STUDY SYNOPSIS

Introduction Summary

Temperament involves stable behavioral and emotional tendencies that differ between individuals, which can be first observed in infancy or early childhood and relate to behavior in many contexts and over many years.1 One of the most rigorously characterized temperament classifications relates to the tendency of individuals to avoid the unfamiliar and to withdraw from unfamiliar people, objects, and unexpected events. This temperament is referred to as behavioral inhibition or inhibited temperament (IT).2 IT is a moderately heritable trait1 that can be measured in multiple species.3 In humans, levels of IT can be quantified from the first year of life through direct behavioral observations or reports by caregivers or teachers. Similar approaches as well as self-report questionnaires on current and/or retrospective levels of IT1 can be used later in life.

Variations in IT are present on a continuous scale within the population, and research suggests that about 20% of young children are characterized by high IT,4 which is in general stable over time.5 Considerable data suggest that this high childhood IT (cIT) has adverse long-term consequences: infants with cIT often become more reserved adults, and, on average, such infants exhibit poorer outcomes than noninhibited infants with respect to social relationships and internalizing psychopathology.6 More specifically, almost half of all children with elevated and stable cIT will develop social anxiety disorder later in life compared with only 12% of noninhibited children.7 Thus, cIT predicts risk for later psychopathology, especially social anxiety disorder.8,9

Several neuroimaging studies have examined neurobiological correlates of cIT. Such research is important, as brain characteristics—including brain structure, function, and connectivity—may mediate the cIT-related risk for poor outcomes.10 Previous studies have linked cIT to the structure and function of brain networks involved in emotion perception, experience, and regulation.1 These brain networks involve the dorsal (caudal) and ventral (rostral) anterior cingulate cortex, insula, amygdala, dorsolateral and medial prefrontal cortex, orbitofrontal cortex, and striatum (cf.1,10), all of which have also been implicated in the familial risk for social anxiety disorder.11 In addition, translational work on anxious temperament has indicated involvement of the hippocampus.3,12 Despite this progress, the few available studies on the neural structural correlates of cIT are often restricted to specific regions of interest, while, to the best of our knowledge, cortical surface area and cortical thickness have been examined in only one study with an exploratory approach.13 Furthermore, most findings with respect to brain structure are unique to a specific sample, and cross-study comparisons are limited by relatively small sample sizes and failure to consider potential modifying variables such as age and biological sex.

In this ENIGMA-Anxiety project,14 we aim to extend previous work by examining brain structure associated with cIT in a large dataset, assembling data acquired at 12 research centers worldwide (17 samples, N = 4,681) (Table 1). Compared with the individual studies, this new study is better powered owing to the larger number of research participants available for analysis. Moreover, by combining data through a mega-analytic approach, the present study facilitates the differentiation of consistent, generalizable findings from false-positive findings that could emerge from studies with smaller samples. Such work has the potential to establish reproducible anatomical correlates and could inform the development of mechanistic studies and intervention research with clinical relevance.15

We expect to corroborate findings in brain circuits found previously (involved in processing fear, reward, and emotion regulation),1,10 with small-to-medium effect sizes. We hypothesize that structural alterations in brain regions involved in these processes, in particular gray matter volumes of multiple subcortical structures (amygdala, hippocampus, striatum including caudate and putamen), and characteristics of several frontal and temporal cortical areas (orbitofrontal cortex, anterior cingulate cortex, insula superior temporal gyrus, transverse gyrus, fusiform gyrus) are neural substrates of cIT.

Method Summary

This ENIGMA-Anxiety Working Group project14 will include individual participant data assembled from studies.
<table>
<thead>
<tr>
<th>Sample (location)</th>
<th>Type of sample</th>
<th>N (n female) with MRI and cIT data</th>
<th>Design*</th>
<th>Age at MRI scan, range (mean ± SD)</th>
<th>Age at cIT phenotype, range (mean ± SD)</th>
<th>Measure of cIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brains study</td>
<td>Oversampled for high/low cIT</td>
<td>130 (72)</td>
<td>C</td>
<td>9.2-13.2 y (10.8 ± 1.0)</td>
<td>9.2-13.2 y (10.8 ± 1.0)</td>
<td>BIQ—parent rated</td>
</tr>
<tr>
<td>Brazilian High Risk Cohort (National Institute of Developmental Psychiatry for Children and Adolescents [INPD], São Paulo, Brazil)</td>
<td>Community sample and high-risk sample of children with increased familial risk for mental disorders</td>
<td>678 (290)</td>
<td>C</td>
<td>5.8-13.0 y (9.7 ± 1.6)</td>
<td>5.8-13.0 y (9.7 ± 1.6)</td>
<td>EATQ-R—shyness scale</td>
</tr>
<tr>
<td>Cohort 3/4 (University of Maryland, College Park, Maryland)</td>
<td>Community sample: prospective longitudinal study of infants thought likely to display behavioral inhibition later in infancy and early childhood</td>
<td>95 (51)</td>
<td>L</td>
<td>13.3-21.1 y (18.0 ± 1.9)</td>
<td>Around 24 mo</td>
<td>(no data at individual level)</td>
</tr>
<tr>
<td>Generation R, sample with behavioral observations (Erasmus University Medical Center, Rotterdam, the Netherlands)</td>
<td>Longitudinal community sample</td>
<td>584 (297)</td>
<td>L</td>
<td>8.7-12.0 y (10.2 ± 0.6)</td>
<td>34.7-44.2 mo (37.4 ± 1.4)</td>
<td>Standard laboratory observations: composite score of stranger, robot, tunnel episodes</td>
</tr>
<tr>
<td>Generation R, sample with questionnaire data (Erasmus University Medical Center, Rotterdam, the Netherlands)</td>
<td>Longitudinal community sample</td>
<td>1,982 (1,030)</td>
<td>L</td>
<td>8.6-12.0 y (10.0 ± 0.5)</td>
<td>4.5-11.8 mo (6.7 ± 1.1)</td>
<td>IBQ-r—fear subscale</td>
</tr>
<tr>
<td>Maryland-PAX (University of Maryland, College Park, Maryland)</td>
<td>30-mo longitudinal study of a sample of first-year university students enriched for internalizing risk</td>
<td>220 (109)</td>
<td>C</td>
<td>18-19 y (18.3 ± 0.4)</td>
<td>Retrospective: remembered inhibited behaviors in childhood</td>
<td>RMBI</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Sample (location)</th>
<th>Type of sample</th>
<th>N (n female) with MRI and cIT data</th>
<th>Design[^a]</th>
<th>Age at MRI scan, range (mean ± SD)</th>
<th>Age at cIT phenotype, range (mean ± SD)</th>
<th>Measure of cIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maryland-TAX</td>
<td>Cross-sectional community sample</td>
<td>53 (28)</td>
<td>C</td>
<td>13-17 y (15.0 ± 1.2) Retrospective: remembered inhibited behaviors in childhood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nijmegen Longitudinal Study</td>
<td>Longitudinal community sample</td>
<td>71 (31)</td>
<td>L</td>
<td>17 y 1.20-1.28 y (1.24 ± 0.02) Standard laboratory observations at age 15 mo; stranger and robot episodes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pittsburgh (University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania)</td>
<td>High- and low-risk (control) children/adolescents from ongoing family studies</td>
<td>15 (3)</td>
<td>L</td>
<td>19.2-24.8 y (21.5 ± 1.7) 4.1-6.4 y (5.1 ± 0.7) Laboratory observations during peer play</td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Raffaele (Vita-Salute San Raffaele University and San Raffaele Scientific Institute, Milan, Italy)</td>
<td>Community sample</td>
<td>20 (8)</td>
<td>L</td>
<td>13-16 y (14.8 ± 1.1) 8-10 y (9.1 ± 0.7) Empirical composite index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDAN (NIMH, Bethesda, Maryland)</td>
<td>Treatment-seeking children and control group of healthy volunteers</td>
<td>55 (26)</td>
<td>C</td>
<td>7.3-14.6 y (10.3 ± 1.7) 8.0-12.8 y (10.4 ± 1.5) BIQ—child rated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stony Brook Temperament Study (Stony Brook University, Stony Brook, New York)</td>
<td>Community sample; MRI subsample oversampled for youth with temperamental high negative emotionality, low positive emotionality, and high behavioral inhibition at age 3</td>
<td>74 (31)</td>
<td>L</td>
<td>9-12 y (10.2 ± 0.9) 2.9-4.0 y (3.4 ± 0.3) Lab-TAB: 3 Kagan-like tasks around age 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Sample (location)</th>
<th>Type of sample</th>
<th>N (n female) with MRI and cIT data</th>
<th>Design</th>
<th>Age at MRI scan, range (mean ± SD)</th>
<th>Age at cIT phenotype, range (mean ± SD)</th>
<th>Measure of cIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTS (University of Maryland, College Park, Maryland)</td>
<td>Longitudinally followed sample of children selected at age 4 mo based on their behavior in the laboratory</td>
<td>96 (56)</td>
<td>L</td>
<td>9.1-19.5 y (11.4 ± 2.1)</td>
<td>1.9-2.7 y (2.1 ± 0.2)</td>
<td>Standard laboratory observations (composite score of stranger, robot, tunnel episodes)</td>
</tr>
<tr>
<td>Vanderbilt—children (Vanderbilt University Medical Center, Nashville, Tennessee)</td>
<td>Study with extreme discordant phenotypes approach: inhibited and uninhibited children at the extreme ends</td>
<td>55 (33)</td>
<td>C</td>
<td>8-12 y (9.3 ± 1.1)</td>
<td>8-12 y (9.3 ± 1.1)</td>
<td>BIQ—child rated</td>
</tr>
<tr>
<td>Vanderbilt—young adults (Vanderbilt University Medical Center, Nashville, Tennessee)</td>
<td>Study with extreme discordant phenotypes approach: inhibited and uninhibited young adults at the extreme ends</td>
<td>150 (83)</td>
<td>C</td>
<td>18-25 y (21.8 ± 2.0)</td>
<td>Retrospective: remembered inhibited behaviors in childhood</td>
<td>RSRI</td>
</tr>
<tr>
<td>Western University (The Brain and Mind Institute, Western University, London, Ontario, Canada)</td>
<td>Children selected based on presence/absence maternal depression</td>
<td>87 (38)</td>
<td>L</td>
<td>9.2-12.4 y (11.1 ± 0.7)</td>
<td>3.0-4.0 y (3.4 ± 0.3)</td>
<td>Lab-TAB: risk room, stranger approach, and jumping spider</td>
</tr>
<tr>
<td>Wisconsin Twin Project—RDoC twin study (University of Wisconsin—Madison, Madison, Wisconsin)</td>
<td>Longitudinally followed samples of twins, recruited from statewide birth records for birth cohorts 1989-2004</td>
<td>316 (145)</td>
<td>L</td>
<td>15.1-23.9 y (17.5 ± 1.6)</td>
<td>6.5-9.0 y (7.5 ± 0.5)</td>
<td>Ratings on approach and shyness from 3-h home visit and scores from videotaped reactions to “Conversation With a Stranger” episode of Lab-TAB</td>
</tr>
<tr>
<td>Total N</td>
<td>4,681 (2,331)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: BIQ = Behavioral Inhibition Questionnaire; cIT = childhood inhibited temperament; EATQ-R = Revised Early Adolescent Temperament Questionnaire; IBQ-r = Infant Behavior Questionnaire - revised; Lab-TAB = Laboratory Temperament Assessment Battery; MRI = magnetic resonance imaging; NIMH = National Institute of Mental Health; PAX = prospective anxiety; RDoC = Research Domain Criteria; RMBI = Retrospective Measure of Behavioural Inhibition; RSRI = Retrospective Self-Report of Inhibition; SDAN = Section on Development and Affective Neuroscience; TAX = teen anxiety; TOTS = Temperament Over Time Study.

*With respect to time point temperament assessment and MRI scan for data used in this study: C = cross-sectional; L = longitudinal.*
in which participants underwent magnetic resonance imaging scanning (T1-weighted anatomical magnetic resonance imaging scans) between 6 and 25 years of age. Regardless of age at the time of scanning, all participants will be phenotyped for cIT (defined as age/C2012 years). These temperament assessments could be behavioral observations in childhood, parental reports, or self-report questionnaires on current or retrospective temperament.

We will perform a mega-analysis with a whole-brain approach (regional and vertex-wise; familywise error rate-corrected)\(^{16}\) and investigate the relation between cIT (continuous) and 3 distinct neuroanatomical metrics (determined using FreeSurfer software [https://surfer.nmr.mgh.harvard.edu/]), namely, volumes of subcortical structures, cortical thickness, and cortical surface area. Additionally, analyses will be performed in 3 subsets, based on the method and thus age at which cIT was determined: first (early-life) behavioral observations, second parental/teacher reports during childhood, and third self-report measures acquired during late childhood/adolescence. A fourth sensitivity analysis will exclude samples with retrospective measures of cIT.

**Significance Summary**

This initiative is the first mega-analysis of brain structure associated with the temperamental risk for developing internalizing psychopathology. This provides the possibility of detecting novel cIT-related brain alterations and clarifying inconsistent findings of prior work.\(^{10}\) Mega-analyses combine existing datasets to increase the overall sample size. This is particularly valuable for data acquired in vulnerable participants, who are often difficult to recruit. Such studies exemplify next-generation science: previous studies within the ENIGMA Consortium have resulted in important insights in the neurobiology of psychiatric conditions.\(^{17}\) These discoveries reflect the advantages of large-scale data analyses for testing the reproducibility and robustness of neuroimaging findings.\(^{17}\) We expect the current project to provide similar insights, increasing our understanding of the development of psychopathology in youth at risk. In addition, by preregistering the study in advance of performing the analyses, we hope to contribute to a reduction of the potential publication bias in the field and to advance a more complete scientific record on this topic (cf.\(^{18}\)).

Janna Marie Bas-Hoogendam, PhD
Rachel Bernstein, BA
Brenda E. Benson, PhD
Kristin A. Buss, PhD
Kelley E. Gunther, PhD
Koraly Pérez-Edgar, PhD

Giovanni A. Salum, PhD
Andrea P. Jackowski, PhD
Rodrigo A. Bressan, PhD
André Zugman, MD, PhD
Kathryn A. Degnan, PhD
Courtney A. Filippi, PhD
Nathan A. Fox, PhD
Heather A. Henderson, PhD
Alva Tang, PhD
Selin Zeytinoğlu, PhD
Anita Harrewijn, PhD
Manon H.J. Hillegers, MD, PhD
Tonya White, MD, PhD
Marinus H. van Ijzendoorn, PhD
Carl E. Schwartz, MD, PhD
Julia M. Felicione, MSc
Kathryn A. DeYoung, MA, MS
Alexander J. Shackman, PhD
Jason F. Smith, PhD
Rachael M. Tillman, PhD
Yvonne H.M. van den Berg, PhD
Antonius H.N. Cillesen, PhD
Karim Roelofs, PhD
Anna Tyborowska, PhD
Shirley Y. Hill, PhD
Marco Battaglia, MD
Marco Tettamanti, PhD
Lea R. Dougherty, PhD
Jingwen Jir, PhD
Daniel N. Klein, PhD
Hoi-Chung Leung, PhD
Suzanne N. Avery, PhD
Jennifer Urbano Blackford, PhD
Jacqueline A. Clauss, MD, PhD
Elizabeth P. Hayden, PhD
Pan Liu, PhD
Matthew R.J. Vandermeer, PhD
H. Hill Goldsmith, PhD
Elizabeth M. Planalp, PhD
Thomas E. Nichols, PhD
Paul M. Thompson, PhD
P. Michiel Westenberg, PhD
Nic J.A. van der Wee, MD, PhD
Ynke A. Groenewold, PhD
Dan J. Stein, MD, PhD
Andersen M. Wickler, MD, DPhip
Daniel S. Pine, MD, PhD

on behalf of the ENIGMA-Anxiety Working Group

Accepted June 6, 2022.

This article was reviewed under and accepted by ad hoc editor Guido K.W. Frank, MD.

Drs. Bas-Hoogendam and Westenberg are with Leiden University, Leiden, the Netherlands. Drs. Bas-Hoogendam and van der Wee are with Leiden University Medical Center, Leiden, the Netherlands. Drs. Bas-Hoogendam, Westenberg, and van der Wee are with Leiden Institute for Brain and Cognition, Leiden, the Netherlands. Drs. Bas-Hoogendam, Benson, Zugman, Filippi, Wickler, and Pine, and Ms. Bernstein are with the National Institute of Mental Health, Bethesda, Maryland. Drs. Buss and Pérez-Edgar and Ms. Gunther are with Pennsylvania State University, University Park. Dr. Salum is with Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
Project administration: Bas-Hoogendam, Bernstein, Benson, Groenewold, Winkler, Resources: Westenberg, Winkler, Pine Sources: Winkler

Supervision: Westenberg, Pine

Visualization: Bas-Hoogendam, Winkler

Writing — original draft: Bas-Hoogendam

Writing — review and editing: Bernstein, Benson, Buss, Gunther, Pérez-Edgar, Salazar, Filippi, Wang, Schwartz, Cillessen, Dougerty, Jin, Klein, Leung, Avery, Blackford, Clauss, Hayden, Liu, Vandermeer, Goldsmith, Planalp, Nichols, Thompson, Westenberg, van der Wee, Groenewold, Stein, Winkler, Pine

The ENIGMA-Anxiety Working Group: https://enigma.ini.usc.edu/ongoing/enigma-anxiety/

Disclosure: Dr. Pérez-Edgar has received funding from the National Institute on Drug Abuse (U01DA055361-01) and NIMH (R56MH162349, R01MH109692, and R01MH130007); royalties from Springer Press; and honoraria for lectures to professional audiences and for editing Developmental Psychology. Dr. Salam has reported that the Brazilian High-Risk Cohort for Mental Conditions was supported with grants from the National Institute of Developmental Psychiatry for Children and Adolescents (INPD), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; 465550/2014-2), and the São Paulo Research Foundation (2014/09197-0). Dr. Jackowski has received grant support from CNPq 427215/2018-8. Dr. Degnan has received grant support from a subaward from NIMH (U01MH093349) via the University of Maryland and honoraria for associate editor duties with Developmental Psychology. Dr. Filippi has received a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation (#28024) and the Intramural Research Program of NIMH through project ZIA-MH-002782. Dr. Fox has received grant support from NIMH (U01MH093349), NICHAD, the National Science Foundation (NSF), and the National Institutes of Health Environmental influences on Child Health Outcomes (NIH ECHO) consortium, the Russell Sage Foundation, and the Lumos Foundation; royalties from Springer Press, Guilford Press, and Harvard University Press; and honoraria for lectures to professional audiences. Dr. Henderson has received grant support from the Social Sciences and Humanities Research Council of Canada and the Canada Foundation for Innovation. Dr. White has received grant or research support from the Sophia Children’s Hospital Foundation. She has served on the advisory board/Data Safety and Monitoring Board of the University of Bergen Center for Brain Plasticity. She has served on the editorial board of Neuroinformatics and as Editor-in-Chief of Aperture Neuro. Drs. Roelofs and Tyborowska have reported that past magnetic resonance imaging scans of the Nijmegen Longitudinal Study (NLS) was funded by European Research Council starting grant (ERC_STG2012_313749 awarded to K.R.) and a F7-HEALTH-2013-INNOVATION grant (602805-2). Dr. Hill has received grant support from the National Institute on Alcohol Abuse and Alcoholism (AA021746 and AA021746-S1-Supplement). Dr. Thompson has received partial research support from Biogen, Inc., unrelated to this work, and NIH Big Data to Knowledge (BD2K) award (U54 EB020403). Dr. van der Wee has received grant support from the EU Innovative Medicines Initiative Psychiatric Ratings using Intermediate Stratified Markers 2 (IMI PRISM 2) with grant agreement number 10103377. Dr. Stein has received research grants and/or consultancy honoraria from Discovery Vitality, Johnson & Johnson, Kanna, Lundbeck, Orion, Sanofi, Servier, and Takeda. Ms. Gunther has received an NSF Graduate Research Fellowship (DGE1255832). Drs. Bas-Hoogendam, Benson, Buss, Buzan, Gunther, Tetyama, Dougerty, Jin, Klein, Leung, Avery, Blackford, Clauss, Hayden, Liu, Vandermeer, Goldsmith, Planalp, Nichols, Westenberg, Groenewold, Winkler, and Pine and Ms. Bernstein, Felicione, and DeYoung have reported no biomedical financial interests or potential conflicts of interest.

Correspondence to Janna Marie Bas-Hoogendam, PhD, Institute of Psychology, Leiden University, Pieter de la Court Building, room 3.B47, Wassemastraat 32 2333 AK Leiden, The Netherlands, e-mail: j.m.bas-hoogendam@sw.leidenuniv.nl

www.jaacap.org

Journal of the American Academy of Child & Adolescent Psychiatry

Volume  |  Number  |  September 2022

BAS-HOOGENDAM et al.
REFERENCES


Registered Report

Structural brain correlates of childhood inhibited temperament: an ENIGMA-Anxiety mega-analysis

Supplemental Materials Stage 1 Manuscript

Bas-Hoogendam et al.
Supplemental Methods

Questionnaires on psychopathology

All research sites were asked to provide as much questionnaire data on psychopathology as possible – cf. the methods described in Bas-Hoogendam et al.¹. This concerned the following questionnaires with respect to anxiety disorders: the Hamilton Anxiety Rating Scale ², Penn State Worry Questionnaire ³, Generalized Anxiety Disorder 7-Item questionnaire ⁴, State Trait Anxiety Inventory ⁵, Anxiety Sensitivity Index ⁶, Beck Anxiety Inventory ⁷, Liebowitz Social Anxiety Scale ⁸, Panic and Agoraphobia Scale ⁹, Agoraphobic Cognition Questionnaire ¹⁰, Panic Disorder Severity Scale ¹¹ and the Screen Child Anxiety Related Disorders ¹².

Furthermore, we asked for data on the Beck Depression Inventory II ¹³, the Children’s Depression Inventory ¹⁴, and Hamilton Depression Rating Scale ¹⁵.

Not all research sites had available data on these variables, therefore, these data were not included in the analyses. Descriptive information and scores on these questionnaires for each sample are provided in Supplemental Table 3.
Information for each sample: diagnostic interviews, clinical questionnaires and assessment of temperament

Brains sample

Brazilian High Risk Cohort (BHRC)

Cohort 3/4

Generation R

Maryland – PAX sample

Maryland – TAX sample

Nijmegen Longitudinal Study on Child and Infant Development

Pittsburgh

San Raffaele

SDAN

Stony Brook Temperament Study

TOTS

Vanderbilt - children

Vanderbilt – young adults

Western University

Wisconsin Twin Project
Brains sample

Publications: 16–25

In- and exclusion criteria 17
The sample consisted of 9–12-year-olds, of families which were recruited through a university database of families interested in participating in research studies, community outreach, and word-of-mouth. The study was part of a larger study on temperament, attention, and anxiety. Participants were screened using parental report on the Behavioral Inhibition Questionnaire (BIQ) 26. Children who met BI cut-off scores (>120 in BIQ Total score or >60 in BIQ Social novelty; ~25% of children screened) were identified and oversampled, while children below cut-off were recruited as a gender- and age-matched non-BI comparison group. Cut-off scores were based on previous studies of extreme temperament in children 27. Exclusionary criteria included severe psychiatric diagnosis (e.g. bipolar disorder), IQ below 70, or severe medical illness. Parents and children provided written consent/assent and the Institutional Review Board approved this study.

Diagnostic interview 16, 17
Social anxiety symptoms and major depressive disorder were assessed via parent-report on the computerized Diagnostic Interview Schedule for Children version 4 (C-DISC 4; 28). A trained research assistant conducted the semi-structured interview, in which parents judged DSM-IV symptoms as either present (‘yes’) or absent (‘no’).

Anxiety symptoms were measured using the parent-report version the Screen for Child Anxiety Related Emotional Disorders (SCARED) 12, a 41-item instrument assessing symptoms of panic disorder, generalized anxiety, separation anxiety, social phobia, and school phobia defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Parents rated the
frequency with which their children experience each symptom on three-point scales (0 = “almost never”, 1 = “sometimes”, and 2 = “often”). Sub-scale scores were summed to create the total score. The SCARED has satisfactory psychometric properties in both clinical and community samples and it offers a valuable tool to predict specific anxiety disorders in clinically-referred youths. It had good internal consistency in the present study (Cronbach’s alpha = .90).

**Measures of temperament**

Behavioral inhibition (BI) was assessed using the Behavioral Inhibition Questionnaire (BIQ), a 30-item instrument that measures the frequency of BI-linked behavior in the domains of social and situational novelty (plus a summed total score) on a seven-point scale ranging from 1 (“hardly ever”) to 7 (“almost always”). Four questions were edited to be more appropriate for the target age range in the current study (e.g., reference to preschool, kindergarten, and childcare was removed for the question: “Happily separates from parent(s) when left in new situations for the first time (e.g., kindergarten, preschool, childcare”). The questionnaire has adequate internal consistency, construct validity, and validity in differentiating behaviorally inhibited from non-inhibited children, parent reports on the BIQ correlate with laboratory observations of BI in social contexts, and the BIQ had good internal consistency in the present study (Cronbach’s alpha = .91).

In the present study, we will use the total score on the BIQ as indicator of cIT.
Brazilian High Risk Cohort (BHRC)

Publications: 32,33

In- and exclusion criteria 32,34,35
As summarized in Axelrud and colleagues 34, “the screening stage for the BHRC took place in 2009/2010 in public schools from two cities in Brazil (Porto Alegre and São Paulo), including a total of 9937 children. Eligibility criteria were being 6–12 years old at enrolment and being registered by a biological parent who could provide information about the children’s behavior. From the screening sample, 2511 children were selected and evaluated at baseline in 2010/2011. Among this subsample, 1554 children were at risk of mental illness, established using family history and the current presence of symptoms, and 957 were randomly selected. A subset of 741 participants underwent MRI scans and 726 of these also underwent psycho-pathology assessment. Parents of the participants and participants provided written or verbal consent. The Ethics Committee of the University of São Paulo approved the study.”
For the present study, we selected participants aged < 13 years at baseline, with imaging data and EATQ-R data.

Diagnostic interview
As described in Hoffmann et al.35, “mental disorders were assessed using the Brazilian Portuguese version 36 of the DAWBA 37. This structured interview was administered to biological parents by trained lay interviewers. Responses, as well as structured answers, were then evaluated by a total of nine certified child psychiatrists, which confirmed, refuted or altered the initial computerized diagnosis. All of them were trained and supervised jointly by a senior child psychiatrist with extensive experience in rating the DAWBA. To perform reliability analysis of the rating procedure, a sub-sample of 200 subjects received a second rating by a
trained child psychiatrist. We selected subjects divided equally into DAWBA bands. DAWBA bands represent computer-generated categories based on answers to the DAWBA questions that provide information to the rater concerning the probability of a positive diagnosis (< 0.1%, ~3%, ~15%, ~50% and higher than 70%). The second rater was informed that the 200 cases (40 cases from each band) did not represent the population distribution of mental disorders. Inter-rater agreement was above 90% for all diagnosis and kappa values ranging from 0.72 for hyperkinetic disorders and 0.84 for emotional disorders. Diagnoses are related to diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders, 4th edition.”

**Measure of temperament**

Temperament was assessed with the Brazilian-Portuguese self-report version of the early adolescent temperament questionnaire (EATQ-R). This questionnaire is a 65-items Likert scale, ranging from 1 (always false) to 5 (always true), containing 12 subscales (4–7 items each). Five temperament factors were used, namely effortful control, fear, frustration, shyness and surgency.

In the present study, we will use the sum score of the shyness items of the EATQ-R as an index of cIT.
Cohort 3/4

Publications: 42–44

In- and exclusion criteria
As described by Shechner et al.43, “participants were a subsample of individuals who were selected at 4 months of age and assessed for BI at ages 14 months and 24 months, and for social reticence at 4 and 7 years of age – cf. Fox and colleagues 42. At each time point, parental ratings of shyness were also collected. Individuals taking psychotropic medications, reporting acute psychopathology in need of immediate treatment, taking recreational drugs, or having any contraindications to MRI (e.g., permanent retainer) were excluded from the current study. All other individuals from the longitudinal study were asked to participate if they were physically healthy based on medical examination and history and had an IQ of > 70.”

Diagnostic interview
The presence of current or lifetime psychiatric disorder was assessed by the Structured Clinical Interview for DSM IV (SCID) 45. Anxiety levels were assessed in two ways: (1) the Beck Anxiety Inventory (BAI) 7 was used as a measure of trait anxiety, and (2) the State subscale of the State Trait Anxiety Inventory (STAI-S) 5 was used as a measure of state anxiety at the time of the scan.

Measures of temperament
Inhibited behavior to novel stimuli was coded at 14 and 24 months 42. Behavioral scores were standardized at each time point. For the present study, we will use the ‘BI-classic-24’ index, composed of scores on the stranger, robot and tunnel episodes of the Laboratory Temperament
Assessment Battery – preschool version 46. Mothers also reported their child’s social fear at 14 and 24 months using the Toddler Behavior Assessment Questionnaire 47.
Generation R

Publications: 48–52

Study design; in- and exclusion criteria

For this project, we will use the data from the second neuroimaging wave of Generation R (children age 9 – 11) 50, because this imaging wave consisted of more children and was more representative of the overall Generation R sample when compared to the participants in the first imaging wave 49. As described in White et al (2018) 50, “the children who were recruited were participants of the Generation R Study, which is a population-based longitudinal cohort study of child health and development based in Rotterdam, the Netherlands. An overview of the Generation R study design and population has been described in detail in 53. In brief, all pregnant women who were living within a well-defined region in Rotterdam (defined by postal codes) with a delivery data between April 2002 and January 2006 were invited to participate. A total of 9,778 mothers provided informed consent and were recruited. Rotterdam is ethnically diverse, with approximately 44% of the population being non-Dutch. Recruitment into Generation R reflects this diversity. Of the 9,778 mothers, 58% were Dutch, 9% Surinamese, 9% Turkish, 7% Moroccan, 3% Dutch Antillean, and 3% of Cape Verdian descent 53. Additional detailed measurements of fetal and postnatal growth and development have been conducted in a randomly selected subgroup of Dutch children (n = 1,232; known as the ‘Focus Cohort’) and their parents at 32 weeks gestational age and at the postnatal ages of 1.5, 6, 14, 24, 36 and 48 months. These additional evaluations on this subgroup were conducted in a Generation R dedicated research center. From the age of 5 years onwards, all willing children and their parents with the Generation R Study have had regular visits to a dedicated research center that includes advanced imaging facilities. The second wave of neuroimaging started in March 2013 with a total of 4,245 children visiting the MRI Centre and 4,087 children received
a brain MRI scan, of which 3,992 fulfilled the Dutch laws of parental consent for research and of these 3,959 children completed a complete T1-weighted sequence.”

**Diagnostic interview**

As outlined by White and colleagues 50, the Diagnostic Interview Schedule for Children-young child version (DISC-YC) was administered in subsample of the Generation R Study that was enriched for psychopathology 54, and this interview took place between the ages of 5-8 years. The DISC-YC is a highly structured DSM-IV-based interview administered to caregivers of children aged 3–8 years. Six trained interviewers (including bilingual interviewers) administered the computer-assisted DISC-YC that determines the presence of disorders for a timeframe of 3 months, or 1-year for dysthymia and conduct disorder, by applying algorithms provided by the developer.

**Measures of temperament**

*Age 6 months (description from Jansen and colleagues, 2009 52)* “At the age of six months, infant temperament was assessed using an adapted version of the infant behavior questionnaire-revised (IBQ-R) 55. A detailed description of the changes has previously been described 56. Briefly, we assessed six scales of the IBQ-R: Activity Level (e.g. movements of arms and legs); Distress to Limitations (e.g. fussing or crying while in caretaking activities); Duration of Orienting (e.g. attention to a single object for extended periods of time); Sadness (e.g. general low mood); Fear (e.g. startle or distress to novelty or sudden changes in stimulation); and Recovery from Distress (e.g. rate of recovery from general arousal; ease of falling asleep). Internal consistencies for the adapted IBQ-R ranged from 0.70 for Duration of Orienting to 0.85 for Fear, which is comparable to the internal consistencies of the original IBQ-R 55.”
Age 36 months A subsample of the Generation R sample (‘Focus Cohort’; all of Dutch origin) visited the lab where stranger fear and the response to a jumping spider were assessed using the Laboratory Temperament Assessment Battery (Lab-TAB).

In the present mega-analysis, we will use the scores on the Lab-TAB as index of cIT (sample with behavioral observations). For participants without these scores, we will use the IBQ-scores on the Fear subscale (sample with questionnaire data).
In- and exclusion criteria
This prospective-longitudinal study focused on the emergence of anxiety disorders and depression and is described in more detail in Hur et al. (2020) 58 and in press. Participants were first-year university students recruited from the University of Maryland. All subjects had normal or corrected-to-normal color vision; and reported the absence of lifetime neurologic symptoms, pervasive developmental disorder, very premature birth, medical conditions that would contraindicate MRI, and prior experience with noxious electrical stimulation. All subjects were free from a lifetime history of psychotic and bipolar disorders; a current diagnosis of a mood, anxiety, or trauma disorder (past 2 months), excepting subclinical (‘other specified’) diagnoses; severe substance abuse; active suicidality; and ongoing psychiatric treatment as determined by an experienced masters-level diagnostician using the Structured Clinical Interview for DSM-5 61.

Diagnostic interview: Structured Clinical Interview for DSM-5 (SCID-5-RV) 61.

Measures of temperament
Participants completed the Adult Measure of Behavioral Inhibition (AMBI) and the Retrospective Measure of Behavioral Inhibition (RMBI) 62. The AMBI is a 16-item clinical research instrument developed in order to measure subjective reports of contemporaneous ‘trait’ inhibition. This instrument provides a dimensional quantitative measurement of the temperamental tendency to respond to social novelty and risk stimuli, with inhibition and
avoidance. The Retrospective Measure of Behavioral Inhibition (RMBI) is an 18-item instrument for the retrospective adult reporting of shyness, reticence, and behavioral inhibition during childhood and early adolescence. This instrument was designed to capture the principal behavioral indices of “behavioral inhibition to the unfamiliar” as measured and observed in children when assessed in play-laboratory settings. This instrument is also a dimensional/quantitative measure. Higher scores on both measures indicate a greater degree of inhibition. Within the PAX-sample, scores on the AMBI and RMBI (total scores) were significantly correlated ($r = 0.47, p < 0.001$).

For the current mega-analysis, we will use the total score for the RMBI as the cIT index.
In- and exclusion criteria

Eighty-four participants between the ages of 13 and 17 years and their caregivers were recruited from advertisements distributed online (i.e., Facebook, listservs), flyers posted at community mental health clinics and broader community settings (i.e., coffee shops, local community centers), and referrals from other university research studies recruiting adolescents. Advertisements were designed to differentially target adolescents with high social anxiety using language inviting “shy” or “socially anxious” adolescents to participate in a study about brain function. Advertisements designed for adolescents without high levels of social anxiety used general language to invite participants to enroll, such as “Are you a teen?” or “Have a teen aged 13-17?”, and these advertisements were only distributed in general community settings.

To ensure inclusion of a clinically enriched sample that comprised both adolescents with social anxiety disorder and adolescents with low levels of social anxiety, participants completed a preliminary screening questionnaire online. The preliminary screening included a measure of the frequency of social anxiety disorder symptoms (the abbreviated Social Phobia and Anxiety Scale for Children; SPAIC-1163) and three additional questions designed to assess interference and distress from social anxiety symptoms using a 5-point Likert scale (1 = Not at all, 5 = Extremely). Individuals were invited to enroll if they met any of the following preliminary inclusion criteria: 1) obtaining a score of 16 or above on the SPAIC-1163; 2) indicating social anxiety interference or distress on the online screener prior to enrollment; and 3) obtaining a
score of 6 or below on the abbreviated SPAIC-11\textsuperscript{63} and indicating low social anxiety interference and distress on the online screener prior to enrollment.

Participants in both groups were right-handed native English speakers with no history of head injury, neurological disorders, psychosis disorders, pervasive developmental disorders (e.g., autism) and bipolar disorder. Participants were free from MRI contraindications and were not currently using any psychotropic medications.

**Diagnostic interview**

The Mini-International Neuropsychiatric Interview For Children And Adolescents (MINI-KID\textsuperscript{64}) based on DSM-IV and ICD-10 criteria.

**Measures of temperament**

Within the TAX sample, five measures of inhibited temperament were acquired. Three were self-reports: the Current Self-Report of Inhibition (CSRI), the Retrospective Self-report of Inhibition (RSRI; focused on elementary and early middle school)\textsuperscript{65} and the BIQ-A\textsuperscript{26}. In addition, a caregiver completed the CSRI and RSRI for their child. All five measures were highly correlated (all within-subject correlations $p < 0.001$, $r > 0.55$).

For the current mega-analysis, we will use the total score for the adolescent-reported RSRI as the cIT index.
In- and exclusion criteria at age 15 months

(As described in Van Bakel et al.68): “The sample consisted of 129 physically healthy 15-month-old infants (67 boys, 62 girls) and their primary caregivers. Because earlier research has recommended that studies of the possible determinants of parenting be conducted in heterogeneous samples 69, the aim was to recruit such a sample in the present study. The recruitment of the families was based on the records from local health-care centers in the city of Nijmegen in The Netherlands. During 9 consecutive months, all families with a 15-month-old baby (i.e., 639 families) living in districts with many young families from various socio-economic backgrounds were contacted. They were sent a recruitment letter explaining the goals of the study and were asked to return a card if interested in participating. Of the 174 families who replied, 129 parent–child dyads (the maximum attainable given the time and resources available for the project) were randomly selected for the study. The sample included 123 two-parent families and 6 single-parent families. In 3 families, the father was the primary caregiver of the child. In these cases, the mothers were the breadwinners and had full-time jobs out of the home. Because these fathers had taken care of the infants from birth on and acted as their primary attachment figures, they were included in the sample of primary caregivers. The patterns of scores of these 3 fathers, moreover, turned out to fall within the normal range in the sample. The percentages of single parents and of fathers acting as primary caregivers were representative of families in The Netherlands with children in this age group. In the sample, 38% of the primary caregivers were homemakers, and only 4% worked out of the home for more than 32 hours a week. The ages of the primary caregivers ranged from 22 to 47 years (M:
32.9 years, SD: 4.42). Their level of education ranged from low (elementary school) to high (college degree or more). The sample contained 73 firstborn infants and 56 later-born infants.”

**Inclusion MRI session – from Tyborowska et al.**

“All actively participating children from the Nijmegen Longitudinal Study on Child and Infant Development (n = 116) were approached to take part in this imaging study. Anatomical scans were obtained from participants at 14 and 17 years of age. Forty-nine at the first imaging time-point and ninety-six at the second imaging time-point agreed to participate. Participants did not have a history of psychiatric disorders or neurological illness (as indicated by parent/guardian report). Written informed consent was obtained from parents and participants during each measurement wave. The study was approved by the local ethics committee (CMO region Arnhem – Nijmegen) and was conducted in compliance with these guidelines.”

**Psychopathology at age MRI scan**

As described in Tyborowska et al. 66, “internalizing symptoms at age 17 were measured using the Child Behaviour Checklist (CBCL) 70. The CBCL is a parent-report questionnaire used to assess the frequency of emotional and behavioral problems exhibited by the adolescent in the past six months. The parent rated each behavior or symptom on a three-point Likert scale (not true, somewhat or sometimes true, very true or often true). Items from the scales anxious/depressive, withdrawn/depressive, and somatic complaints were summed to provide a score for internalizing symptoms.”

No diagnostic interview was performed, precluding establishing diagnoses according to the DSM-5.
Measures of temperament

Procedure 68 “The caregivers and children were visited in their homes for 2 hours when the child was 15 months of age. During the visit, the primary caregiver completed a Q-sort and a set of questionnaires assessing his or her ego-resiliency and attachment style, network and partner support, and child temperament. In addition, the caregiver was administered a verbal intelligence test. At the end of the visit, the caregiver and child were videotaped during the performance of four consecutive interaction tasks, lasting 3 or 4 min each. The parent was asked to have the child unlock a puzzle box, put a puppet together, do a jigsaw puzzle, and “read” a set of picture books. The parents were also told that they could help the child whenever they felt the need to.

Questionnaire data on temperament The Toddler Behavior Assessment Questionnaire (TBAQ; 71) was used to characterize children in terms of five dimensions of temperament: activity level, pleasure, social fearfulness, anger proneness, and interest/persistence. The caregiver indicates along a 7-point scale how often he or she observed particular behaviors on the part of the child during the past month; for example, “When your child was being approached by an unfamiliar adult while shopping or out walking, how often did your child show distress or cry?” The internal consistency of the five scales was satisfactory; Cronbach’s .86 (20 items) for activity level, .82 (19 items) for pleasure, .77 (19 items) for social fearfulness, .88 (28 items) for anger proneness, and .79 (22 items) for interest/persistence.

Behavioral observations of temperament from 67: “Within 1 week of the home visit, the parent and child visited the University laboratory. First, the child’s cognitive development was assessed. This assessment was followed by a 25-min parent–child interaction episode that was not used in the present study. After that, the 14-min ‘stranger/robot episode’ was set up to measure cortisol reactivity in the infants. This episode was an adapted version of a procedure described by Mullen, Snidman, and Kagan (1993) 72 i.e., 3 min of free play, an encounter with
a female stranger (4 min) “stranger episode”, a confrontation with a moving robot (4 min) “robot episode”, and again 3 min of free play. At the beginning of the session, the child was placed at the center of the room with a set of age-appropriate toys while the parent was sitting on a chair at a distance of about 1m. The parent was given a questionnaire to fill out and was instructed to respond to the child naturally, but to refrain from initiating interaction. The child played freely for 3 min. Next, an unfamiliar woman entered the room with a toy ladybird containing colorful blocks. She sat quietly for 1 min within .3 m from the child. Then she played with the ladybird and the blocks and invited the child to play with the toy (3 min). Next, she went to a cabinet in the corner of the room and placed a colorful mechanical robot (10 in. high) on the floor in front of the cabinet. The experimenter, who was sitting behind the cabinet, turned the robot’s light and frightening sounds on and off and moved the robot forwards and backwards using remote control. The unfamiliar woman invited the child to come and play with the robot (4 min). After this episode, the robot was put away and a new set of age-appropriate toys was spread out on the floor. The stranger left the room, and the child was free to play for another 3 min. The entire lab session was recorded on videotape.

To assess additional behavioral measures of infant fearfulness during the stranger/robot episode, the occurrence of three infant behaviors (adapted from Nachmias et al. 73) was rated from the videotapes, separately for the 4-min stranger episode and the subsequent 4-min robot episode. The behaviors were looking/referencing to parent (i.e., looking between parent and stranger with a questioning expression), proximity seeking/maintaining physical contact with parent (i.e., increasing or actively maintaining proximity to parent within one arm’s length), and crying. Each behavior was rated on a scale of 1 (not at all), 2 (sometimes), or 3 (often/most of the time) by a graduate student who was trained by the first author. Inter-observer agreement was computed on a 20% sample of randomly chosen tapes and ranged from Cohen’s k .84 to .98 for the six scores (i.e., three scores in two situations). Principal Component Analysis with
Varimax rotation yielded two factors on the six behavioral ratings. The first factor, fear of stranger, had an eigenvalue of 2.98 (38.29% of the variance explained) and variable loadings of .82 for looking to parent during the stranger episode, .89 for proximity seeking during the stranger episode, and .71 for crying during the stranger episode. The second factor, fear of robot, had an eigenvalue of 1.50 (25.01% of the variance explained) and variable loadings of .83 for looking to parent during the robot episode, .75 for proximity seeking during the robot episode, and .68 for crying during the robot episode. To create two composite scores, the three scores loading on each factor were summed after standardization. These two scores, fear of stranger and fear of robot, were used as behavioral measures of infant social fearfulness. The correlation between the two composite scores was .19 ($p < .05$).”


**Pittsburgh**

**Publications:** 74–77

**Study design; in- and exclusion criteria**

The present set of structural MRI (sMRI) scans of third generation offspring are part of an ongoing family study that selected families through their parents’ generation. The goal of the larger longitudinal study was to contrast offspring from high and low-risk for alcohol dependence families on the basis of neurobiological and clinical status. Accordingly, offspring were followed through childhood at approximately annual intervals and through young adulthood, biennially. Extensive assessment of psychiatric disorders including alcohol and drug use information was obtained at each follow-up wave using age appropriate instruments. All participants provided consent with each visit. Children provided assent with parental consent. The study has ongoing approval from the University of Pittsburgh Institutional Review Board. Although 102 individuals were scanned at 3T before age 25, only 64 signed consent for sharing their data, and 15 of them had data on childhood temperament (peer play, see next page). The high-risk families were identified through a proband pair of alcohol dependent brothers or pair of sisters as previously described 74,75. Both members of the proband pair were screened using an in-person structured interview (Diagnostic Interview Schedule; DIS 78) to determine the presence of alcohol dependence and other Axis I psychopathology. Selection of control pedigrees was based on availability of a nuclear family with children between the ages of 8-18 and through parents who were screened for absence of alcohol and drug dependence using the DIS.
Diagnostic interview

Each child/adolescent and his/her parent were separately administered the Schedule for Affective Disorders and Schizophrenia (K-SADS) by trained, Masters’ level clinical interviewers and an advanced resident in child psychiatry at each annual evaluation. A reliable best-estimate diagnosis was obtained for all major DSM-III diagnoses at approximately yearly intervals. Quantity and frequency of use of commonly used substances (e.g., alcohol, cannabis, benzodiazepines, opioids) was also obtained. Young adult assessments included the Composite International Diagnostic Interview (CIDI) and CIDI-Substance Abuse Module (CIDI-SAM), providing diagnoses for all DSM-IV diagnoses. Information concerning lifetime use of substances prior to the MRI scan was derived from the K-SADS, CIDI, and CIDI-SAM interview data. The presence of a SUD diagnosis was determined based on the outcome of the K-SADS or CIDI interviews.

Because multiple evaluations were available for each participant, the clinical evaluation closest in time but preceding the scan was chosen. Only those diagnoses that occurred prior to the scan and within 1 year of the scan were included as current diagnoses. Because neuropsychological testing that included IQ assessment occurred less frequently than the clinical interviews during childhood, only those occurring within 2 years prior to the scan were included. The educational attainment at the point in time where IQ was selected was chosen to indicate the level of education at the time of the scan.

Measures of temperament

Peer play procedure (4 – 6 year olds): The peer play study included 36 children who were assessed with different pairings with other children, totaling 100 sessions in all. In each pairing, a child was paired with one other child whom he/she had never met, in up to three separate sessions. Both children had mothers present within the test room who were asked to quietly observe. Observations were made during the 30-minute play session through a one-way
mirrored window supplemented by cameras which provided additional views of the playroom though monitors in the observation room where coders were located. All coders met an interrater reliability criterion of $r = 0.90$ with other coders. The sessions were scored for: (1) amount of time spent proximal to the parent (within the parent's reach); (2) the amount of time staring at the other child, neither speaking nor playing with the child at the time staring occurred. Also, latency to speak, latency to touch the playroom toys and the total amount of speech were recorded. Most children participated in three sessions of peer play. For this mega-analysis, we created a sum score of 1) average (over all sessions) total amount of time staring at the other child, 2) average amount of time spent proximal to the parent, 3) average latency to speak, as an index of cIT.

**Adolescent temperament measures (around time of scan)** As part of the longitudinal follow-up, subjects were administered the Multidimensional Personality Questionnaire (MPQ). The MPQ provides 11 personality scales and 3 higher order scales. Assessment was completed within 1 year of the scans.
San Raffaele

Publications: 87–90

In- and exclusion criteria (from 87)

“Subjects were drawn from a sample of 49 normally developing children who had participated in an ERP study and shyness at age 8–9 89. The 49 ERP study participants had been drawn from a general population cohort (n = 149) assessed at age 7 for shyness 88. In 2007–2008, we invited all 49 children and their families to a new phase of the study, which encompassed fMRI sessions and direct psychiatric interviews: 38 (78%) accepted, 4 (8%) refused, and 7 (14%) were unavailable due to relocation. Amongst the 38 acceptant subjects, 17 withdrew for the presence of orthodontic apparels, health/family problems, or for sickness/unexpected constraints on the experiment day. This left 21 participants to this study. The procedures were accepted by the ethical committee of the participating institutions and, after complete description of the study to the subject, parental written informed consent was obtained.”

Diagnostic interview

“The presence of symptoms of DSM-IV childhood disorders was established by consensus of the first two authors via blinded reviews. K-SADS interviews were administered to parents while their children were undergoing fMRI on the day of the experiment. For all diagnostic categories, the K-SADS instructions 91 were followed and applied rigorously.” 87

Measures of temperament

As described in the baseline-paper of this longitudinal study88, the assessment of cIT consisted of several steps.
Training of Teachers as Informants and Psychometric Indices

“Before beginning the study, teachers were invited to a lecture on childhood behavioral inhibition and social anxiety disorder, and they also participated in a hands-on seminar on the format and wording of the questionnaire. All items were presented in detail and examples of behaviors that applied to items were provided. Further training of teachers was provided through question times and educational papers on childhood behavioral inhibition and social anxiety disorder. The questionnaire sought to identify (1) symptoms of possible social anxiety disorder proper and (2) temperamental disposition to behavioral inhibition as possible correlates and external validators of social anxiety symptoms. Three different scales were used in the questionnaire: (1) the Liebowitz Social Anxiety Scale adapted for children (LSAS) to evaluate symptoms of fear and avoidance of social situations using DSM-IV criteria of social anxiety disorder, (2) the Shyness-to-the-Unfamiliar (SU) Scale to evaluate temperamental shyness and the modality of approach to the unfamiliar, and (3) the Harm Avoidance (HA) scale of Cloninger’s Junior Temperament and Character Inventory, Parent version, to measure temperamental disposition toward avoidant behaviors in the face of uncertainty.

LSAS Scale

The LSAS is usually administered to children or to informants by an interviewer, but for the purposes of this study the wording was adapted to allow the scale’s completion by trained teachers who acted as informants. The original LSAS includes 24 items rated 0 (“no fear/never avoids”) to 3 (“severe fear/usually avoids”), but for our purposes items 5, 9, and 21 (“talking/answering telephone” and “urinating in public restroom”) were excluded because teachers could not rate them adequately, so that in our study the LSAS ranged from 0 to 63. Previous studies showed validity and clinical usefulness of the LSAS in (1) assessing the reduction of fear and avoidance of social contests in school-based behavioral treatments for social anxiety disorder in adolescents and (2) clarifying the presence of social
anxiety disorder as the salient clinical feature of childhood selective mutism \textsuperscript{94}

\textit{SU Scale}

The SU assesses the initial approach to/withdrawal from unfamiliar people: it encompasses six items (e.g., “readily plays with a new child,” “avoids new guests/visitors”) rated 0 to 5 (from “almost never” to “almost always”), and ranges from 0 to 30. Behavioral inhibition evaluated by the SU scale has revealed good temporal stability ($r \sim 0.5$) from age 4.5 to age 7 \textsuperscript{99}. The SU scale has been linked with some physiological correlates of behavioral inhibition and generally has predicted familial social phobia; subjects rated as behaviorally inhibited with the SU scale tend to have a higher heart rate \textsuperscript{99}. Furthermore, in a community study of schoolchildren, an association was found between children’s higher rates on the SU and mothers’ heightened risk for social phobia \textsuperscript{96}.

\textit{HA Scale}

The HA scale encompasses 22 true/false items and thus ranges from 0 to 22; it is organized into four subscales: Fear of Uncertainty, Worry and Pessimism, Shyness With Strangers, and Fatigability. As with the LSAS, the HA parent version was modified to allow teachers’ use. Harm avoidance has a heritability of about 0.5 \textsuperscript{98}, is relatively stable from childhood to adult life \textsuperscript{100}, and heightens the risk of developing symptoms of anxiety/depression for people with extreme scores on the HA scale \textsuperscript{101}.

The teachers were asked to base their judgment on the instructions received from our group at seminars and on direct observation. On the basis of an anonymous review of pupils’ individual reports available from the school archive, children were excluded from assessment if they (1) had joined the class less than 6 months earlier, (2) displayed mental/physical handicaps that would require special attention, such as a remedial teacher, and (3) revealed learning disabilities and/or overt attention difficulties. This left 149 subjects (mean age $7.5 \pm 0.5$ years) who
underwent an expression discrimination trial that was administered at school by four psychologists trained in the evaluation of childhood behavior.”

For the present mega-analysis, we will use an empirical composite index of eIT encompassing latency of first spontaneous comment, items from the Stevenson-Hinde and Glover Shyness to the Unfamiliar, Cloninger's Harm Avoidance and the Liebowitz Social Anxiety Scale adapted for children."
In- and exclusion criteria (from \textsuperscript{102,104}); diagnostic information

The sample comprised healthy volunteers and youth diagnosed with an anxiety disorder, disruptive mood dysregulation disorder, or ADHD by licensed clinicians using the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) \textsuperscript{91}. Exclusion criteria were neurological disorders, autism and bipolar spectrum disorders, psychosis, substance use, MRI contraindications, and Full Scale IQ below 70. Anxiety was assessed by using the parent- and youth-reported ratings of the five subscales of the Screen for Child Anxiety Related Disorders (SCARED) \textsuperscript{12}.

Measures of temperament

Behavioral Inhibition Questionnaire \textsuperscript{105}.
In- and exclusion criteria (described in \textsuperscript{110})

For this longitudinal study, “participants were recruited from the community utilizing commercial mailing lists, screened for any major medical conditions, and required to have at least one English-speaking biological parent. Exclusionary criteria included any developmental disabilities, metal or electronic implants, a history of head trauma, or use of medications known to affect brain functioning (e.g., antihistamines, pain killers). Participants were oversampled based on their temperamental negative emotionality, low positive emotionality, or behavioral inhibition, assessed observationally when they were 3 years old (see \textsuperscript{111}). This oversampling was done as the broader goal of the study was to understand early childhood risk factors for later depressive and anxiety disorders, for which high negative emotionality, low positive emotionality, and high behavioral inhibition are risk factors (see Olino et al. \textsuperscript{112} for details). Negative and positive emotionality as well as behavioral inhibition were assessed via the Laboratory Temperament Assessment Battery (LabTAB) \textsuperscript{46}, which involves a standardized set of tasks designed to elicit children’s bodily, vocal, and facial expressions of a range of emotions (see \textsuperscript{113}).”

The MRI sample was a subsample of the age 3 sample and were selected on age 3 temperament traits based on the LabTAB at age 3.

Diagnostic interview

Parents completed the Preschool Age Psychiatric Assessment interview at ages 3 and 6 \textsuperscript{114}; parents and youth completed the K-SADS at ages 9, 12, and 15.
Measures of temperament

(Quoted from 113): “Each child and a parent (95.0 % mothers) visited the laboratory for a 2-hour observational assessment of temperament that included a standardized set of 12 episodes selected to elicit a range of temperament-relevant behaviors. Eleven episodes were from the Laboratory Temperament Assessment Battery (Lab-TAB) 46 and one was adapted from a Lab-TAB episode. Using an independent sample, we previously reported moderate stability of laboratory ratings of temperament from ages 3 to 7 (r = .46 and .45 for positive emotionality and negative emotionality, respectively), and moderate concurrent and longitudinal associations between Lab-TAB ratings and home observations 115. Each task was videotaped through a one-way mirror and later coded. To prevent carryover effects, no episodes presumed to evoke similar affective responses occurred consecutively and each episode was followed by a brief play break to allow the child to return to a baseline affective state. The parent remained in the room with the child for all episodes except Stranger and Box Empty, but was instructed not to interact with the child (except in Pop-Up Snakes), and was seated facing at a right angle from the experimenter and child and given questionnaires to complete.

The episodes, in order of presentation, were: (1) Risk Room. Child explored a set of novel and ambiguous stimuli, including a Halloween mask, balance beam, and black box; (2) Tower of Patience. Child and experimenter alternated turns in building a tower. The experimenter took increasing amounts of time before placing her block on the tower during each turn; (3) Arc of Toys. Child played independently with toys for five minutes before the experimenter asked the child to clean up the toys; (4) Stranger Approach. Child was left alone briefly in the room before a male accomplice entered, speaking to the child while slowly walking closer; (5) Make that Car Go. Child and experimenter raced remote-controlled cars; (6) Transparent Box. Experimenter locked an attractive toy in a transparent box, leaving the child alone with a set of non-working keys. After a few minutes, the experimenter returned and told the child that she
had left the wrong set of keys. The child used the new keys to open the box and play with the toy; (7) Exploring New Objects. Child was given the opportunity to explore a set of novel and ambiguous stimuli, including a mechanical spider, a mechanical bird, and sticky soft gel balls; (8) Pop-up Snakes. Child and experimenter surprised the parent with a can of potato chips that actually contained coiled snakes; (9) Impossibly Perfect Green Circles. Experimenter repeatedly asked the child to draw a circle on a large piece of paper, mildly criticizing each attempt; (10) Poppin Bubbles. Child and experimenter played with a bubble-shooting toy; (11) Snack Delay. Child was instructed to wait for the experimenter to ring a bell before eating a snack. The experimenter systematically increased the delay before ringing the bell; and (12) Box Empty. Child was given an elaborately wrapped box to open under the impression that a toy was inside. After the child discovered the box was empty, the experimenter returned with several toys for the child to keep.

**Coding Procedures** Behavioral inhibition (BI) was coded using an approach that was similar to most previous studies. The three episodes specifically designed to assess BI (Risk Room, Stranger Approach, Exploring New Objects) were divided into 20 or 30 second epochs, and a series of affective and behavioral codes were rated for each epoch. Within each epoch, a maximum intensity rating of facial, bodily, and vocal fear was coded on a scale of 0 (absent) to 3 (highly present and salient). Based on previous studies using the Lab-TAB, BI was computed as the average standardized ratings of latency to fear (reversed); and facial, vocal, and bodily fear (Risk Room, Stranger Approach, and Exploring New Objects); latency to touch objects; total number of objects touched (reversed); tentative play; referencing the parent; proximity to parent; referencing the experimenter; and time spent playing (reversed) (Risk Room and Exploring New Objects); startle (Exploring New Objects); sad facial affect (Exploring New Objects and Stranger Approach); and latency to vocalize; approach towards
the stranger (reversed); avoidance of the stranger; gaze aversion; and verbal/nonverbal interaction with the stranger (reversed; Stranger Approach).”

At age 3, a parent also completed the Behavioral Inhibition Questionnaire 26.

For the present mega-analysis, we will use a sum score (log-transformed) from 3 Kagan-like tasks in Goldsmith's Lab-TAB as an index of cIT.
In- and exclusion criteria

This study concerns a longitudinal project. The selection of participants was as follows (described by Hane et al., 2008): “Families identified via commercially available mailing lists were sent a letter about the project and were asked to complete a form and send it back to the laboratory. Interested mothers of developmentally healthy infants were scheduled for a laboratory visit between their infant’s 15th and 17th weeks.

Four-Month Selection. 779 infants were screened for degree of reactivity to visual and auditory stimuli at four months (see 42,121). Infant behavior during the reactivity paradigm was subsequently coded as follows: A motor reactivity score was obtained by summing the frequencies of arm waves, arm wave bursts (several waves in rapid succession), leg kicks, leg kick bursts, back arches and hyper extensions throughout the paradigm. A negative affect score was derived by summing the frequencies of fussing and crying and a positive affect score was obtained by summing the frequencies of smiling and positive vocalizations.

The first 100 infants screened were used as a criterion group, i.e., their negative, positive, and motor reactivity scores were used to set the selection criteria for all subsequent infants as follows: Infants who scored above the criterion group mean on both negative affect and motor arousal and below the mean on positive affect served as the negatively reactive (NR) group (n = 75). Infants who scored above the criterion group mean on both positive affect and motor arousal and below the mean on negative affect served as the positively reactive (PR) group (n = 73). Eighty-six infants who did not meet the criteria for either temperament group served as the control sample. Four reliable raters coded the four-month reactivity paradigm, with pairs of coders achieving intraclass correlation coefficients ranging from .80 to .92. A MANOVA
comparing the three temperament groups on the three reactivity dimensions was significant \((p<.001)\). The NR group manifested significantly more negative affect than both the PR and control groups \((F(2, 231) = 75.08, p < .001; \text{Tukey’s HSD both } p’s < .001)\). The PR group displayed significantly more positive affect than the NR and the control groups \((F(2, 231) = 41.94, p < .001; \text{Tukey’s HSD both } p’s < .001)\). The control group showed significantly less motor activity than both the NR and PR groups \((F(2, 231) = 51.17, p < .001; \text{Tukey’s HSD both } p’s < .001)\).

Based on four-month temperament group status, 278 infants were invited to continue participation. 268 children returned to the laboratory at 2 and 3 years of age for BI assessment. At ages 10 and 12, eligible children were invited to participate in brain imaging visits. Brain imaging visits were conducted at the National Institute of Mental Health as part of the longitudinal assessment. Participants were excluded if they were taking any psychotropic medications at the time of scanning. However, subjects on psychostimulant medications who could tolerate a 24-h medication-free period prior to scanning were included. Children were also deemed ineligible to participate in imaging visits if they had an MRI contraindication (i.e., metal in their body). Parental consent was obtained prior to all visits and child assent was obtained prior to 10 and 12 year visit.”

**Diagnostic interview**

Psychopathology around the time of scan was assessed using the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS). Trait anxiety symptoms were measured using the SCARED, a reliable child- and parent-report questionnaire with 42 items \(^{30,122}\).

**Measures of temperament**

At age nine months, data were collected using 6 LabTAB tasks: 2 anger/frustration tasks (arm restraint, toy barrier), 2 fear tasks (masks, unpredictable toy), and 2 joy tasks (peek-a-boo,
Furthermore, inhibited behavior to novel stimuli was coded at 24 and 36 months. Behavioral scores were standardized at each time point.

For the present study, we will use the ‘BI-classic-24’ index, composed of scores on the stranger, robot and tunnel episodes of the Laboratory Temperament Assessment Battery – preschool version.
Vanderbilt - children

Publications: 124 and unpublished

In- and exclusion criteria
Quoted from Clauss et al 124; “Consistent with the extreme discordant phenotypes approach 125, we compared inhibited children and uninhibited children at the extreme ends to maximize our chances of identifying differences. To obtain pure risk groups (not confounded by existing disorders), children were excluded from the study for having any current or past psychiatric diagnoses, as measured by the Kiddie-Schedule for Affective Disorders and Schizophrenia–Present and Lifetime Version (KSADS-PL) 91 or having received treatment for anxiety symptoms. Children were also excluded if they had cognitive deficits that might affect task performance (developmental delay, repeating a grade, or receiving special assistance in school), contraindications to MRI scanning, or factors that might affect blood oxygen level-dependent (BOLD) signal (psychotropic medications, history of head injury, major medical or neurological conditions). Intelligence quotient (IQ) was assessed using the Kaufman Brief Intelligence Test 126. Handedness was assessed using the Edinburgh Handedness Inventory 127. Participants were recruited from the Vanderbilt University Medical Center and surrounding community using flyers, e-mails, and research recruitment databases. Advertisements were for children who were “quiet,” “cautious,” “shy,” “outgoing,” and general recruitment for a study on “temperament and brain function.” Before the first study visit, parents completed a brief online screening, including the Behavioral Inhibition Questionnaire-Parent (BIQ-P) 26, a validated measure of childhood inhibited temperament, which shows convergent validity with behavioral measures and other measures of social inhibition 27,31. Although 4 questions in the questionnaire refer to younger age groups, these questions were highly correlated with other items in the scale and therefore were retained as written. Children
were selected based on a temperament score plus or minus 1 standard deviation from the mean based on published norms (inhibited >123; uninhibited <59)\textsuperscript{26}; these norms were similar to those identified in children and adolescents 4 to 15 years of age\textsuperscript{39} and those used in a recent similar neuroimaging study \textsuperscript{16}.

**Diagnostic interview and psychiatric symptom measures**

Kiddie-Schedule for Affective Disorders and Schizophrenia–Present and Lifetime Version (KSADS-PL) \textsuperscript{91} - see above. To further characterize participants, both parents and children reported on a number of psychiatric symptom measures, including, among others the Screen for Child Anxiety-Related Disorders \textsuperscript{122} and the Children’s Depression Inventory \textsuperscript{128}.

**Measures of temperament**

Children completed a self-report of temperament, the Behavioral Inhibition Questionnaire-Child (BIQ-C). \textsuperscript{27}. The total score will be used as an index of cIT in the present mega-analysis.
In- and exclusion criteria

The sample in this mega-analysis consist of participants from multiple previously published studies. In general, as described in Clauss et al. 132, subjects were recruited by seeking individuals ages 18–25 who were “extremely shy or outgoing.” Consistent with prior studies 130,131,134, individuals with an extreme inhibited or extreme uninhibited temperament were identified using the Adult Self-Report of Inhibition (ASRI) and the Retrospective Self-Report of Inhibition (RSRI) 65. Subjects were selected for having a stable temperament (i.e., being extremely inhibited or extremely uninhibited as both an adult and a child), defined by scores on both the ASRI and RSRI that were greater than one standard deviation from published means. Other inclusion criteria included: passing an MRI safety screen, being free of psychoactive medications within the past 6 months, having no history of brain trauma, and having no psychiatric illness (based on clinical interview), except anxiety disorders in the inhibited temperament group. Inhibited subjects who met criteria for a current or past anxiety disorder were not excluded.

Diagnostic interview

Psychiatric diagnosis was assessed by a trained clinical interviewer using the Structured Clinical Interview for DSM-IV 135.

Measures of temperament

To focus on a stable trait and ensure that valid groups were identified, inhibited temperament was assessed retrospectively (childhood) and currently using the Retrospective Self-Report of
Inhibition (RSRI) and the Adult Self-Report of Inhibition (ASRI)\textsuperscript{65}, respectively. Both questionnaires have excellent reliability (Cronbach’s alpha = 0.79 for the RSRI and 0.78 for the ASRI), demonstrate convergent validity\textsuperscript{65,136} and minimize self-report bias by focusing on reports of concrete behaviors in specific situations instead of subjective feelings\textsuperscript{137}. In the sample for the current mega-analysis ($n = 150$), scores on the RSR and ASRI were highly correlated ($r = 0.91, p < 0.001$); we will use the total scores on the RSRI as index of cIT.
Western University

Publications: 138–144

In- and exclusion criteria
As described in Vandermeer et al., 2020 142, children (n = 87) and their mothers were recruited from a larger longitudinal study of children’s depression risk (n = 409) that began when children were 3-year-olds. At baseline, children with major medical or psychological problems were excluded, and typical cognitive development was verified using the Peabody Picture Vocabulary Test-Fourth 145. For the current study, children were recruited from the larger longitudinal sample based on maternal history of depression (MH+) drawn from data collected at a previous round of data collection for this study 138. Children were considered high-risk based on a maternal history of recurrent major depression (n = 26), or a maternal lifetime history of a single major depressive episode and a serious anxiety disorder (i.e., any anxiety disorder except a specific phobia; n = 3). Low-risk children had no maternal history of major depression or anxiety disorder. From this sample, 237 families were contacted (58 MH+). Children with any contraindications to the MRI scan (e.g., braces, metallic objects implanted in the body, claustrophobic) were deemed ineligible, leaving a pool of 231 families, from which 110 families agreed to participate (36 MH+). Children from these families were screened as described in the following section to ensure the absence of current or lifetime depressive disorder. Eighty-seven children (29 MH+; 49 boys) participated in the MRI session.

Diagnostic interview
Children were administered the K-SADS-PL and completed self-reported symptom and severity measures, including the Children’s Depression Inventory 2nd Edition
Measures of temperament

A detailed description of the assessment of temperament is provided in previous publications on the study, for example Vandermeer et al.\textsuperscript{140} and Liu et al.\textsuperscript{138}. Of interest for the present work is the baseline measurement of the study, in which temperament was assessed during a standardized lab visit based on the Laboratory Temperament Assessment Battery (Lab-TAB;\textsuperscript{46}) and an age-adapted version of the Lab-TAB\textsuperscript{115}. During this lab visit, children participated in tasks drawn directly from the preschool-aged version of the Lab-TAB. The assessment of behavioral inhibition (BI) at age three consisted of three Lab-TAB tasks: Risk Room, Stranger Approach, and Exploring New Objects, as described in\textsuperscript{138,140}. Age 3 BI scores were a composite score based on the average of z-scores for coded variables for the three tasks. These procedures for computing BI composite scores are consistent with other studies using observational coding (e.g.,\textsuperscript{107,148}). These age 3 scores will be used as an index of cIT in the present mega-analysis.
Wisconsin Twin Project

Publications: 149–151

Study design (from 150)

“The Wisconsin Twin Project sample is based on birth-record based cohorts of twins born in the state of Wisconsin during the years 1989–2004 151–154. After nearly 30 years, the research program encompasses a series of longitudinal studies that span infancy to early adulthood. Twin family recruitment and early results were covered in prior overviews of the project 151,152. Briefly, initial contact was attempted with a mailed letter and contact form. Contact was maintained with multiple phone numbers, email addresses, a toll-free phone number and secondary contact information from a family friend or relative. Sample retention efforts included newsletters and a website devoted to participant communication. Web-based tracing methods (e.g., public court records) were used to locate families with whom we had lost contact. The University of Wisconsin Survey Center also provided tracing services. All of these procedures helped maintain the research sample longitudinally. The research is conducted at the University of Wisconsin–Madison’s Waisman Center and the Department of Psychology (https://goldsmithtwins.waisman.wisc.edu/). Procedures in studies under the Wisconsin Twin Project were approved by University of Wisconsin–Madison Internal Review Boards and comply with the Helsinki Accords of 1975, as revised in 2008.”

In the present mega-analysis, we will use data from the RDoC twin study (participants < 13 years of age), as these data are available through the National Institute of Mental Health Data Archive (NDA). As summarized in Schmidt et al. 150, “the RDoC twin study used longitudinal an quantitative genetic approaches to establish developmental antecedents and neural substrates for the RDoC positive valence systems (e.g., anticipatory positive affect and contentment) an negative valence systems (e.g., acute fear, potential threat/anxiety, frustrating non-reward and
loss). The RDoC twin study aimed to (a) establish distinctiveness, stability and external validity of each RDoC construct during childhood and adolescence; (b) investigate the relationship between brain structure and function (via MRI) and concurrent and longitudinal RDoC measures; and (c) utilize the MZ difference design to highlight early environmental contributions to later brain structure and function. In the MRI analyses, we focused more on white matter microstructure and on resting state and task-related functional measures (i.e., circuitry and networks) than on gray matter structure. The RDoC twin study enrolled 518 MZ (56%) and dizygotic (DZ) twin individuals (mean age = 17.4 years, SD = 2.2 years). Approximately 70% were under 18 years of age at the time of assessment. Parents (88% mothers) of adolescents completed surveys. Data collection concluded in early 2019.

**Diagnostic interview**

Not available.

**Measure of temperament**

In the present analysis, we will use the temperament assessments which took place during middle childhood (age 7), as described in\(^{155}\). “The four-hour, in-home assessment involved additional parent questionnaires and interviews, child interviews, observer ratings, and the Lab-TAB. Data collection for this middle childhood phase took place across more than five years.

**Laboratory Temperament Assessment Battery.** Lab-TAB\(^{156}\) is a laboratory-based behavioral assessment that comprises multiple episodes designed to tap observable elements of temperament dimensions. Lab-TAB was administered during the childhood home visit and was modified slightly for use in homes\(^{157}\). During Lab-TAB administration, children’s behavior was videotaped and later coded by individuals blind to other information about the child. Individual raters did not rate both twins from the same family. 10% of the videos were rated by
a master coder, and agreement between master coder and the other coder (Kappa > 0.70) was required. Each Lab-TAB episode (3-10 minutes duration), provided multiple responses scored in 5-30 second epochs or in discrete trials. Parameters included latency to first response, occurrence of a target response within an epoch or trial (mean response), and the magnitude or intensity of a target response (peak response). In general, positivity was coded as absence/presence (0/1) while facial, bodily, and vocal angry, sad, and fearful responses were coded on a 0-2 or 0-3 scale. For detailed descriptions of each episode and of scoring procedures see the Lab-TAB manual 158.

In **Storytelling**, the child stands in front of multiple child testers and is asked to talk about what they did the prior day, with least one prompt given by the child tester 155. In **Stranger approach**, social interaction with an unfamiliar adult wearing hat and sunglasses is investigated 157.

**Post-visit observer ratings** (from Moore et al. 155). “Two child testers from each middle childhood home visit independently completed post-visit ratings for each twin on 28 items related to child behavior, where “1” indicates the absence of the characteristic or behavior and “5” describes an extreme reaction. Behavior was observed throughout the visit, including times before, between, and after administration of Lab-TAB episodes. Some items include modified content from the Behavior Rating Scales (BRS) from the Bayley Scales of Infant Development. Child tester ratings were averaged for each item; item-specific correlations between raters ranged from .38 to .49.”

**Additional questionnaire data (parental report) on inhibited temperament**

**MacArthur Health and Behavior Questionnaire**: inhibition scale from the MacArthur Health and Behavior Questionnaire (HBQ; 159). From Moore et al 155. “Parents rated their child’s behavior over the past six months using a 3-point scale (0 = rarely, 2 = certainly applies). Internal consistency reliability (alpha) for age 7 HBQ subscales ranged from .67 to .84 for mother-report and from .62 to .85 for father-report. Mother and father scores were moderately
and significantly correlated (age 7 rs ranged from .27 to .53, all ps < .001) and were mean-averaged into a single parent-report score at each age.”

**Children’s Behavior Questionnaire (CBQ)**

An abridged 80-item version of the CBQ was completed by both parents. As described in Gagne et al.\(^{157}\), the CBQ requires parents to judge their children’s reactions to a variety of situations over the last 6 months (e.g., “Can lower his/her voice when asked to do so”) and is appropriate for children from 3 years to 7 years of age.\(^{160}\) Each item is rated on a 1–7 scale, with 1 indicating the reaction is extremely untrue of the child and 7 indicating that the reaction is extremely true. CBQ scores have shown high internal consistency, parental agreement, and convergent validity with socialization-relevant traits\(^{160}\) and have been used in numerous studies with a wide range of empirical correlates. The eight CBQ scales that we used were selected for overlap with temperament dimensions assessed in the Lab-TAB, and each CBQ scale had 10 items. Estimates of internal consistency for each CBQ scale were as follows: Anger (α = .78), Fear (α = .73), Shyness (α = .92), Sadness (α = .63), Approach (α = .74), Activity Level (α = .73), Attentional Focusing (α = .78), and Inhibitory Control (α = .82).”

For the present mega-analysis, we used a sum-score of the behavioral observations related to inhibited behavior (ratings on Approach and Shyness from the home visit by 2 observers and the scores from videotaped reactions to the “Conversation with a Stranger” episode of Lab-TAB). This sum-score correlated significantly with the parental reports on inhibited temperament (correlation sum-score with HBQ-Mother (inhibition): \( r = 0.28, p < 0.001 \); with HBQ-Father (inhibition): \( r = 0.18, p = 0.004 \); with CBQ-Mother (shyness): \( r = 0.29, p < 0.001 \); with CBQ-Father (shyness): \( r = 0.30, p < 0.001 \).
Supplemental Tables

Supplemental Table 1  STROBE checklist case-control studies
Supplemental Table 2  Scanner characteristics per sample
Supplemental Table 3  Clinical characteristics per sample
Supplemental Table 4  Overview of included independent variables per sample
Supplemental Table 1  STROBE Statement—Checklist of items that should be included

in reports of case-control studies

<table>
<thead>
<tr>
<th>Item No</th>
<th>Recommendation</th>
<th>Location in manuscript</th>
</tr>
</thead>
</table>
| **Title and abstract**  | 1 | *(a)* Indicate the study’s design with a commonly used term in the title or the abstract  
*(b)* Provide in the abstract an informative and balanced summary of what was done and what was found | Title  
Synopsis |
| **Introduction**  | 2 | Explain the scientific background and rationale for the investigation being reported | Synopsis and Registered Report-Introduction |
| **Objectives**  | 3 | State specific objectives, including any prespecified hypotheses | Registered Report-Introduction |
| **Methods**  | 4 | Present key elements of study design early in the paper | Registered Report-Methods |
| **Setting**  | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | Registered Report-Methods |
| **Participants**  | 6 | *(a)* Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls  
*(b)* For matched studies, give matching criteria and the number of controls per case | Registered Report-Methods  
*Not applicable* |
| **Variables**  | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | Registered Report-Methods |
| **Data sources/ measurement**  | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | Registered Report- Supplemental Methods |
| **Bias**  | 9 | Describe any efforts to address potential sources of bias | Registered Report-Methods |
| **Study size**  | 10 | Explain how the study size was arrived at | Registered Report-Methods |
| **Quantitative variables**  | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | Registered Report-Methods |
| **Statistical methods**  | 12 | *(a)* Describe all statistical methods, including those used to control for confounding  
*(b)* Describe any methods used to examine subgroups and interactions  
*(c)* Explain how missing data were addressed  
*(d)* If applicable, explain how matching of cases and controls was addressed  
*(e)* Describe any sensitivity analyses | Registered Report-Methods  
Registered Report-Methods  
Registered Report-Methods  
*Not applicable*  
Registered Report- Methods |
### Results

**Participants**

- **13***
  - (a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
  - (b) Give reasons for non-participation at each stage
  - (c) Consider use of a flow diagram

**Descriptive data**

- **14***
  - (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders
  - (b) Indicate number of participants with missing data for each variable of interest

**Outcome data**

- **15***
  - Report numbers in each exposure category, or summary measures of exposure

**Main results**

- **16**
  - (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included
  - (b) Report category boundaries when continuous variables were categorized
  - (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

**Other analyses**

- **17**
  - Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses

### Discussion

**Key results**

- **18**
  - Summarise key results with reference to study objectives

**Limitations**

- **19**
  - Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias

**Interpretation**

- **20**
  - Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

**Generalisability**

- **21**
  - Discuss the generalisability (external validity) of the study results

### Other information

**Funding**

- **22**
  - Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls.*
**Supplemental Table 2  Scanner characteristics per sample**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Scanner type</th>
<th>Field-strength</th>
<th>Structural MRI scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brains study</td>
<td>Siemens Trio</td>
<td>3 T</td>
<td>High-resolution T1-weighted structural scans with a magnetization prepared gradient echo sequence (MP-RAGE) (176 1 mm slices, TR = 1700, TE = 2.01, FA = 9°, FOV = 256 mm, voxel size = 1 × 1 × 1 mm; 256 × 256 matrix, T1 = 850 ms).</td>
</tr>
<tr>
<td>Brazilian High Risk Cohort</td>
<td>Siemens Prismafit</td>
<td>1.5 T</td>
<td>T1-weighted structural MR images were acquired with the following parameters: TR=10.916 ms, TE=4.2 ms, slice thickness=1.2 mm, FA =15°, matrix size=2563192, FOV=245 mm, max=156 slices.</td>
</tr>
<tr>
<td>Cohort 3 / 4</td>
<td>GE Healthcare MR750</td>
<td>3 T</td>
<td>High-resolution, T1-weighted structural imaging sequence (MPRAGE; sagittal acquisition; 176 slices; 1 mm³ isotropic voxels; 256 * 256 matrix; flip angle = 7°; TR = 7.7 ms; TE = 3.42 ms; TI = 425 ms).</td>
</tr>
<tr>
<td>Generation R - sample with behavioral</td>
<td>GE MR750</td>
<td>3 T</td>
<td>T1-weighted inversion recovery fast spoiled gradient recalled (IR-FSPGR) sequence with the following parameters: TR = 8.77 ms, TE = 3.4 ms, TI = 600 ms, FA = 10°, matrix 220x220, slice thickness 1.0 mm, in-plane resolution (mm) 1.0 mm.</td>
</tr>
<tr>
<td>observations and Generation R - sample with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>questionnaire data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland-PAX</td>
<td>Siemens Magnetom TIM Trio</td>
<td>3 T</td>
<td>Sagittal T1-weighted anatomic images with MPRAGE sequence (TR = 2400 ms; TE = 2.01 ms; TI = 1060 ms; FA = 8°; sagittal slice thickness = 0.8 mm; in-plane = 0.8 x 0.8 mm; matrix = 300x300, FOV= 240x240).</td>
</tr>
<tr>
<td>Maryland-TAX</td>
<td>Siemens Magnetom TIM Trio</td>
<td>3 T</td>
<td>High-resolution anatomical (T1-weighted) images with a magnetization-prepared, rapid-acquisition, gradient-echo sequence (TR = 1900 ms; TE = 2.32 ms; TI = 900 ms; FA = 9°; sagittal slice thickness = 0.9 mm; voxel size in plane = 0.449 x 0.449 mm; matrix = 512 x 512; FOV = 230 x 230).</td>
</tr>
<tr>
<td>Nijmegen Longitudinal Study</td>
<td>Siemens Magnetom = Trio or PRISMA</td>
<td>3 T</td>
<td>MPRAGE sequence (TR = 2300 ms; TE = 3.03 ms; 192 sagittal slices; 1.0 x 1.0 x 1.0 mm voxels; FOV = 256 mm).</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>Siemens Trio scanner</td>
<td>3 T</td>
<td>7-min 3D T1-weighted Magnetization Prepared Rapid Gradient Echo Imaging (MPRAGE) sequence (TR = 2300 ms, TE = 2.98 ms, FA = 9°, field of view FOV=240 mm, acquisition matrix=240 x 256, in-plane resolution 1.0x1.0 mm², yielding 160 transversal slices with a thickness of 1.2 mm).</td>
</tr>
<tr>
<td>San Raffaele</td>
<td>Philips Achieva</td>
<td>3 T</td>
<td>T1-weighted scan, 150 axial slices, resolution 1mm x 1mm x 1mm</td>
</tr>
<tr>
<td>SDAN</td>
<td>General Electric MR750</td>
<td>3 T</td>
<td>Whole-brain, high-resolution, T1-weighted anatomical scan (MPRAGE; 176 axial slices, 256 x 256 matrix,1 mm³ isotropic slices; FA = 7°, FOV = 220 mm; TR = 7.7ms, TE = 3.42s).</td>
</tr>
<tr>
<td>Stony Brook Temperament Study</td>
<td>Siemens Trio</td>
<td>3 T</td>
<td>T1-weighted high resolution structural images with the magnetization prepared gradient echo (MPRAGE) sequence: slices = 176, slice thickness =1 mm, TR = 2400 ms, TE = 3.16 ms, FA=8°; matrix size = 256 × 256, FOV=256 × 256 mm, resolution=1 x 1 x 1 millimeters</td>
</tr>
<tr>
<td>TOTS</td>
<td>General Electric MR750</td>
<td>3 T</td>
<td>High-resolution T1-weighted whole-brain volumetric scan with a high-resolution magnetization prepared gradient echo sequence (MPRAGE; TE = min full; TI = 425 ms; FA = 7°; FOV = 256 mm; matrix = 256 x 256; in-plane resolution = 1 x 1 x 1mm).</td>
</tr>
<tr>
<td>Vandybilt - children</td>
<td>Philips</td>
<td>3 T</td>
<td>T1-weighted structural data were acquired using the following parameters: 256 mm field of view (FOV), 170 slices, 1-mm slice thickness, 0-mm gap, 2-second TR, 22- millisecond TE, 90 ° flip angle, 1.8 SENSE factor, 240-mm FOV, 3 * 3 mm in-plane resolution.</td>
</tr>
<tr>
<td>Vandybilt - young adults</td>
<td>Philips</td>
<td>3 T</td>
<td>High resolution T1-weighted anatomical images (256mm FOV, 170 slices, 1-mm slice thickness, 0-mm gap)</td>
</tr>
<tr>
<td>Western University</td>
<td>Siemens Trio</td>
<td>3 T</td>
<td>T1-weighted 3D magnetization prepared rapid gradient echo (MPRAGE) sequence (1* 1 * 1 mm), voxel size, TR = 2300 ms, TE = 2.98 ms, FOV = 256 mm), 192 slices.</td>
</tr>
<tr>
<td>Wisconsin Twin Project - RDoC twin study</td>
<td>GE SIGNA (Discovery MR750)</td>
<td>3 T</td>
<td>T1-weighted structural images (1 mm³ voxels) were also acquired axially with an isotropic 3D Bravo sequence (TE = 3.2 ms, TR = 8.2 ms, Ti = 450 ms, flip angle = 12°).</td>
</tr>
</tbody>
</table>

**Abbreviations:** FA: flip angle; FOV: field of view; T: Tesla; TE: echo time; TI: inversion time; TR: repetition time.
## Supplemental Table 3a  Clinical characteristics per sample – psychopathology at (around) time of scan

### Psychopathology based on clinical interview (current / lifetime but not current)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Clinical interview</th>
<th>Clinical interview n (n female)</th>
<th>Anxiety disorders</th>
<th>Other psychopathology</th>
<th>Psychotropic medication at time scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brains study</td>
<td>C-DISC-4 administered to parents</td>
<td>130 (72)</td>
<td>SAD 3 SAD, 4 SAD, 7 SAD</td>
<td>Other disorders 0, 0, 0, 0, 0</td>
<td>C-DISC-4 administered to parents 13 current 9 ADHD</td>
</tr>
<tr>
<td>Brazilian High Risk Cohort</td>
<td>Development and Well-678 (290) (DAWBA) Assessment</td>
<td>678 (290)</td>
<td>Other anxiety 3/na 24/na 31/na</td>
<td>Other anxiety nos 1/5 0/0 0/0 1/5</td>
<td>1 current 1 ADHD, current 7 ADHD</td>
</tr>
<tr>
<td>Cohort 3 / 4</td>
<td>Structured Clinical 95 (51) Interview for DSM-IVTR</td>
<td>584 (297)</td>
<td>Other anxiety nos 2/10 0/0 0/0</td>
<td>Other anxiety nos 2/10 0/0 0/0</td>
<td>1 current ADHD, current ODD, lifetime ADHD, lifetime ADD, lifetime ODD, lifetime Tourettes</td>
</tr>
<tr>
<td>Generation R - sample with behavioral observations</td>
<td>Diagnostic Interview Schedule for Young Children (DISC-YC) (note: at age 5 – 8)</td>
<td>53 (28)</td>
<td>Other anxiety nos 12/na 4/na 8/na</td>
<td>Other anxiety nos 12/na 4/na 8/na</td>
<td>2 current ADHD</td>
</tr>
<tr>
<td>Generation R - sample with questionnaire data</td>
<td>Diagnostic Interview Schedule for Young Children (DISC-YC) (note: at age 5 – 8)</td>
<td>1982 (1030 12</td>
<td>Other anxiety nos 37/3 3/17 30/21</td>
<td>Other anxiety nos 37/3 3/17 30/21</td>
<td>1 current ADHD, current ODD, lifetime ADHD, lifetime ADD, lifetime ODD, lifetime Tourettes</td>
</tr>
<tr>
<td>Maryland-PAX</td>
<td>Structured Clinical 220 (109) Interview for DSM-5 (SCID-5-RV)</td>
<td>220 (109)</td>
<td>Other anxiety nos 1/4 0/3 14/2 14/45</td>
<td>Other anxiety nos 1/4 0/3 14/2 14/45</td>
<td>1 current ADHD, current ODD, lifetime ADHD, lifetime ADD, lifetime ODD, lifetime Tourettes</td>
</tr>
<tr>
<td>Maryland-TAX</td>
<td>Mini-International Neuropsychiatric Interview For Children And Adolescents (MINI-KID; Sheehan et al, 1998) based on DSM-IV and ICD-10 criteria</td>
<td>53 (28)</td>
<td>Other anxiety nos 12/na 4/na 8/na</td>
<td>Other anxiety nos 12/na 4/na 8/na</td>
<td>2 current ADHD</td>
</tr>
<tr>
<td>Sample</td>
<td>Clinical interview</td>
<td>n (n female)</td>
<td>SAD</td>
<td>PD</td>
<td>AG</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------</td>
<td>--------------</td>
<td>-----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Nijmegen Longitudinal Study</td>
<td>na</td>
<td>71 (31)</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>K-SADS-PL administered to parent and child</td>
<td>15 (3)</td>
<td>0 / 1</td>
<td>0 / 1</td>
<td>0 / 0</td>
</tr>
<tr>
<td>San Raffaele</td>
<td>K-SADS-PL administered to parents</td>
<td>20 (8)</td>
<td>5 / na</td>
<td>0 / na</td>
<td>0</td>
</tr>
<tr>
<td>SDAN</td>
<td>K-SADS-PL administered to parent and child</td>
<td>55 (26)</td>
<td>18 / 2</td>
<td>0 / 0</td>
<td>0 / 0</td>
</tr>
<tr>
<td>Stony Brook</td>
<td>K-SADS-PL administered to parent and child</td>
<td>74 (31)</td>
<td>1 / 2</td>
<td>0 / 0</td>
<td>0 / 0</td>
</tr>
<tr>
<td>TOTS</td>
<td>K-SADS-PL administered to parent and child</td>
<td>96 (56)</td>
<td>4 / 2</td>
<td>0 / 0</td>
<td>0 / 0</td>
</tr>
<tr>
<td>Vanderbilt children</td>
<td>K-SADS-PL</td>
<td>55 (33)</td>
<td>4 / 0</td>
<td>0 / 0</td>
<td>0 / 0</td>
</tr>
<tr>
<td>Vanderbilt young adults</td>
<td>Structured Clinical Interview for DSM-IV</td>
<td>150 (83)</td>
<td>15 / 1</td>
<td>2 / 0</td>
<td>1 / 0</td>
</tr>
<tr>
<td>Western University</td>
<td>K-SADS-PL administered to parent and child</td>
<td>87 (38)</td>
<td>1 / 1</td>
<td>0 / 0</td>
<td>0 / 0</td>
</tr>
</tbody>
</table>
Psychopathology based on clinical interview (current / lifetime but not current)

<table>
<thead>
<tr>
<th>Anxiety disorders</th>
<th>Other psychopathology</th>
<th>Psychotropic medication at time scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>Clinical interview</td>
<td>n (n female)</td>
</tr>
<tr>
<td>Wisconsin Twin</td>
<td>316 (145)</td>
<td>na</td>
</tr>
<tr>
<td>Project - RDoC</td>
<td>twin study</td>
<td>na</td>
</tr>
</tbody>
</table>
Supplemental Table 3b  Clinical characteristics per sample – symptomatology at (around) time of scan

Questionnaires – scores represent mean ± SD

<p>| Sample                        | STAI-trait | ASI | BAI | LSAS - total | PSWQ | BDI | CDI | SCARED-PT | SCARED-ST | SCARED-SC | SCARED-PN | SCARED-GD | SCARED-SH | YSR - total | Note                                                                 |
|-------------------------------|------------|-----|-----|--------------|------|-----|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|---------------------------------------------------------------------|
| Brains study                  | na         | na  | na  | na           | na   | 11.3±9.1 | 17.2±11.8 | 4.8±3.5  | 4.8±3.6  | 3.1±3.5  | 3.8±3.5  | 3.8±0.4   | 1.0±1.2   | na         | na          | Data availability for SCARED subscales: n = 104 (child-report) and n=118 (parent-report) |
| Brazilian High Risk Cohort    | na         | na  | na  | na           | na   | 24.6±5.7  | 3.5±3.5   | 4.7±na   | 6.0±na   | 4.5±na   | 1.5±1.7   | na         | na          | na          | Data availability for diagnostic information: n = 69 participants; data availability for SCARED: range 40 - 43 participants. |
| Cohort 3 / 4                  | na         | na  | na  | na           | na   | 6.9±6.5   | 12.3±9.7  | 3.5±2.7  | 2.4±2.3  | 0.9±0.9  | 0.9±1.1   | na         | na          | na          | na                                    |
| Generation R - sample with behavioral observations | na | na | na | na | na | na | na | na | na | na | na | na | na | na | na | Data on SCARED for 48 participants. |
| Generation R - sample with questionnaire data | na | na | na | na | na | na | na | na | na | na | na | na | na | na | na | Data on SCARED for 48 participants. |
| Maryland-PAX                  | na         | na  | na  | 42.2±22.3    | na   | 17.1±27.9 | 6.0±7.1   | 2.9±6.4  | 8.7±13.7 | 2.2±2.3  | na         | na          | na          | na          | Data on SCARED for 48 participants. |
| Maryland-TAX                  | 45.5±13.7  | na  | na  | 48.2±14.6    | na   | 17.1±20.3 | 4.4±4.9   | 4.1±6.8  | 4.8±5.5  | 5.5±18   | na         | na          | na          | na          | Data on SCARED for 48 participants. |
| Nijmegen Longitudinal Study   | na         | na  | na  | na           | na   | 3.9±3.2   | 3.9±3.2   | 3.9±na   | 3.2±na   | 3.0±na   | 0.6±na    | 0.6±na    | na         | na          | Data on PSWQ for 49 participants. |
| Pittsburgh                    | na         | na  | na  | na           | na   | 27.6±32.1 | 8.6±7.4   | 2.7±2.9  | 0.7±0.7  | 3.0±1.0  | na         | na          | na          | na          | CDI: rated by child |
| San Raffaele                  | na         | na  | na  | na           | na   | 21.5±21.8 | 21.8±21.8 | 5.6±5.3  | 2.6±3.7  | 7.1±6.1  | 1.6±1.6   | 1.6±1.7   | na         | na          | Data on SCARED for 48 participants. |
| SDAN                          | 34.9±8.3   | na  | na  | 17.5±9.9     | na   | 21.5±15.1 | 15.1±15.1 | 3.8±3.0  | 4.2±5.1  | 4.5±1.8  | na         | na          | na          | na          | CDI: rated by child |
| Stony Brook Temperament Study | na         | na  | na  | na           | na   | 8.6±3.8   | 2.7±na   | 3.0±na   | 3.0±na   | 0.6±na   | na         | na          | na          | na          | Data on SCARED for 48 participants. |</p>
<table>
<thead>
<tr>
<th>Sample</th>
<th>STAI-trait</th>
<th>ASI</th>
<th>BAI</th>
<th>LSAS - total</th>
<th>PSWQ</th>
<th>BDI</th>
<th>CDI</th>
<th>SCARED-T-P</th>
<th>SCARED-C</th>
<th>SCARED-SC-P</th>
<th>SCARED-SC-C</th>
<th>SCARED-PN-P</th>
<th>SCARED-PN-C</th>
<th>SCARED-GD-P</th>
<th>SCARED-GD-C</th>
<th>SCARED-SH-P</th>
<th>SCARED-SH-C</th>
<th>YSR - total</th>
<th>YSR - internalizing</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTS</td>
<td>28.0 ± 6.3</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>9.6 ± 7.7</td>
<td>17.1 ± 9.7</td>
<td>3.5 ± 3.6</td>
<td>5.5 ± 3.4</td>
<td>1.7 ± 1.8</td>
<td>2.7 ± 2.6</td>
<td>2.8 ± 2.8</td>
<td>4.3 ± 3.5</td>
<td>0.6 ± 0.9</td>
<td>1.3 ± 1.2</td>
<td>na</td>
<td>na</td>
<td>Data availability for psychopathology: n = 93; for SCARED: 35 - 43 participants.</td>
</tr>
<tr>
<td>Vanderbilt children</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>5.3 ± 5.5</td>
<td>11.5 ± 10.6</td>
<td>17.7 ± 13.5</td>
<td>4.4 ± 4.5</td>
<td>5.0 ± 3.6</td>
<td>1.0 ± 1.8</td>
<td>2.8 ± 3.2</td>
<td>4.0 ± 3.4</td>
<td>0.3 ± 0.6</td>
<td>1.3 ± 1.5</td>
<td>na</td>
<td>na</td>
<td>Data availability for questionnaires varies from n = 44 to n = 47</td>
</tr>
<tr>
<td>Vanderbilt young adults</td>
<td>31.8 ± 10.9</td>
<td>13.5</td>
<td>6.4</td>
<td>34.8 ± 10.9</td>
<td>16.5</td>
<td>5.5</td>
<td>6.6</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>Data availability for questionnaires varies from n = 29 to n = 149</td>
</tr>
<tr>
<td>Western University</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>5.0 ± 5.3</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>Data availability for psychopathology: n = 82</td>
</tr>
<tr>
<td>Wisconsin Twin Project - RDoC twin study</td>
<td>Not yet available</td>
<td>na</td>
<td>na</td>
<td>Not yet available</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Not yet available</td>
</tr>
</tbody>
</table>
Supplemental Table 4  

Overview of included independent variables per sample

*Will be included in final submission (i.e. after performing the analyses)*
References Supplemental Materials


26. Bishop G, Spence SH, Mcdonald C. Can Parents and Teachers Provide a Reliable and


41. Rothbart MK. Temperament, development, and personality. *Curr Dir Psychol Sci*.


50. White T, Muetzel RL, El Marroun H, et al. Paediatric population neuroimaging and the


111. Kann SJ, O’Rawe JF, Huang AS, Klein DN, Leung H-C. Preschool negative emotionality predicts activity and connectivity of the fusiform face area and amygdala


Blackford JU, Avery SN, Cowan RL, Shelton RC, Zald DH. Sustained amygdala
response to both novel and newly familiar faces characterizes inhibited temperament.


139. Liu P, Vandermeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Neural


147. Achenbach TM, Rescorla L. *Manual for the ASEBA School-Age Forms & Profiles: An*


155. Moore MM, Planalp EM, Van Hulle CA, Goldsmith H. Pediatric assessment of Research Domain Criteria positive and negative valence systems: Partial genetic mediation of
links to problem behaviors. *J Abnorm Psychol (in Press).*


