizing symptoms (also see Supplementary Table 1).

DAWBA also provides a diagnostic output for common psychiatric disorders based on the likelihood of a clinical diagnosis being made following rating. Out of the 1,750 IMAGEN participants at age 14, 134 were found to have a high risk for at least one diagnosis (i.e., scored 4 or 5, with over

50% chance of being diagnosed), and 39 participants met the criteria for two or more diagnoses. More specifically, 93 participants were likely to have one or more externalizing disorders, including 24 with attention-deficit/hyperactivity disorder (ADHD), 45 with oppositional defiant disorder (ODD), 59 with conduct disorder (CD), and 1 with autism spectrum disorder (ASD). In addition, 46 participants were likely to have one or more internalizing disorders, including 16 with general anxiety disorder (GAD), 21 with depression, 5 with eating disorder (ED), and 14 with specific phobia (SP) (see **Extended Table 1** for more details).

**Monetary incentive delay (MID) task.** Participants performed a modified version of the Monetary Incentive Delay task (**Supplementary Fig.1**) to examine neural responses to reward anticipation and reward outcome<sup>4</sup>. The task consisted of 66 trials, each lasting 10 seconds. In each trial, participants were presented with one of three cue shapes (cue, 250 ms), denoting whether a target (white square) would subsequently appear on the left or right side of the screen and the potential number of points (0, 2 or 10) that could be won in that trial. After a variable delay (4,000–4,500 ms) of fixation on a white crosshair, participants were instructed to respond with a left/right button-press as soon as the target appeared. Feedback on the number of points won during the trial was presented for 1,450 ms after the response (**Supplementary Fig. 1**). Task difficulty, determined by the target duration (ranging from 100 to 300 ms), was adjusted individually using a tracking algorithm to ensure that each participant successfully responded on approximately 66% of trials. Participants had first completed a practice session outside the scanner (~5 minutes), during which they were informed that for every 5 points won, they would receive one food snack in the form of small chocolate candies. The current study used the task conditions consisting of hit anticipation, hit feedback and miss feedback.

**Stop-signal task (SST).** Participants performed an event-related stop-signal task (**Supplementary Fig.2**) designed to study neural responses to successful and unsuccessful inhibitory control<sup>5</sup>. The task was composed of Go trials and Stop trials. During Go trials (83%; 480 trials), participants were presented with arrows pointing either to the left or to the right and were instructed to make a button response with their left or right index finger according to the arrow direction. In the unpredictable Stop trials (17%; 80 trials), the arrows pointing left or right were followed (on average 300 ms later) by arrows pointing upwards; participants were instructed to inhibit their motor responses during these trials. A tracking algorithm adjusted the time interval between the Go signal and Stop signal onsets

according to each subject's performance on previous trials (average percentage of inhibition over previous Stop trials, recalculated after each Stop trial), resulting in 50% successful and 50% unsuccessful inhibition trials. The inter-trial interval was 1,800 ms. The tracking algorithm of the task ensured that subjects were successful on 50% of Stop trials and worked at the edge of their own inhibitory capacity. The current study used the SST task measures consisting of stop success, stop failure and go wrong.

**Emotional face task (EFT).** The emotional face task was adapted from Grosbras et al.<sup>6</sup>. Participants watched 18-second blocks of either a face movie (depicting anger or neutrality) or a control stimulus. Each face movie showed black and white video clips (200-500ms) of a male or female face. Five blocks of angry and neutral expressions were interleaved with nine blocks of the control stimulus. Each block contained eight trials of 6 face identities (3 female). The same identities were used for the angry and neutral blocks. The control stimuli were black and white concentric circles expanding and contracting at various speeds that roughly matched the contrast and motion characteristics of the face clips. The current study used the EFT task conditions consisting of neutral and angry faces.

**Image acquisition.** Functional MRI data were acquired at eight IMAGEN assessment sites with 3T MRI scanners of different manufacturers (Siemens, Philips, General Electric, Bruker). The scanning variables were specifically chosen to be compatible with all scanners. The same scanning protocol was used in all sites. In brief, high-resolution T1-weighted 3D structural images were acquired for anatomical localisation and co-registration with the functional time series. In addition, Blood-oxygen-level-dependent (BOLD) functional images were acquired with gradient-echo, echo-planar imaging (EPI) sequence. For any tasks, each volume consisted of 40 slices aligned to the anterior commission/posterior commission line (2.4 mm slice thickness, 1 mm gap). The echo-time (TE) was optimised (TE=30 ms, repetition time (TR)=2,200 ms) to provide reliable imaging of subcortical areas.

**Task-based functional image pre-processing.** Task-based functional MRI data were first pre-processed with SPM8 (Statistical Parametric Mapping, <http://www.fil.ion.ucl.ac.uk/spm>). Spatial preprocessing included: 1) slice time correction to adjust for time differences due to multi-slice imaging acquisition, 2) realignment to the first volume in line, 3) non-linearly warping to the MNI space (based on a custom EPI template (53x63x46 voxels) created out of an average of the mean images of 400 adolescents), 4) resampling at a resolution of 3x3x3 mm<sup>3</sup> and 5) smoothing with an isotropic Gaussian

kernel of 5 mm full-width at half-maximum.

**Network construction.** The CONN toolbox (Version 16.h) was utilised to estimate the condition-specific functional connectivity with the weighted generalised linear model (wGLM) method. Task-condition regressors and 21 covariate regressors were first regressed out from the raw BOLD signal of each region of interest (ROI). The 21 covariate regressors included 12 motion regressors (3 translations, 3 rotations, 3 translations shifted 1 TR before, and 3 translations shifted 1 TR later) and 9 additional columns corresponding to the long-term effects of the movement (3 nuisance variables for the white matter and 6 nuisance variables for ventricles, commonly referred to as CompCor correction<sup>7</sup>). The residual signals were further fed into wGLMs to investigate conditional time-series correlations (i.e., the conditional functional connectivity) between any pairs of ROIs. For each condition, a temporal weight function was calculated as the corresponding but now rectified task-condition regressor, considering only time-points expected with positive BOLD signals. This approach not only amplifies the expected hemodynamic delay to each task condition but also de-weights the initial and final scans within when estimating functional correlation measures to avoid spurious jumps in BOLD signal, as well as reducing the potential cross-talk between adjacent task conditions<sup>8</sup>. Following the above procedure, ROI-to-ROI functional connectivities were calculated based on the 268-node functional brain atlas template<sup>9</sup> (**Supplementary Fig. 3**).

**Reliability assessment using permutation tests.** To investigate which task conditions could provide reliable cross-disorder edges, we implemented permutation tests to evaluate if identified cross-disorder edges from each task condition were indeed informative, i.e., if the number of edges identified for the given condition was significantly larger than a random discovery (**Supplementary Fig. 5**). Due to the time-consuming of the proposed CPM analysis (1000 repetitions of fifty-fold cross-validation as described above), the number of permutations was set as 1000, which is sufficient to provide an accurate estimation of P-value as small as 0.01. Besides, this permutation process was also used to provide unbiased P-values for the following association of the stratificational network with behavioural symptoms.

**Generalization datasets.** To investigate whether the *NP* factor identified with the adolescent IMAGEN dataset using the task-based connectomes could be generalised into other developmental periods and fMRI states, we utilised multiple large-scale population-based datasets (the Adolescent Brain Cognitive Development cohort<sup>10</sup>, **ABCD**; the Human Connectome Project<sup>11</sup>, **HCP**) and clinical case-control datasets (The Stratify/ESTRA project<sup>12</sup>, **STRATIFY/ESTRA**; **ADHD-200**<sup>13</sup>, **ABIDE II**; **XiNan**).

**ADHD-200.** The ADHD-200 Sample is a grassroots initiative dedicated to accelerating the scientific community's understanding of the neural basis of ADHD (ages 7-21). Inclusion criteria included: no history of neurological diseases and other chronic medical conditions; estimates of full-scale IQ above 80; psychostimulant drugs were withheld at least 24-48 hours before scanning. Data was downloaded from the ADHD-200 consortium website ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200](http://fcon_1000.projects.nitrc.org/indi/adhd200)). In the present study, we used data from 4 sites (Peking University, Kennedy Krieger Institute, New York University Child Study Center and Oregon Health & Science University) which recruited both patients with ADHD and typically-developed controls (TD).

In total, there were 228 patients and 292 controls. For all analyses of the ADHD-200 data, we included site, handedness and sex as covariates.

**ABCD.** The dataset used for this study was selected from Annual Curated Data Release (<https://data-archive.nimh.nih.gov/abcd>) of the Adolescent Brain Cognitive Development (ABCD) cohort, which recruited 11,875 children between 9–11 years of age from 21 sites across the United States<sup>14</sup>. Magnetic resonance imaging (MRI) data in the ABCD study were collected from different 3T scanner platforms (i.e. Siemens Prisma, General Electric (GE) MR750 and Philips Achieva dStream). To minimize biases introduced by multiple platforms, we only included MRI data from the most frequent manufacturer Siemens Prisma, i.e. 5968 participants from 13 sites. By examining the similarity of brain activations across these 13 sites, we further selected 2326 participants from 4 sites with consistent activation patterns. After quality control<sup>15</sup>, 1966 participants of the monetary incentive delay (MID) task and 1837 participants of the stop signal task (SST) were included in the following analysis. ABCD has balanced sample sizes for males and females (based on self-reported sex) (Table 1). To construct the *NP* factor in the ABCD dataset, with the same 'positive-positive' edges that were used to establish the *NP* factor in the IMAGEN cohort, we extracted the corresponding functional

connectivity (FC) of reward anticipation and reward positive feedback from the MID task, and FC of the stop success and stop failure from the SST. The sum of FC for the MID and SST was the corresponding *NP* factor for the ABCD, respectively. For the psychiatric symptoms, the Parent Child Behavior Checklist Scores (abcd\_cbcls01) was used to assess the dimensional psychopathology in children<sup>16</sup>. The summed scores of externalizing and internalizing symptoms were used in the following analysis. The ABCD Parent Diagnostic Interview for DSM-5 Full provides a diagnostic output for common psychiatric disorders (abcd\_ksad01). The diagnosis of ASD was provided in a clinical assessment questionnaire (abcd\_screen01). Please note that as the morbidity of specific phobia (SP) (21.5%) with abcd\_ksad01 in the ABCD dataset was much higher than that of other pediatric epidemiologic investigations of SP (4.8%)<sup>17,18</sup>, we thus excluded this diagnosis information in the clinical relevance analysis. For all analyses of the ABCD data, we included site, family, handedness and sex as covariates in a mixed model according to the practice with the ABCD dataset<sup>19</sup>.

**ADHD-200.** The ADHD-200 Sample is a grassroots initiative dedicated to accelerating the scientific community's understanding of the neural basis of ADHD (ages 7-21). Inclusion criteria included: no history of neurological diseases and other chronic medical conditions; estimates of full-scale IQ above 80; psychostimulant drugs were withheld at least 24-48 hours before scanning. Data was downloaded from the ADHD-200 consortium website ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200](http://fcon_1000.projects.nitrc.org/indi/adhd200)). In the present study, we used data from 4 sites (Peking University, Kennedy Krieger Institute, New York University Child Study Center and Oregon Health & Science University) which recruited both patients with ADHD and typically-developed controls (TD).

In total, there were 228 patients and 292 controls. For all analyses of the ADHD-200 data, we included site, handedness and sex as covariates.

**ABIDE II.** The Autism Brain Imaging Data Exchange (ABIDE) II sample was used in the present study<sup>20</sup>. Approval was required by the respective site Institutional Review Board (IRB). All participants' autistic symptoms (measured using ADOS), as well as resting-state functional magnetic resonance imaging (rsfMRI) data, were collected from multiple data acquisition sites across the globe. More details on data acquisition can be found on the initiative's website ([fcon\\_1000.projects.nitrc.org/indi](http://fcon_1000.projects.nitrc.org/indi)). In the present study, we used data from 6 sites (Erasmus University Medical Center Rotterdam, Georgetown University, Kennedy Krieger Institute, NYU La

ngone Medical Center, Oregon Health and Science University and San Diego State University). For all analyses of the ABIDE II data, we included site, handedness and sex as covariates.

**XiNan.** Patients with MDD were recruited from the outpatient department of the First Affiliated Hospital of Chongqing Medical School in Chongqing, China. All were diagnosed according to the Structured Clinical Interview for DSM-IV. They were also assessed for disease severity using the Hamilton Depression Rating Scale (HAMD) and Beck Depression Inventory (BDI). All of the subjects in the control group did not meet DSM-IV criteria for any psychiatric disorders and did not use any drugs that could affect brain function. This study was approved by the Research Ethics Committee of the Brain Imaging Center of Southwest University and First Affiliated Hospital of Chongqing Medical School. Informed written consent was obtained from each subject.

**STRATIFY/ESTRA.** The Stratify project recruited patients (ages 19–25) with anorexia nervosa, alcohol use disorder, bulimia nervosa, major depression, and controls at three sites (Berlin, London, and Southampton). The proportions of males and females (based on self-reported sex) varied across different mental health disorder groups (Table 1). Furthermore, the protocol of Stratify was harmonized to match with the IMAGEN protocol. The stratify datasets collected task-based neuroimaging data of the SST and MID task. After quality control (the same QC procedures with the ABCD dataset <sup>15</sup>), there are 267 patients and 46 controls of the MID and 380 patients and 64 controls of the SST were included in the following analysis. For all analyses of the Stratify data, we included site, handedness and sex as covariates.

**Genotyping for the IMAGEN study.** DNA purification and genotyping were performed by the Centre National de Génotypage. DNA was extracted from whole-blood samples (~10 mL) preserved in BD Vacutainer EDTA Tubes (Becton, Dickinson and Company) using the Gentra Puregene Blood Kit (QIAGEN Inc.) according to the manufacturer's instructions. SNPs with call rates of <98%, minor allele frequency <1%, or deviation from the Hardy–Weinberg equilibrium ( $p < 1.00 \times 10^{-4}$ ) were excluded from the analyses. Individuals with an ambiguous sex code, excessive missing genotypes (failure rate >2%), and outlying heterozygosity (heterozygosity rate of 3 SDs from the mean) were also

excluded.

**Polygenic risk scores.** To calculate the polygenic risk scores (PRS) of depression, ADHD and intelligence, we used previously published GWASs of depression<sup>11</sup>, and ADHD<sup>12</sup> in *Nature Genetics*. Regarding these discovery samples for GWAS, the depression study consists of 135,458 cases and 344,901 controls and the ADHD study consists of 20,183 cases and 35191 controls. We then employed the PRSice software (<http://prsice.info/>) to calculate the corresponding PRS. The clumping process was applied to only keep SNPs with the smallest p-value for each linkage disequilibrium block (combining with a sliding window process to exclude any less significant SNPs with an  $r^2 < 0.1$  in 250 kb windows). PRSs were calculated at P-value thresholds between  $P_T = 0$  and  $P_T = 0.5$  at increments of 0.01, and we used the mean PRSs of depression and ADHD for subsequent analyses<sup>14</sup>.

**Cognition-behavior phenotypes: CANTAB.** The Cambridge Cognition Battery (CANTAB, <http://www.cambridgecognition.com/>) comprised of the spatial working memory (**SWM**) task (number of errors and strategies), the Cambridge Guessing Task (**CGT**) (risk taking, quality of decision making, delay aversion, deliberation time, overall proportion bet, risk adjustment), the Rapid Visual Information Processing (**RVP**) task, and the Affective Go-No go (**AGN**) task (mean correct latency for positive and negative stimuli, number of omissions errors for positive and negative stimuli). The CGT quality of decision making is the proportion of trials on which the subject chooses the most likely outcome. The CGT deliberation time is the reaction time to choose the colour of the box. The overall bet is the overall bet across the trials. The CGT risk-taking: mean proportion of available points the subject stakes at each trial. The CGT delay aversion is the difference between the risk-taking score in the descending and the ascending conditions. The CGT risk adjustment is the degree to which a subject adjusts the risk taking according to the ratio of coloured boxes, calculated as:  $[2 \times (\text{proportion of points staked (\% at 9:1)} + (\text{\% 8:2}) - (\text{\% 7:3}) - 2 \times (\text{\% 6:4}))] / \text{CGT risk taking}$ . The RVP is a 10-minute test which measures sustained attention by presenting a rapid stream of digits and requiring participants to detect target sequences. A white box is displayed in the centre of the screen in which digits 2-9 are rapidly presented at 100 digits per minute (Supplementary Figure 2C). Participants are required to detect target sequences (e.g. 2-4-7, 3-5-7 or 4-6-8) and respond to this target sequence as quickly as possible. Outcome measures include a signal detection theor

y measure of target sensitivity and mean response latency. **IQ.** The intelligence was measured with the Wechsler Intelligence Scale for Children-IV (WISC-IV)<sup>15</sup>, consisting of the fluid intelligence and the verbal intelligence. **Delay discounting.** We used the monetary-choice questionnaire as described by Kirby<sup>16</sup>, an efficient and reliable measurement of delay discounting that has been validated in adolescents<sup>17</sup>. We estimated for each subject the  $k$  values which reflect how one discounts a reward value with the delay required to obtain it. The questionnaire contains 27 dichotomous-choice items pitting a smaller immediate reward against a larger delayed reward for three levels of reward magnitude (small, medium, and large). Higher  $k$  coefficients in a hyperbolic discounting equation for each reward level represent greater preference for small immediate rewards and higher impulsivity. The geometric mean was calculated and logarithmically transformed to use in our analyses.

**Personality: SURPS.** The Substance Use Risk Personality Scale (SURPS, 23 items, self-questionnaire) was used to measure “sensation seeking”, “impulsivity”, “anxiety sensitivity” and “negative thinking” sub-scores, and has been shown to be related to substance use in adolescents<sup>18</sup>. **NEO.** The NEO Personality inventory (NEO-PI, 60 items, self-questionnaire) explores the big five domain of personality: Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness<sup>19</sup>. **TCI.** The Temperament and Character Inventory-Revised (TCI, 36 items)<sup>20</sup> was utilised to measure “excitability”, “impulsiveness”, “reserve”, “disorderliness”, and their sum-up “novelty seeking”.

**Substance use: Alcohol.** The alcohol abuse was assessed using the screening questions from the Alcohol Use Disorders Identification Test (AUDIT, 10 items)<sup>21</sup>. The AUDIT was developed by the World Health Organization as a simple way to screen and identify people who are at risk of developing alcohol problems. The AUDIT test focuses on identifying the preliminary signs of hazardous drinking and mild dependence. It is used to detect alcohol problems experienced within the last year. It is one of the most accurate alcohol screening tests available. **Smoking.** The smoking behaviour was assessed as the frequency (i.e. cigarettes per day) of smoking during the last 30 days with the ESPAD (European School Survey Project on Alcohol and Other Drugs)<sup>22</sup>.

**Environment risk: CTQ.** The childhood trauma questionnaire (CTQ)<sup>23</sup> was used to assess childhood maltreatment across childhood and adolescence. It consists of five domains: emotional abuse,

emotional neglect, physical abuse, physical neglect, and sexual abuse. The scores of five domains was summed for a total CTQ score; the higher the score the greater the severity of maltreatment. **School Bully.** The school bully behaviour was measured with the adapted questionnaire, based on the Health Behaviour in School-aged Children (HBSC). These questions were initially utilized in the revised Olweus Bully/Victim Questionnaire<sup>24</sup>. **Family Stress.** Family stress was measured by the family stress and socio-economic item from the Development And Well-Being Assessment (DAWBA). Greater score at this item indicates greater family stress. **Family drinking.** The family drinking was measured with the parent AUDIT.

**Other risk: BMI.** Recorded weight and height were used to calculate the BMI [weight(kg)/height(m)<sup>2</sup>]. **PBQ.** The pregnancy and Birth Questionnaire (PBQ) was used to collect information during the pregnancy, consisting of **Mather and Father Data**, **Medical Condition of Mother** ("during pregnancy, did the mother take any prescribed medication during pregnancy?"), **Smoking Exposure** ("How many cigarettes did the MOTHER smoke per day before pregnancy?"), **Birth Weight** ("What was the birth weight of the child?").

## References

1. Goodman, R. The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry* **38**, 581-586 (1997).
2. Galinowski, A., *et al.* Resilience and corpus callosum microstructure in adolescence. *Psychol Med* **45**, 2285-2294 (2015).
3. Crijnen, A.A., Achenbach, T.M. & Verhulst, F.C. Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. *J Am Acad Child Adolesc Psychiatry* **36**, 1269-1277 (1997).
4. Knutson, B., Fong, G.W., Adams, C.M., Varner, J.L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**, 3683-3687 (2001).
5. Bari, A. & Robbins, T.W. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog Neurobiol* **108**, 44-79 (2013).
6. Grosbras, M.H. & Paus, T. Brain networks involved in viewing angry hands or faces. *Cereb Cortex* **16**, 1087-1096 (2006).
7. Behzadi, Y., Restom, K., Liau, J. & Liu, T.T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90-101 (2007).
8. Whitfield-Gabrieli, S. & Nieto-Castanon, A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect* **2**, 125-141 (2012).
9. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R.T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* **82**, 403-415 (2013).
10. Marek, S., *et al.* Identifying Reproducible Individual Differences in Childhood Functional Brain Networks: An ABCD Study. *Developmental Cognitive Neuroscience* (2019).
11. Van Essen, D.C., *et al.* The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222-2231 (2012).
12. Quinlan, E.B., *et al.* Identifying biological markers for improved precision medicine in psychiatry. *Mol Psychiatry* **25**, 243-253 (2020).
13. Consortium, H.D. The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. *Front Syst Neurosci* **6**, 62-62 (2012).
14. Casey, B.J., *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci* **32**, 43-54 (2018).
15. Xiang, S., *et al.* A novel analytical decoder of BOLD signals for dissociating latent neurobehavioral processes. *bioRxiv*, 2021.2008.2025.457728 (2021).
16. Achenbach, T.M. & Rescorla, L.A. The Achenbach system of empirically based assessment (ASEBA) for ages 1.5 to 18 years. in *The use of psychological testing for treatment planning and outcomes assessment* 179-214 (Routledge, 2014).
17. Wardenaar, K.J., *et al.* The cross-national epidemiology of specific phobia in the World Mental Health Surveys. *Psychol Med* **47**, 1744-1760 (2017).
18. salehi, M., *et al.* The Lifetime Prevalence, Risk Factors, and Co-Morbidities of Specific Phobia Among Pediatric Population: A Cross-Sectional National Survey. *Clinical Medicine Insights: Psychiatry* **13**(2022).
19. Dick, A.S., *et al.* Meaningful associations in the adolescent brain cognitive development study. *Neuroimage* **239**, 118262 (2021).
20. Di Martino, A., *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry* **19**, 659-667 (2014).

21. Jia, T., *et al.* Neurobehavioural characterisation and stratification of reinforcement-related behaviour. *Nat Hum Behav* **4**, 544–558 (2020).
22. Greene, A.S., Gao, S., Scheinost, D. & Constable, R.T. Task-induced brain state manipulation improves prediction of individual traits. *Nature communications* **9**, 2807 (2018).

## Supplementary figures

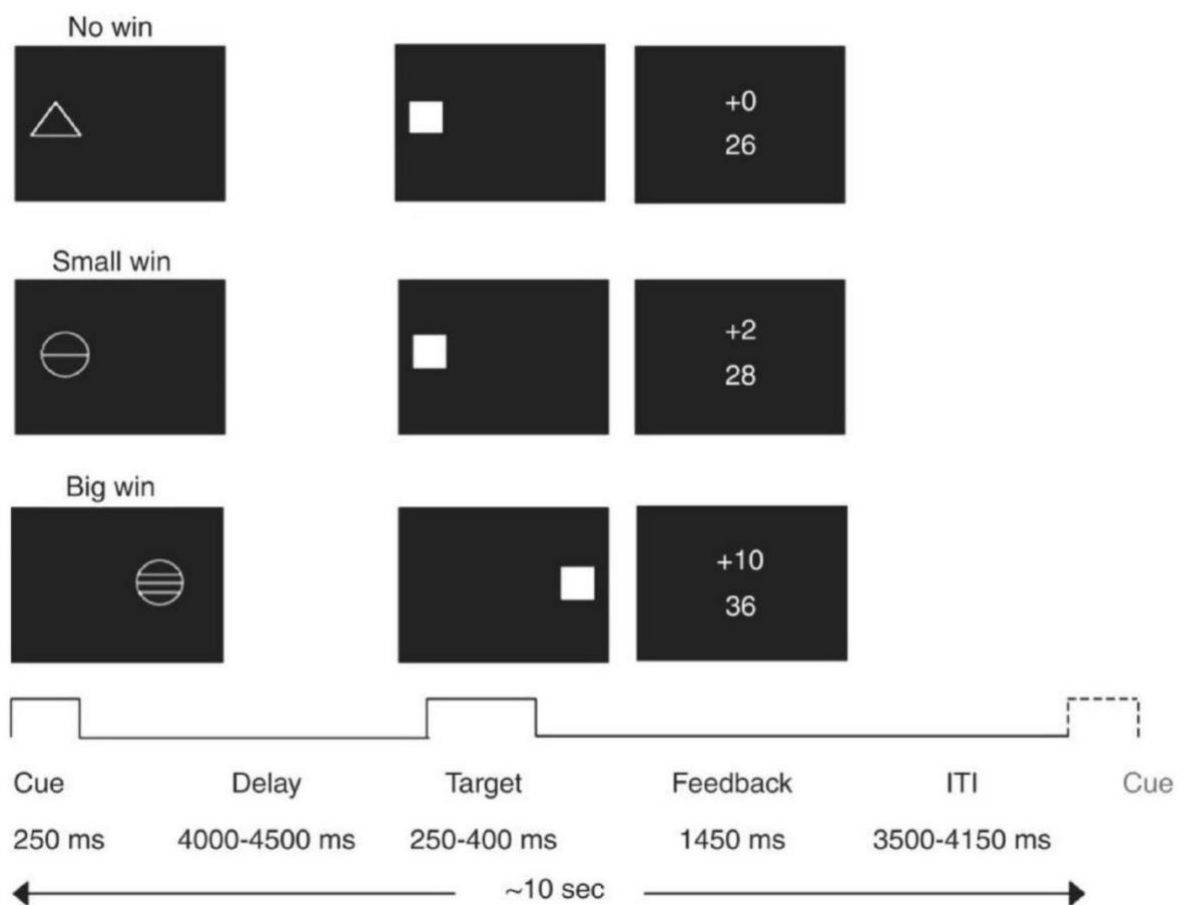
Figure S1. The task design of the Monetary Incentive Delay (MID) Task.

Figure S2. The task design of the stop signal Task (SST).

Figure S3. Ten large-scale canonical networks for anatomical interpretation.

Figure S4. The analysis flow of the connectome-based predictive model (CPM).

Figure S5. The analysis flow of the permutation analysis.



**Figure S1. The task design of the Monetary Incentive Delay (MID) Task.**

The figure of experimental paradigm was adapted from a previous publication <sup>21</sup>.

Number of Go trials/ Stop trials:  
480/80

The diagram illustrates the sequence of events in a Stop-Signal task. It shows a series of trials progressing along a diagonal timeline labeled 'time' at the bottom left. The trials alternate between 'Go trial' and 'Stop trial'. Each trial consists of a stimulus (a white arrow pointing right for Go, and a white arrow pointing left for Stop) followed by an 'ITI' (Inter-Trial Interval) screen with a white plus sign. The duration of the ITI is specified as 300 msec. The duration of the Stop trial is specified as 'RT' (Reaction Time). The duration of the Go trial is specified as 'duration = ( 1800 msec ) + (hold period - RT)'. A box contains the text: '\*adapted according to performance in preceding stop trial: StopInhibit: + 50 msec StopFail : - 50 msec'. An arrow points to the right, indicating the sequence continues.

Stop trial

ITI

StopInhibit: duration = 300 msec  
StopFail: duration = RT

duration = variable SSD\*

\*adapted according to performance in preceding stop trial:  
StopInhibit: + 50 msec  
StopFail : - 50 msec

Go trial

ITI

duration = ( 1800 msec ) + (hold period - RT)

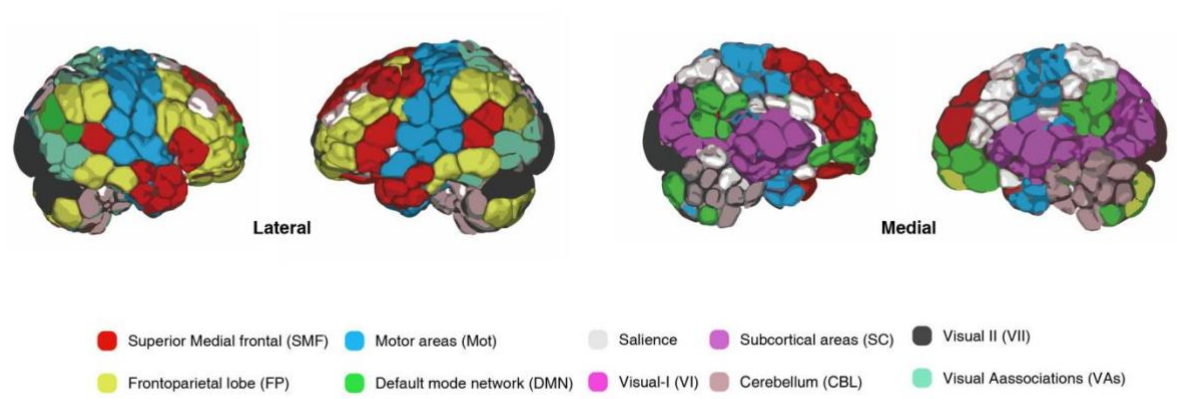
Go trial

ITI

duration = RT (hold period = max. 1000 msec)

time

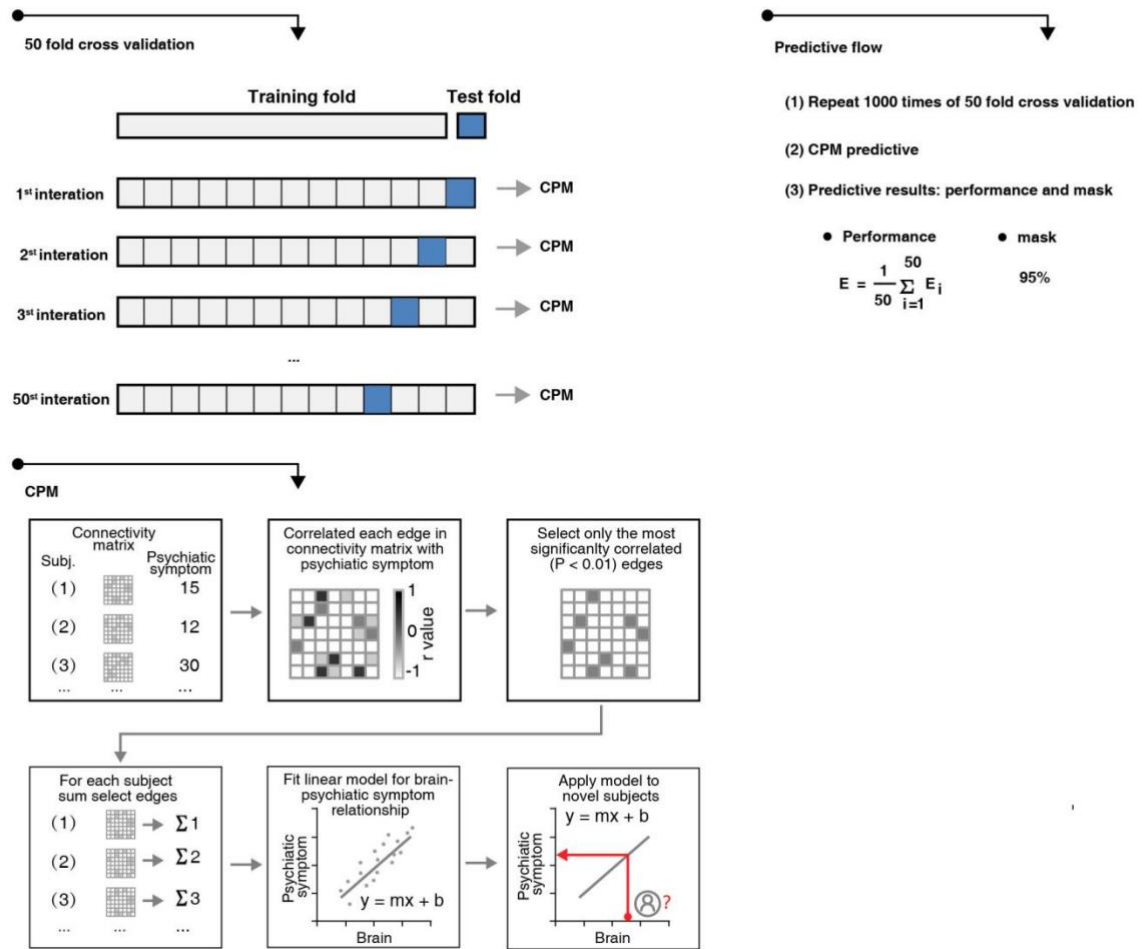
The figure of experimental paradigm was adapted from a previous publication <sup>21</sup>.



**Figure S3. Ten large-scale canonical networks for anatomical interpretation.**

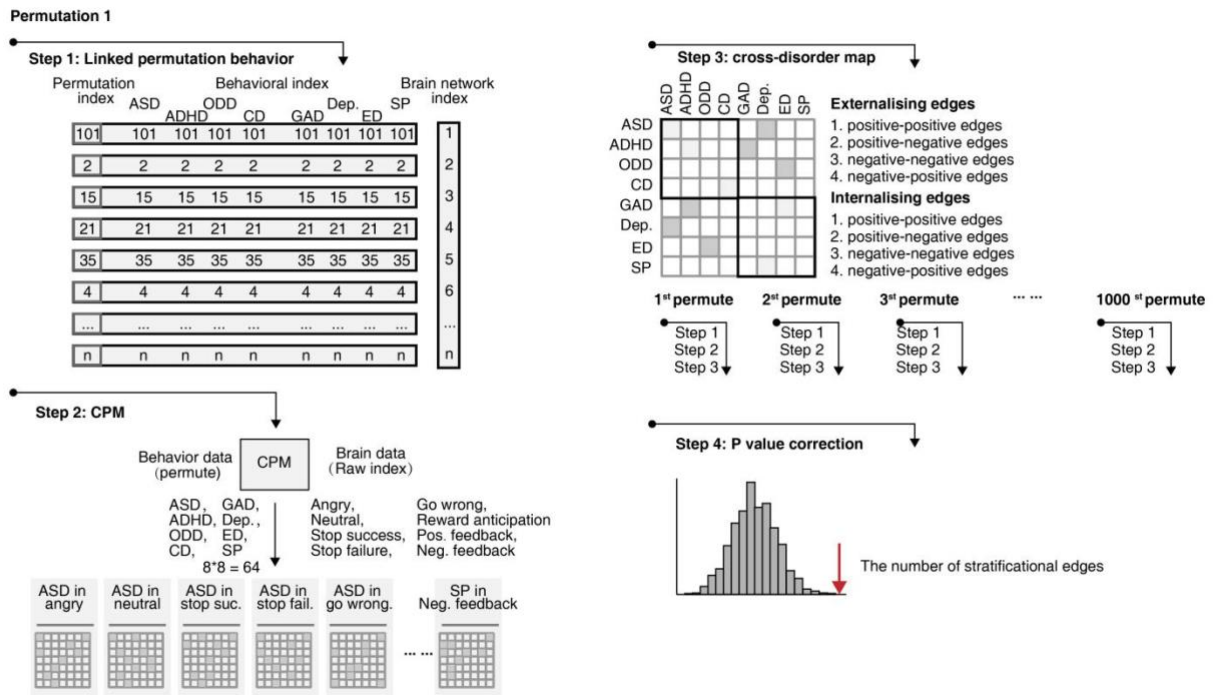
The figure of shen-268 template was adapted from a previous publication<sup>22</sup>.

# Connective predictive model (CPM)



**Figure S4. The analysis flow of the connectome-based predictive model (CPM).**

# Linked Permutation analysis



**Figure S5. The analysis flow of the permutation analysis.**

ADHD, attention deficit/hyperactivity disorder; ASD, autism spectrum disorder; CD, conduct disorder; ODD, oppositional defiant disorder; Pos. Feedback, positive feedback; Neg. Feedback, negative feedback.