NEW RESEARCH

Check for updates

Structural Brain Correlates of Childhood Inhibited Temperament: An ENIGMA-Anxiety Mega-Analysis

Janna Marie Bas-Hoogendam[a,b,c,d,*], PhD, Rachel A. Bernstein[d], BA, Brenda E. Benson[d], PhD, Samuel E.C. Frank[d], BA, Kristin A. Buss[e], PhD, Kelley E. Gunther[e,f], PhD, Koraly Pérez-Edgar[e], PhD, Giovanni A. Salum[g,h,i], PhD, Andrea Jackowski[j], PhD, Rodrigo A. Bressan[j], PhD, André Zugman[d], PhD, Kathryn A. Degnan[k], PhD, Courtney A. Filippi[l], PhD, Nathan A. Fox[f], PhD, Heather A. Henderson[m], PhD, Alva Tang[n], PhD, Selin Zeytinoglu[f,o], PhD, Anita Harrewijn[p], PhD, Manon H.J. Hillegers[q], PhD, MD, Ryan L. Muetzel[q], PhD, Tonya White[d,q], MD, PhD, Marinus H. van IJzendoorn[r,s], PhD, Carl E. Schwartz[t], MD, Julia M. Felicione[t,u], PhD, Kathryn A. DeYoung[f], MA, MS, Alexander J. Shackman[f], PhD, Jason F. Smith[f], PhD, Rachael M. Tillman[f], PhD, Yvonne H.M. van den Berg[v], PhD, Antonius H.N. Cillessen[v], PhD, Karin Roelofs[v], PhD, Anna Tyborowska[v], PhD, Shirley Y. Hill[w], PhD, Marco Battaglia[x,y], MD, Marco Tettamanti[z], PhD, Lea R. Dougherty[f], PhD, Jingwen Jin[aa], PhD, Daniel N. Klein[bb], PhD, Hoi-Chung Leung[bb], PhD, Suzanne N. Avery[cc], PhD, Jennifer Urbano Blackford[cc,dd], PhD, Jacqueline A. Clauss[t,ee], PhD, MD, James M. Bjork[ff], PhD, John M. Hettema[gg], MD, PhD, Ashlee A. Moore[hh], PhD, Roxann Roberson-Nay[ii], PhD, Chelsea Sawyers[ff], PhD, Elizabeth P. Hayden[jj], PhD, Pan Liu[jj,kk], PhD, Matthew R.J. Vandermeer[jj], PhD, H. Hill Goldsmith[ll], PhD, Elizabeth M. Planalp[ll], PhD, Thomas E. Nichols[mm], PhD, Paul M. Thompson[nn], PhD, P. Michiel Westenberg[a,c], PhD, Nic J.A. van der Wee[b,c], PhD, MD, Nynke A. Groenewold[oo], PhD, Dan J. Stein[oo], PhD, MD, Anderson M. Winkler[d,pp], MD, DPhil, Daniel S. Pine[d], MD

Objective: Childhood inhibited temperament (cIT) is associated with an increased risk for developing internalizing psychopathology. Neurobiological characteristics identified by structural magnetic resonance imaging (MRI) may elucidate the neural substrates for cIT, but studies are scarce and often focus on particular regions of interest. Moreover, current findings lack replication. This preregistered analysis from the ENIGMA-Anxiety Working Group examined structural brain characteristics associated with cIT using a comprehensive whole-brain approach.

Method: Temperament assessments (behavioral observations, parent/teacher reports or self-reports on cIT before age 13 years) and MRI data (age at scan, 6-25 years) from international research sites (Europe, North America, South America) were pooled for mega-analysis. Following image processing and quality control, associations between cIT and brain structure were examined in 3,803 participants. Subcortical volumes, cortical thickness, and surface area (main analyses) and detailed subcortical characteristics (eg, subnuclei, subfields, partial volume effects; exploratory analyses) were considered.

Results: In the full sample, cIT showed no relation with brain structure, either as a main effect or in interactions with sex or age. Subgroup analyses (based on cIT assessment type) revealed cIT by sex interactions on mean cortical thickness ($p_{MC\text{-}FWER} = .037$) and thickness of the right superior parietal region ($p_{MC\text{-}FWER} = .029$) in youth with parent/teacher reports on cIT levels. Exploratory analyses revealed findings in the hippocampus, putamen, and caudate, but most did not survive statistical correction for multiple testing.

Conclusion: This mega-analysis found no consistent associations between cIT and regional brain structure, although the role of parietal regions warrants further investigation. Future studies should consider brain function in cIT, preferably using longitudinal designs.

Plain language summary: Inhibited temperament during childhood is a risk factor for the development of anxiety and depression later in life. A preregistered study from the international ENIGMA-Anxiety Working Group investigated whether characteristics of brain structure are associated with the level of childhood inhibited temperament, using brain scans and data on temperamental traits from participants aged 6 to 25 years, which have been previously acquired at research sites worldwide (total sample > 3,800 subjects). Analyses revealed no consistent correlations between brain structure and inhibited temperament.

Study Registration Information: Structural Brain Correlates of Childhood Inhibited Temperament: An ENIGMA-Anxiety Mega-analysis. https://www.jaacap.org/article/S0890-8567(22)00299-4/fulltext

Key words: temperament; magnetic resonance imaging (MRI); childhood; adolescence; anxiety disorders

J Am Acad Child Adolesc Psychiatry 2026;65(4):577-604.

The term "temperament" refers to "a biological bias for particular feelings and actions that first appear during infancy or early childhood and that are sculpted by environments into a large, but still limited, number of personality traits."[1] One of the most rigorously characterized temperament classifications distinguishes infants and young children based on their tendencies to approach or avoid unfamiliar people, objects, and unexpected events, especially in social contexts.[2,3] Some infants explore new toys with enthusiastic curiosity, whereas others react in a more cautious or avoidant way; when meeting unfamiliar people, some toddlers approach them eagerly, whereas others cling to their parents. Individuals with the tendency to avoid the unfamiliar are labeled as manifesting "behavioral inhibition" or "inhibited temperament" (IT).[4,5] IT is a moderately heritable trait that can be measured in multiple species, providing opportunities for translational research.[6-8] In human beings, 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 IT,[2] can be used later in life.

Variations in IT are present on a continuous scale within the population, and research suggests that around 20% of young children are characterized by high IT.[9] Although temperament in unselected samples shows at least moderate continuity over time, these individuals with high levels of childhood IT (cIT) have much higher levels of stability,[3,10-12] Considerable data suggest that temperament predicts personality traits later in life,[13,14] and that high cIT has adverse long-term consequences[15,16]: 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.[13] Multiple studies have shown cIT (especially "social cIT," when compared to "nonsocial cIT"[17]) to be associated with an elevated risk of developing social anxiety.[18-20] These findings have recently been strengthened by the results of a longitudinal twin sample (868 families) revealing that behavioral inhibition robustly predicts social anxiety.[21] More specifically, almost half of all children with elevated and stable cIT will develop social anxiety disorder (SAD) later in life, compared with only 12% of noninhibited children.[15] Furthermore, a twin study showed that cIT was associated with preadolescent social anxiety symptoms; social anxiety shared a substantial proportion of genetic and environmental variance with cIT, providing evidence for early cIT as a potential developmental endophenotype for later social anxiety.[22] In addition, a recently published study reporting on 110,367 children from a population-based pregnancy cohort study in Norway presented an association between early childhood temperament (shyness at age 5 years reported by the mother) and the presence of emotional disorders in adolescence.[23] Taken together, these findings indicate that cIT predicts risk for later psychopathology, especially SAD (large effect size; odds ratio = 5.84, 95% CI = 3.38-10.09, $p < .001$, as reported by Sandstrom *et al.*, 2020),[16,24-28] although it should be noted that not all children with high levels of cIT early in life become anxious adults.[29-32]

Several neuroimaging studies have examined neurobiological correlates of cIT.[30] Such research is important, because brain characteristics—including brain structure, activity, and connectivity—may mediate the cIT-related risk for poor outcomes.[33] Some studies have used a cross-sectional approach, including children and early adolescents with high IT[34-36] or investigating young adults who displayed inhibited behavior as a child (determined retrospectively) and at the time of MRI assessment.[37-42] Other studies had a longitudinal design, in which infant temperament was assessed early in life, whereas neuroimaging was performed during late childhood, adolescence, or young adulthood.[43-56] These previous studies have connected cIT to structure and function of brain networks involved in emotion perception, experience, and regulation.[2] These brain networks involve the dorsal (caudal) and ventral (rostral) anterior cingulate cortex (ACC), insula, amygdala, dorsolateral and medial prefrontal cortex (PFC), orbitofrontal cortex (OFC) and striatum,[2,33] all of which have also been implicated in familial risk for SAD.[57-61] In addition, translational work has indicated involvement of the hippocampus.[8,62-64]

Despite this progress, the few available studies on the neuroanatomical correlates of cIT are often restricted to specific regions of interest, and cortical surface area and cortical thickness have been examined in only 1 study, with an exploratory approach.[55] Furthermore, most findings with respect to brain structure are unique to a specific sample[33] (Table 1[34-37,39-42,44,45,51-55,65-79]), 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,[80] we aim to extend previous work by examining brain structure associated with cIT in a large dataset, assembling data acquired at research centers worldwide (18 samples, n = 4,810 before image processing and quality control). Compared to the individual studies (on relatively small [sample sizes ranging from 23 to 130] [Table 1] and homogeneous samples), this new study

TABLE 1 Previous Neuroimaging Findings Related to Childhood Inhibited Temperament, From Samples Included in the Present Mega-Analysis

	Publications on cIT and brain structure				Publications on cIT and brain function/connectivity	
Sample	**Sample size in present mega-analysis[a]**	**Sample size in previous work**	**Approach**	**Findings**	**Sample size in previous work**	**Findings**
BRAINS study	n = 130	n = 130[35]	ROI-based: bilateral amygdala, posterior insula, anterior insula, ACC, OFC, and vlPFC, as well as comparison occipital ROIs (inferior and middle occipital gyri).	Left posterior insula volume was positively correlated with total cIT score (no effect size reported).	n = 42[34] n = 67[36] n = 56[65]	Increased connectivity between putative "salience processing" regions (amygdala and insula) and putative internal processing regions (vmPFC). Emotion-face masked dot-probe task: non-cIT children displayed greater activation vs cIT children in several regions in response to threat faces vs neutral faces, including striatum and prefrontal and temporal lobes. Dot-probe task (analysis in 3 prefrontal ROIs based on task activation): greater activation in the right dlPFC cluster in children with high cIT; no differences in amygdala, vlPFC, and mPFC ROIs.
Brazilian High Risk Cohort	n = 678	None	—	—	None	—
Cohort 3/4	n = 95	n = 53 [52]	ROI analysis of brain structure in adulthood. Cortical thickness: middle anterior part of cingulate gyrus and sulcus (dACC), short insular gyrus, subcallosal gyrus, and left orbitofrontal and right ventromedial ROIs. Bilateral volumes of the amygdala and hippocampus. Vertex-wise exploratory analyses in prefrontal cortex.	Early cIT predicted thinner cortex in the dACC (large effect size, partial η^2 = 0.26) and subcallosal gyrus (small effect size, partial η^2 = 0.10, uncorrected for multiple comparisons); other regions no relationship with cIT.	n = 32[66] n = 32[44] n = 39[67] n = 44[68] n = 32[69] n = 35[70] n = 35[71] n = 27[45] n = 38[72] n = 50[73] n = 83[74]	Reward-contingency task: Adolescents characterized by an enduring pattern of cIT demonstrated enhanced sensitivity of the reward-related neural system. Monetary Incentive Delay task: greater striatal activation to incentives in adolescents with cIT; no significant interactions between early inhibited temperament and activitiy in the bilateral nucleus accumbens with changes in anxiety/anxiety levels at age 26 y. Social evaluation task: striatal sensitivity in adolescents varied as a function of temperament, the peer delivering the feedback, and feedback valence. Attention-bias task: young adults with cIT exhibited greater strength in threat-

(*continued*)

TABLE 1 Continued

Publications on cIT and brain structure					Publications on cIT and brain function/connectivity	
Sample	**Sample size in present mega-analysis[a]**	**Sample size in previous work**	**Approach**	**Findings**	**Sample size in previous work**	**Findings**
						related connectivity, differences manifested in connections between the amygdala and dlPFC and anterior insula. Emotional conflict task: adults with cIT exhibited greater dorsomedial prefrontal cortex activity during conflict detection and greater putamen activity during conflict adaptation. Implicit emotion-processing task: in the presence of fearful faces, adults with cIT exhibited greater activity in cingulate cortex, dlPFC, and striatum for high attention control trials compared with low attention control trials. The opposite pattern emerged in the presence of happy faces. Face processing task: adolescents with cIT showed exaggerated amygdala response during subjective fear ratings and deactivation during passive viewing, across all emotion faces. In addition, the cIT group showed an abnormally high amygdala response to a task condition marked by novelty and uncertainty. Connectivity: cIT was associated with differences in intrinsic functional connectivity in adulthood, between 3 amygdala subdivisions and prefrontal cortex, striatum, anterior insula, and cerebellum. Extinction recall task: cIT was associated with greater activation in subgenual ACC in response to cues signaling safety.

(*continued*)

TABLE 1 Continued

Publications on cIT and brain structure					Publications on cIT and brain function/connectivity	
Sample	**Sample size in present mega-analysis[a]**	**Sample size in previous work**	**Approach**	**Findings**	**Sample size in previous work**	**Findings**
Generation R-behavioral observations	n = 584	None	—	—	None	—
Generation R-questionnaire data	n =1,982	None	—	—	None	—
Maryland-PAX	n = 220	None	—	—	None	—
Maryland-TAX	n = 53	None	—	—	None	—
Nijmegen Longitudinal Study	n = 71	None	—	—	None	—
Pittsburgh	n = 15	*n* = 23 [51]	ROI-based approach with manually traced amygdala and OFC, followed by whole-brain VBM.	cIT related to greater right OFC volume and greater total amygdala volume in adolescence (no effect size reported).	None	—
San Raffaele	n = 20	None	—	—	None	—
SDAN	n = 55	None	—	—	None	—
Stony Brook Temperament Study	n = 74	None	—	—	None	
TOTS	n = 96	n = 75[55]	ROI-based: bilateral amygdala volume, with negative reactivity at 4 mo as predictor. Exploratory analysis: vertex-wise whole-brain cortical thickness and cortical surface area.	In children between 10 and 12 y of age, left amygdala volume increased more slowly in those with cIT (no effect size reported).	n = 43[75] n = 87[54] n = 53[53] n = 53[76] n = 55[77]	Extinction recall task: cIT predicts a distinct pattern of hemodynamic—autonomic covariation when recalling extinguished threat and safety cues; interactions present in anterior insular cortex, anterior subdivision of the medial cingulate cortex, and dlPFC. Connectivity: in children with a history of high cIT, anxiety symptoms became more negatively correlated with dlPFC—amygdala connectivity when processing salient, proximal threats; the opposite developmental pattern was observed in low-cIT children. Virtual school paradigm: in adolescents with preadolescent

(*continued*)

TABLE 1 Continued

Publications on cIT and brain structure					Publications on cIT and brain function/connectivity	
Sample	**Sample size in present mega-analysis[a]**	**Sample size in previous work**	**Approach**	**Findings**	**Sample size in previous work**	**Findings**
						social anxiety, greater cIT was associated with enhanced bilateral insula engagement while anticipating unpredictable-vs-nice social evaluation. High cIT predicted greater activity in dorsal ACC and bilateral insula. High cIT was associated with negative functional connectivity between insula and vmPFC, and negative evaluation was associated with increased amygdala activity (during feedback from unpredictable peers). Modified flanker task: significant cIT-by-anxiety-by-error condition interactions in cuneus, fusiform gyrus, lingual gyrus, orbitofrontal gyrus, and middle occipital gyrus.
Vanderbilt—children	n = 55	None	—	—	n = 37[78]	During anticipation and viewing of threat stimuli and social stimuli: high cIT is related to widespread alterations in prefrontal cortex function and connectivity
Vanderbilt—young adults	n = 150	n = 84[42]	ROI-based approach focused on amygdala, with 3 complementary methods: manual segmentation, surface mapping, and VBM.	Inhibited adults had significantly larger volume in right amygdala, with a similar trend for left amygdala (manual segmentation), regions of increased convexity located primarily in basolateral and lateral subnuclei (surface mapping), and greater gray matter volume in	n = 20[41] n = 33[39] n = 39[40] n = 34[37] n = 32[79]	Faces task: cIT participants had faster amygdala responses to novel compared with familiar faces, and both longer and greater amygdala response to all faces; cIT young adults had increased BOLD response in amygdala when viewing both novel and recently familiarized faces (so sustained amygdala activation). In individuals with an inhibited temperament, the amygdala and hippocampus failed to habituate across repeated presentations of faces. Young adults with cIT: greater

(*continued*)

TABLE 1 Continued

Publications on cIT and brain structure					Publications on cIT and brain function/connectivity	
Sample	**Sample size in present mega-analysis[a]**	**Sample size in previous work**	**Approach**	**Findings**	**Sample size in previous work**	**Findings**
				both left and right amygdalae (VBM) (no effect size reported). Furthermore, inhibited adults had larger caudate volume (left; no effect size reported).		activation of a prefrontal network when anticipating viewing fear faces (but no functional differences in amygdala), and more negative connectivity between the rostral ACC and the bilateral amygdala. Higher social fearfulness was associated with slower habituation across regions of the social brain, including the hippocampus, amygdala, vmPFC, medial OFC, fusiform face area, primary visual cortex, and extrastriate visual cortex.
Virginia Commonwealth University Juvenile Anxiety Study (VCU-JAS)	n = 129	None	—	—	None	—
Western University	n = 87	None	—	—	None	—
Wisconsin Twin Project-RDoC twin study	n = 316	None	—	—	None	—

Note: ACC = anterior cingulate cortex; BOLD = blood oxygen level–dependent; cIT = childhood inhibited temperament; dACC = dorsal anterior cingulate cortex; dlPFC = dorsolateral prefrontal cortex; mPFC = medial prefrontal cortex; OFC = orbitofrontal cortex; ROI = region of interest; VBM = voxel-based morphometry; vlPFC = ventrolateral prefrontal cortex; vmPFC = ventromedial prefrontal cortex.

[a]*Sample size before image processing and quality control.*

is better powered because of 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-positives that could emerge from smaller-sampled studies.[81] Such work has the potential to establish reproducible anatomical correlates, and could inform the development of mechanistic studies and intervention research with clinical relevance.[82]

We performed a mega-analysis of T_1-weighted anatomical MRI scans of the human brain with a whole-brain approach (regional and vertex-wise) on the total dataset, and considered the relationship between cIT and 3 distinct neuroanatomical metrics: volumes of subcortical structures, cortical thickness, and cortical surface area. As cortical thickness and cortical surface area are genetically and phenotypically independent, it is important to investigate them separately.[83] In addition to the main analyses, sensitivity analyses were 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 and adolescence.[2] Importantly, the association between cIT and later (social) anxiety has been established for all 3 types of assessments (for some examples, see behavioral observations,[18,19,84-86] parental reports,[22,26,87,88] and self-report questionnaires[89,90]; see also a recent meta-analysis reporting no significant effect of the method of measuring inhibited temperament on the cIT-anxiety association[16]). A fourth sensitivity analysis included only samples in which temperament was assessed during childhood (not retrospectively).

We expanded previous work by performing exploratory analyses on the relationship between cIT and amygdalar subnuclei, thalamic subnuclei, and hippocampal subfields,[91-93] the amount of gray matter inside each subcortical structure,[94] and the volumes of additional subcortical limbic structures that were more recently included in the FreeSurfer software package.[95]

We expected to corroborate findings in brain circuits found previously (involved in processing fear, reward, and emotion regulation),[2,33] with small-to-medium effect sizes.[81,96-98] That is, based on earlier work on inhibited temperament (Table 1), we hypothesized that structural characteristics of the amygdala (larger volume[42,51]), caudate (larger volume[42]), caudal and rostral ACC (thinner cortex[52]), insula (increased cortical thickness[35]), and OFC (increased cortical thickness, especially in right OFC[51]) are neural substrates of cIT. Additional hypotheses, based on an endophenotype study in socially anxious families that revealed heritable brain alterations related to social anxiety,[57] were the following: we expected that cIT is associated with increased volume of the putamen, decreased cortical thickness of the superior temporal gyrus, increased cortical thickness of the transverse temporal gyrus, and decreased surface area of the fusiform gyrus. In addition, we expected to find decreased volumes of the hippocampus.[8,62,99-101] Furthermore, the whole-brain approach of the proposed study enabled us to explore and to potentially discover novel substrates for the risk-conferring cIT phenotype.

To the best of our understanding, 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 in prior work.[33] We anticipated the large sample size to provide precise and relatively unbiased estimates of true effect sizes for multiple indices of cIT, providing a solid foundation to guide future research by individual investigators.

Furthermore, 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[102]: previous studies within the ENIGMA-Consortium have resulted in important insights in the neurobiology of psychiatric conditions,[103-107] and mega-analyses within the ENIGMA-Anxiety Working Group have revealed novel brain characteristics related to SAD,[108,109] specific phobia,[110] and anxiety in youth.[111] These insights reflect the advantages of large-scale data analyses for testing reproducibility and robustness of neuroimaging findings.[103] We expected the current project to provide similar insights concerning an important risk factor for social anxiety, 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 aimed to contribute to a reduction of the potential publication bias in the field, and to advance a more complete and reliable scientific record on this topic.[112]

METHOD

Study Design and Setting

This preregistered study concerns a mega-analysis of T_1-weighted anatomical MRI scans of the human brain that have been previously acquired at research sites in Europe, North America, and South America (for scan characteristics, we refer to Table S2, available online). The project is part of the ENIGMA-Anxiety Working Group,[80] and analyses took place at the National Institute of Mental Health

(NIMH; Bethesda, MD). See Table S1, available online, for Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Participants

Individual participant data from studies where participants underwent MRI scanning between 6 and 25 years of age (inclusion criterion 1) and possessed at least 1 measurement of childhood inhibited temperament (cIT; inclusion criterion 2) were considered for inclusion. These inclusion criteria are based on the course of normative brain development into young adulthood[113-115] and the emergence of internalizing psychopathology during adolescence.[116,117] Regardless of age at scan, all participants were required to have data on cIT (childhood defined as age $\leq$ 12 years). These temperament assessments should include measures of the tendency to withdraw from novel stimuli or to avoid unknown people, indices of fear toward novelty, and/or scores of social reticence. To make optimal use of the available data, various methods to acquire this information were allowed: we included behavioral observations in childhood, parent or teacher reports, self-report questionnaires on current temperament (children or young adolescents), and self-report questionnaires on retrospective temperament (young adults). Several studies acquired information using multiple methods.

Studies varied in their designs, with temperament assessments performed at or around the time of scan ("cross-sectional") or preceding the MRI scan ("longitudinal"). For studies in which MRI scans were acquired at multiple time points, we selected the scan closest to the time point of the temperament assessment. (Internalizing) psychopathology was not an exclusion criterion, to allow for investigation of the full spectrum of cIT, and was included as a descriptive variable when available. An overview of the temperament measures acquired in each sample, as well as a description of the design of the datasets included in this analysis, are provided in Table 2[118-174] and in Supplement 1, available online.

Results on structural brain characteristics of cIT have been reported previously for several samples included in the present mega-analysis[35,42,51,52,55]; however, only 1 study investigated the neural substrates of cIT using a whole-brain vertex-wise approach.[55]

Ethics

The individual research protocols were approved by local institutional review boards and ethics committees. All adult participants and parents of participants younger than 18 years of age provided written informed consent at their local research site. Principal investigators from the individual research sites signed a memorandum of understanding, which included regulations about data use, participant deidentification, data transfer methods, data ownership, and confidentiality and security practices.[80,175] Each site also obtained approval from their local officials to share data.

Variables

Independent Variable: Data on Childhood Inhibited Temperament (cIT). As summarized in Table 2, studies varied in the way in which cIT was assessed. To optimally use all of the available information, we used a continuous approach to investigate the relationship between measures of cIT (predictor) and structural brain characteristics, based on the sample-specific temperament measures as provided by the participating sites (see Statistical Analyses).

Dependent Variable: Subcortical Volumes, Cortical Thickness, and Cortical Surface Area Derived From Structural MRI Data

Ten ENIGMA-Anxiety sites sent individual participant structural MRI data to the corresponding author and the research group at NIMH between January 2021 and December 2021. In addition, structural MRI data from the Wisconsin Twin project[171,172] were downloaded from the NIMH data archive (September 2021). In September 2022, data from a new ENIGMA-Anxiety site (Virginia Commonwealth University) were added to the dataset and shared with the corresponding author and the research group at NIMH. Because of data-sharing restrictions, data from the Generation R study could not be shared internationally; therefore, analyses of these data took place locally in Rotterdam, the Netherlands, and group-level outcomes were merged with the results obtained at NIMH.

Additional Descriptive Data

Research sites were asked to provide information with respect to variables of interest: namely, demographic information (age, sex, IQ, socioeconomic status [SES], ancestry), information from clinical interviews concerning anxiety (generalized anxiety disorder [GAD], panic disorder [PD], social anxiety disorder [SAD], specific phobia [SP], other anxiety disorders) and other psychiatric disorders (major depressive disorder [MDD], obsessive compulsive disorder [OCD], posttraumatic stress disorder [PTSD], substance use dependence [SUD], other psychiatric disorders), psychotropic medication use at the time of scan, and several questionnaires on psychopathology (see Supplement 2, available online). Availability of these variables varied per sample (Table S3, available online).

TABLE 2 Characteristics of Samples After Image Processing and Quality Control

Sample (location)	Type of sample	n (n female) with MRI and cIT data	Design[a]	Age MRI scan (range; mean ± SD)	Age cIT (range; mean ± SD)	Measure of cIT (range; mean ± SD)	Sub-group[b]	SES	Ancestry[c]	IQ MRI scan (range; mean ± SD)	Notes
BRAINS study (Pennsylvania State University, University Park, Pennsylvania) [34,36,65,118-124]	Oversampled for high/low cIT.	129 (72)	C	9.2-13.2 y (10.8 ± 1.0)	9.2-13.2 y (10.8 ± 1.0)	BIQ, parent rated; 34-165 (95.3 ± 32.6); cross-sectional measure.	2	NA	African American (2.8%), Asian/Pacific Islander (2.8%), Hispanic (2.8%), Mixed Race (3.7%), White (Non-Hispanic; 88.1%).	71-149 (112.2 ± 14.4)	IQ data available for 124 participants; data on ancestry available for 109 participants.
Brazilian High Risk Cohort (National Institute of Developmental Psychiatry (INPD), São Paulo, Brazil)[125]	Community sample and a high-risk sample of children with increased familial risk for mental disorders.	502 (233)	C	5.8-13.0 y (9.8 ± 1.7)	5.8-13.0 y (9.8 ± 1.7)	EATQ-R-shyness scale; 1.0-5.0 (2.7 ± 1.1); cross-sectional measure.	3	Low (10.3%), middle (76.0%) and high (13.7%).	Asian (0.4%), Between White and Black (Brown; 32.0%), Black (12.4%), Indigenous (0.3%), White (54.9%).	57-152 (102.8 ± 16.4)	IQ data available for 502 participants.
Cohort 3/4 (University of Maryland, College Park Maryland,[44,45,48,52,66-71,126]	Community sample: prospective longitudinal study on infants thought likely to display behavioral inhibition later in infancy and early childhood.	88 (50)	L	13.3-21. 1 y (18.0 ± 1.9)	Around 24 mo (no data at individual level).	Standard laboratory observations at age 2: composite score of stranger, robot, tunnel episodes; −1.3 −1.2 (−0.04 ± 0.6); significant correlation with scores obtained at age 14 mo ($r = 0.3$).	1	Parents of the infants were in the middle to upper-middle class; 61.5% of mothers held college degrees.	Predominately Caucasian (98% White).	83-137 (114.3 ± 10.4)	IQ data available for 84 participants.
Generation R —behavioral observations (Erasmus University Medical Center, Rotterdam, the Netherlands) [127-130]	Community sample.	498 (248)	L	8.7-12.0 y (10.2 ± 0.6)	34.7-44.2 mo (37.4 ± 1.4)	Standard laboratory observations: stranger approach and jumping spider episode from the Lab-TAB; −1.16 to 1.36 (−0.01 ± 0.37); cross-sectional measure.	1	Educational level mother at 5 y: 2.8% primary, 31.2% secondary, 52.3% higher; (missing: 13.8%).	55.8% Dutch, 8.1% Non-Dutch Western, 33.9% Non-Dutch non-Western, 2.2% missing.	No IQ scores at (around) time of scan.	
Generation R-questionnaire data (Erasmus University Medical Center, Rotterdam, the Netherlands) [127-129]	Community sample.	1,604 (833)	L	8.6-12.0 y (10.0 ± 0.5)	4.5-11.8 mo (6.7 ± 1.1)	Infant Behavior Questionnaire —Revised (IBQ-r) —fear subscale; maternal report; 0.0-1.8 (0.4 ± 0.3); cross-sectional measure.	2	Educational level mother at 5 y: 2.8% primary, 31.2% secondary, 52.3% higher; (missing: 13.8%).	55.8% Dutch, 8.1% Non-Dutch Western, 33.9% Non-Dutch non-Western, 2.2% missing.	No IQ scores at (around) time of scan.	Age of IBQ-R assessment was missing for 271 children.

(*continued*)

TABLE 2 Continued

Sample (location)	Type of sample	n (n female) with MRI and cIT data	Design[a]	Age MRI scan (range; mean ± SD)	Age cIT (range; mean ± SD)	Measure of cIT (range; mean ± SD)	Sub-group[b]	SES	Ancestry[c]	IQ MRI scan (range; mean ± SD)	Notes
Maryland-PAX (University of Maryland, College Park, Maryland)[131-134]	30-m Longitudinal study on a sample of first-year university students enriched for internalizing risk.	139 (81)	C	18-19 y (18.2 ± 0.4)	Retrospective: remembered inhibited behaviors in childhood	RMBI; 2-31 (15.0 ± 6.7); measures repeated at 3 follow-ups, with very good within-subject stability (from baseline to 6 mo, from baseline to 24 mo, and from baseline to 30 mo follow-up: all $r > 0.76$, $p < .001$).	3	NA	African American (8.6%), Asian (17.3%), Hispanic/Latino (4.1%), Multiracial/other (9.0%) White (61%).	NA	
Maryland-TAX (University of Maryland, College Park, Maryland)[135]	Cross-sectional community sample enriched for elevated social anxiety symptoms.	53 (28)	C	13-17 y (15.0 ± 1.2)	Retrospective: remembered inhibited behaviors in childhood.	RSRI-adolescent rated 1. 0-3.6 (2.3 ± 0.5); cross-sectional measure.	3	NA	African American (27.8%), Asian (5.6%), Hispanic (9.2%), Multiracial/other (7.4%), White (50%).	NA	
Nijmegen Longitudinal Study (Radboud University, Nijmegen, the Netherlands)[136-138]	Longitudinal community sample.	68 (31)	L	17 y	1.20-1.28 y (1.24 ± 0.02)	Behavioral observations at 15 mo of age: 6-17 (9.5 ± 2.6); cross-sectional measure.	1	NA	NA	NA	
Pittsburgh (University of Pittsburgh School of Medicine, Pittsburgh, USA)[51,139-141]	High and low-risk (control) children/adolescents from ongoing family studies.	15 (3)	L	19.2-24. 8 y (21.5 ± 1.7)	4.1-6.4 y (5.1 ± 0.7)	Laboratory observations during peer play; sum score of amount of time staring at the other child, amount of time spent proximal to the parent, and latency to speak; 14.0-951.3 (213.5 ± 260.5); average over 3 sessions that were separated by no less than 1 wk and no more than 2 mo. Lack of session effect indicates stability of behavioral measures; cf. Hill *et al.*[139]	1	Majority of children from parents with professional, semiprofessional, and skilled occupation.	"All children were from white families."	88-124 (105.6 ± 12.5)	IQ data available for 11 participants.

(*continued*)

TABLE 2 Continued

Sample (location)	Type of sample	n (n female) with MRI and cIT data	Design[a]	Age MRI scan (range; mean ± SD)	Age cIT (range; mean ± SD)	Measure of cIT (range; mean ± SD)	Sub-group[b]	SES	Ancestry[c]	IQ MRI scan (range; mean ± SD)	Notes
San Raffaele (Vita-Salute San Raffaele University and San Raffaele Scientific Institute, Milan, Italy) [142-145]	Community sample.	20 (8)	L	13-16 y (14.8 ± 1.1)	Around age 7 y (no data at individual level) [143]	Empirical composite index of multiple scales with moderate to high cross-correlations [143]; 0-23 (8.8 ± 6.9); cross-sectional measure.	2	Majority high SES (low/middle/high: 11%/33%/55%).	"All children were White and of Italian ancestry."	NA	
SDAN (National Institute of Mental Health, Bethesda, Maryland)[146-148]	Treatment- seeking children and control group of healthy volunteers.	41 (20)	C	7.3-14.6 y (10.4 ± 1.8)	8.1-12.8 y (10.5 ± 1.6)	BIQ-child rated; 47-171 (114.5 ± 28.6); cross-sectional measure.	3	Range 20-82, mean 30.9 (data for n = 36).	5.5% Asian, 10.9% Black or African American, 16.4% multiple, 5.5% unknown, 61.8% White.	78-145 (113.5 ± 14.9).	IQ data available for 37 participants.
Stony Brook Temperament Study (Stony Brook University, Stony Brook, New York)[149-157]	Community sample; MRI subsample oversampled for youth with temperamental high negative emotionality, low positive emotionality, and high behavioral inhibition at age 3 y.	74 (31)	L	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 (log-transformed sum-score); 0.2-1.4 (0.7 ± 0.2); Lab-TAB was repeated at age 6 (2 tasks), small but significant within-subject correlations.[150]	1	Majority of sample middle-class based on the Hollingshead Four Factor Index of Social Status. Majority of parents were married (94.6%) and well educated (67.6% of families had at least 1 parent who graduated from college).	"Majority of the MRI subsample was White and Non-Hispanic (77.0%)."	NA	
TOTS (University of Maryland, College Park, Maryland) [52-55,77,158-160]	Longitudinally followed sample of children selected at 4 mo of age based on their behavior in the laboratory.	27 (15)	L	9.1-19.5 y (12.4 ± 3.3)	1.9-2.7 y (2.2 ± 0.2)	Standard laboratory observations at age 2 y (composite score of stranger, robot, tunnel episodes); −1.0 to 1.2 (0.0 ± 0.5); tasks were repeated at age 3 y, significant within-subject correlations (r = 0.3).	1	For the majority of the children, the mother graduated from graduate school (34.4%) or college (44.8%).	14.6% African American, 65.6% Caucasian, 5.2% Hispanic, 14.6% other/mixed.	78-134 (111.0 ± 15.8)	IQ scores available for 24 participants.
Vanderbilt—children (Vanderbilt University Medical Center, Nashville, Tennessee) [78]	Study with extreme discordant phenotypes approach: inhibited and uninhibited children at the extreme ends.	55 (33)	C	8-12 y (9.3 ± 1.1)	8-12 y (9.3 ± 1.1)	BIQ-child rated; 33-181 (111.4 ± 31.7); cross-sectional measure.	3	NA	NA	93-136 (115.7 ± 10.3)	IQ scores available for 47 participants.

(*continued*)

TABLE 2 Continued

Sample (location)	Type of sample	n (n female) with MRI and cIT data	Design[a]	Age MRI scan (range; mean ± SD)	Age cIT (range; mean ± SD)	Measure of cIT (range; mean ± SD)	Sub-group[b]	SES	Ancestry[c]	IQ MRI scan (range; mean ± SD)	Notes
Vanderbilt—young adults (Vanderbilt University Medical Center, Nashville, Tennessee) [39,41,42,79,161]	Study with extreme discordant phenotypes approach: inhibited and uninhibited young adults at the extreme ends.	145 (79)	C	18-25 y (21.8 ± 2.0)	Retrospective: remembered inhibited behaviors in childhood.	RSRI; 1.1-4. 4 (2.3 ± 0.9); cross-sectional measure.	3	NA	NA	93-141 (112.2 ± 13.8)	IQ scores available for 14 participants.
Virginia Commonwealth University Juvenile Anxiety Study (VCU-JAS) (VCU, Richmond, Virginia) [22,162-164]	Twin study of pre-adolescents, using an epidemiological sampling design unselected for any particular outcome phenotypes.	126 (75)	C	9.2-14.3 y (11.3 ± 1.4)	Retrospective: remembered inhibited behaviors in early childhood (2-6 y).	Retrospective BIQ —parent rated, range 30-186 (88.9 ± 30.0).	2	NA	Caucasian	76-139 (111.9 ± 12.9)	IQ scores available for 108 participants.
Western University (The Brain and Mind Institute, Western University, London, Ontario, Canada)[165-170]	Children selected based on presence/absence maternal depression.	82 (36)	L	9.2-12.4 y (11.1 ± 0.7)	3.0-4.0 y (3.4 ± 0.3)	Lab-TAB around age 3 y: risk room, stranger approach and jumping spider; log-transformed composite scores; −0.7-1.1 (0.0 ± 0.4); stable over time as indicated by Lab-TAB at age 5 y.[166]	1	Majority (50.6%) of the families were middle class with an annual family income of $40,000-$100,000 CAD.	"Majority of the sample was White (97%)."	82-147 (113. 0 ± 14.0)	IQ based on scores on the Peabody Picture Vocabulary Test (data for 81 participants).
Wisconsin Twin Project-RdoC twin study (University of Wisconsin—Madison, Madison, Wisconsin) [21,171-174]	Longitudinally followed samples of twins, recruited from statewide birth records for birth cohorts 1989—2004.	152 (93)	L	15.1-23.9 y (17.9 ± 1.8)	6.5-9.1 y (7.5 ± 0.5)	Ratings on Approach and Shyness from a home visit, and scores from videotaped reactions to the "Conversation with a Stranger" episode of Lab-TAB; 0. 7-5.4 (2.9 ±1.2) cross-sectional measure.	1	Median family income above $50,000; majority of parents completed college. [174]	"Majority of the sample was White/ Caucasian."[174]	NA	
Total		3,818 (1,969)									

Note: *BI = behavioral inhibition; BIQ = Behavioral Inhibition Questionnaire; CBQ = Child Behavior Questionnaire; cIT = childhood inhibited temperament; EATQ-R = Early Adolescent Temperament Questionnaire; Lab-TAB = Laboratory Temperament Assessment Battery; MRI = magnetic resonance imaging; NA = not available; RMBI = Retrospective Measure of Behavioral Inhibition; RSRI = Retrospective Self-report of Inhibition; SES = socioeconomic status.*

[a]*With respect to timepoint temperament assessment and MRI scan, for data used in this study: cross-sectional (C) or longitudinal (L).*

[b]*Subgroups for sensitivity analysis: 1: (early-life) behavioral observations; 2: parental/teacher reports during childhood; 3: self-report measures acquired during late childhood and adolescence.*

[c]*Information on ancestry was derived from the original papers and is thus reflective of the concept of "ancestry" as it was defined by the individual research sites. No attempt was made to acquire new information for the current work.*

Study Size and Bias

We aimed to assemble the largest dataset possible, consisting of previously acquired samples with structural MRI data of the human brain from participants 6 to 25 years of age (inclusion criterion 1) and a characterization of cIT for each participant (inclusion criterion 2). Therefore, we reached out to all members within the ENIGMA-Anxiety Working Group, based on information about their samples that they had previously provided to the Working Group (see the description of this procedure in Bas-Hoogendam *et al.*[80]). In addition, we contacted research groups based on literature searches and personal contacts of the coordinators (JMBH and DSP). This secured unpublished data, which minimizes the risk of publication biases. The initial sample consisted of 5,098 MRI scans (Table 3), whereas temperament data were available for 4,810 participants, leading to a pooled dataset of 4,810 MRI scans for further processing (Table 3).

Image Processing

Image processing took place following the procedure described previously.[97] To start, structural MRI scans were organized according to the Brain Imaging Data Structure (BIDS)[176] specification, and MRI Quality Control (MRIQC)[177] was used for initial quality checking (QC). Next, images were processed with FreeSurfer software (version 7.0.0)[178] to obtain volumes of subcortical structures and regional measures of cortical surface area (CSA) and cortical thickness (CT). In addition, advanced methods were used to obtain volumes of amygdalar subnuclei, thalamic subnuclei, and hippocampal subfields,[91-93] to compute the amount of gray matter inside each subcortical structure (partial volume effect [PVE]; see the description in the Supplemental section of Abend *et al.*[94]), and to acquire volumes of subcortical limbic structures that were recently included in the FreeSurfer software package (hypothalamus, mammillary bodies (part of hypothalamus), basal forebrain, septal nuclei, nucleus accumbens, and fornix).[95] These outcome measures were investigated in preregistered exploratory analyses (reported in Supplement 2, available online), and the results of these analyses can advance the field, as they have the potential to enable the generation of new hypotheses.

Data were visually checked for gross over- or underestimation of the white/pial surfaces by 2 independent raters, and additional semi-automated QC was performed by using the ratio between the Euler characteristic and the number of vertices in the surfaces before topology correction, defining site-specific thresholds using a receiver operating characteristic curve constructed using the results of the visual inspection.[175] After registration to a common space ("FsAverage"), measurements of CT and CSA were resampled to an icosahedron recursively subdivided 4 times ("fsaverage4"), which was used as a common grid for interpolation.[179] Data availability after processing and quality control is shown in Table 3.

Statistical Analyses

cIT as Continuous Predictor. Studies included in this analysis varied in the way in which cIT was assessed (Table 2), and data required for between-sample harmonization of cIT measures (ie, data on multiple measures acquired within the same participants) are lacking. Here, we used a dimensional approach and used continuous scores of cIT as predictor in the analyses. To allow for joint inference across all samples while accounting for the variability in cIT measures between samples, a variation of the method of nonparametric combination (NPC)[180,181] was used. Specifically, we allowed the relationship between cIT and brain structures to be modeled as variable across sites, and then nonparametrically combined the resulting statistics across sites using the Stouffer combining function. The modification over the original NPC is that we allowed each subject to make 1 contribution to the joint result across samples, as opposed to each subject contributing with multiple metrics within a given sample. Models included cIT (continuous variable), sex, age, age squared, and their interactions (all allowed to vary among sites), as well as scanner (Table 4). It should be noted that in each sample, higher scores reflect a higher level of cIT. We tested 6 contrasts: main effect of cIT (positive and negative), 2-way interaction between cIT and sex (positive and negative), 2-way interaction between cIT and age, and 3-way interaction between cIT, age and sex (Table 4).

Linear and quadratic effects of age were combined using an F test. As the design is fully separable, effects across sites could be combined nonparametrically, thus allowing a joint test that benefits from the various cIT measures in their native scales, eschews the need for explicit data harmonization across the multiple samples, and, together with inference through permutation testing, allows multiple-testing correction with minimal assumptions.[182] Moreover, the use of NPC with a separable design allowed data analysis to occur locally for sites that were unable to share data (as was the case with the Generation R study); all that was needed was the group-level test statistics, as well as the same statistics after permutation of the data. It should be noted, however, that these advantages come with a few drawbacks: effect sizes are not computed, as NPC uses a combination of statistics based on different measures of cIT, each with a different scale; furthermore, interaction effects are hard to explore, given the restrictions on data sharing from some sites.

TABLE 3 Data Inclusion for Each Sample; Sample Size per (Sub)-Analysis

Total remaining (n) following pre-processing and quality control

Sample	Initial no. MRI scans (n)[a]	Initial no. with cIT data (n)[a]	Main analysis (n)	Subgroup analysis 1 (n)[b]	Subgroup analysis 2 (n)[b]	Subgroup analysis 3 (n)[b]	Subgroup analysis 4 (n)[b]
BRAINS study	131	130	129		129		129
Brazilian High Risk Cohort	688	678	502			502	502
Cohort 3/4	121	95	88	88			88
Generation R-sample with behavioral observation	584[a]	584[a]	498[a]	498[a]			498[a]
Generation R-sample with questionnaire data	1,982[a]	1,982[a]	1,604[a]		1,604[a]		1,604[a]
Maryland-PAX	220	220	139			139	
Maryland-TAX	54	53	53			53	
Nijmegen Longitudinal Study	71	71	68	68			68
Pittsburgh	64	15	0[c]				
San Raffaele	20	20	20		20		20
SDAN	55	55	41			41	41
Stony Brook Temperament Study	74	74	74	74			74
TOTS	129	96	27	27			27
Vanderbilt—children	55	55	55			55	55
Vanderbilt—young adults	150	150	145			145	
VCU-JAS	133	129	126		126		
Western University	87	87	82	82			82
Wisconsin Twin Project	480	316	152	152			152
Total	5,098	4,810	3,803	989	1,879	935	3,340

[a]*Data with superscript letter "a" were processed locally in Rotterdam, the Netherlands. All other numbers available at Section on Development and Affective Neuroscience (SDAN), National Institute of Mental Health, Bethesda, Maryland.*

[b]*Subgroup sensitivity analyses: 1: (early life) behavioral observations; 2: parent/teacher reports during childhood; 3: self-report measures acquired during late childhood /adolescence; 4: cIT measured during childhood.*

[c]*Data from Pittsburgh needed to be dropped from the final analyses, because the number of participants made the design rank deficient (statistical issue unrelated to data quality).*

TABLE 4 Variables and Contrasts for the Analyses

Global analysis		Regional analysis		Regional analysis with global brain measures		Vertex-wise analysis		Vertex-wise analysis with global brain measures	
Dependent variables	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**
Total ICV	Of interest	SV (16 regions)	Of interest	SV (16 regions)	Of interest	Vertex-wise CSA (2,562 vertices)	Of interest	Vertex-wise CSA (2,562 vertices)	Of interest
Total CSA	cIT	CSA (68 regions)	cIT	CSA (68 regions)	cIT	Vertex-wise CT (2,562 vertices)	cIT	Vertex-wise CT (2,562 vertices)	cIT
Mean CT	Age[a]	CT (68 regions)	Age	CT (68 regions)	Age		Age		Age
	Sex		Sex		Sex		Sex		Sex
	Age^2		Age^2		Age^2		Age^2		Age^2
	Sex by age		Sex by age		Sex by age		Sex by age		Sex by age
	Sex by age^2		Sex by age^2		Sex by age^2		Sex by age^2		Sex by age^2
	cIT by age		cIT by age		cIT by age		cIT by age		cIT by age
	cIT by sex		cIT by sex		cIT by sex		cIT by sex		cIT by sex
	cIT by age^2		cIT by age^2		cIT by age^2		cIT by age^2		cIT by age^2
	cIT by sex by age		cIT by sex by age		cIT by sex by age		cIT by sex by age		cIT by sex by age
	cIT by sex by age^2		cIT by sex by age^2		cIT by sex by age^2		cIT by sex by age^2		cIT by sex by age^2
	Nuisance: Scanner		**Nuisance:** Scanner		**Nuisance:** Scanner, total CSA, mean CT, total ICV		**Nuisance:** Scanner		**Nuisance:** Scanner, total CSA, mean CT, total ICV
Contrasts of interest		**Contrasts of interest**		**Contrasts of interest**		**Contrasts of interest**		**Contrasts of interest**	
1	Positive effect cIT	1	Positive effect cIT	1	Positive effect cIT	1	Positive effect cIT	1	Positive effect cIT
2	Negative effect cIT	2	Negative effect cIT	2	Negative effect cIT	2	Negative effect cIT	2	Negative effect cIT
3	cIT by sex (positive)	3	cIT by sex (positive)	3	cIT by sex (positive)	3	cIT by sex (positive)	3	cIT by sex (positive)
4	cIT by sex (negative)	4	cIT by sex (negative)	4	cIT by sex (negative)	4	cIT by sex (negative)	4	cIT by sex (negative)

(*continued*)

TABLE 4 Continued

Global analysis		Regional analysis		Regional analysis with global brain measures		Vertex-wise analysis		Vertex-wise analysis with global brain measures	
Dependent variables	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**	**Dependent variables**	**Independent variables**
5	cIT by age[b]	5	cIT by age[b]	5	cIT by age[b]	5	cIT by age[b]	5	cIT by age[b]
6	cIT by sex by age	6	cIT by sex by age	6	cIT by sex by age	6	cIT by sex by age	6	cIT by sex by age

Note: cIT = childhood inhibited temperament; CSA = cortical surface area; CT = cortical thickness; ICV = intracranial volume; SV = subcortical volumes.
[a]Age represents age at timepoint of magnetic resonance imaging scan.
[b]Linear and quadratic effects of age are combined using an F test.

Analyses were performed using the Permutation Analysis of Linear Models software (PALM; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM) with 2,000 permutations, followed by the fitting of a generalized Pareto distribution to the tail of the permutation distribution,[181] thus dispensing with the need to perform a computationally prohibitive larger number of permutations.

Global, Regional, and Local (Vertex-wise) Analyses. We investigated the association between cIT and 3 global brain measures: total intracranial volume (ICV), total cortical surface area (CSA), and mean cortical thickness (CT).

We also investigated the relationship between cIT and subcortical volumes (SV; 8 regions in each hemisphere) and regional estimates of CSA and CT (34 bilateral regions according to the Desikan–Killiany parcellation[183]), in 2 models, one with and one without global brain measures (ie, total CSA, mean CT, and total intracranial volume) as additional nuisance variables (Table 4). Symptom scores and diagnostic information were not included in the models, as availability of clinical information for specific diagnoses varied widely across samples (Table S3, available online). Vertex-wise CSA and CT were investigated (2,562 vertices per hemisphere) in models similar to those used in the regional analyses (Table 4).

Sensitivity Analyses per Subgroup Based on cIT Assessment Type. To examine the neurobiological substrates of cIT in more homogenous but smaller samples, we repeated the analyses described above in subgroups of the dataset based on the type of cIT assessment (which closely varied with age at assessment). We defined the following subgroups: first, (early-life) behavioral observations; second, parent/teacher reports during childhood; and third, self-report measures acquired during late childhood and adolescence. Allocation of the samples to the subgroups is provided in Table 2, and sample sizes for each analysis are described in Table 3.

Sensitivity Analysis Excluding Retrospective Measures of cIT. In a fourth sensitivity analysis, we selected only those samples in which temperament was assessed during childhood, meaning that samples with retrospective measures on temperament (Maryland-TAX, Maryland-PAX, Vanderbilt-adults, and Virginia Commonwealth University Juvenile Anxiety Study) (Table 2) were excluded.

Correction for Multiple Testing. Multiple testing correction used the distribution of the maximum statistic,[182,184] thus allowing control over the family-wise error rate (FWER). Correction considered all tests within each metric

(ie, 68 cortical regions each for CSA and CT, and 16 SV), all 3 sets of metrics, and all contrasts (ie, MC-FWER).[184] Results at more liberal levels of correction for multiple testing (eg, only within a metric [M-FWER] or only across contrasts [C-FWER]) are reported in Supplement 3, available online.

Timeline for Completion of the Study

Data collection (ie, retrieval of MRI data, cIT information, and further descriptive data from participating sites) was locked on December 31, 2022. Image processing and organization of data took place from January 2022 to December 2023; analyses as described above took place from January to November 2024.

Data Access Certification

The first and last authors of this paper (JMBH, RB, BB, AMW, DSP) declared that they did not perform any cIT analyses on the MRI data for the purpose of the mega-analysis described in this preregistration before submitting the registered report of this study, although AMW and DSP were involved in a previous study on the relationship between cIT and brain structure (subset of the TOTS sample).[55]

RESULTS

Analyses in Full Mega-Analytic Sample

Following image processing and quality control, the analysis included cIT and MRI data from 3,803 participants (1,966 female), originating from 17 independent samples (Table 3).

Analyses revealed no significant effects of cIT, nor interactions between levels of cIT, sex, or age, on any of the brain measures when applying FWER correction considering all imaging metrics and contrasts (MC-FWER) (Figure 1). There were no significant effects either when more liberal levels of correction were used (C-FWER correction, considering only the 6 contrasts of interest; M-FWER correction, considering only the number of tests within each metric). Results of exploratory analyses are reported in Supplement 3, available online.

Sensitivity Analyses in Subgroups

As preregistered, we repeated analyses in subsamples based on cIT assessment type (Table 3). The first sensitivity analysis included data from 989 participants for whom cIT was determined based on behavioral observations, mostly early in life. Regional analyses in this subgroup did not yield any significant results when MC-FWER correction was applied; vertex-wise analyses on CT revealed 1 isolated vertex (located at the border of the right superior parietal region and supramarginal gyrus) with a significant cIT by sex by age interaction at the MC-FWER-corrected significance level ($p_{\text{MC-FWER}} = .020$). Findings at M-FWER corrected significance level (for both main and exploratory analyses) are reported in Supplement 3, available online.

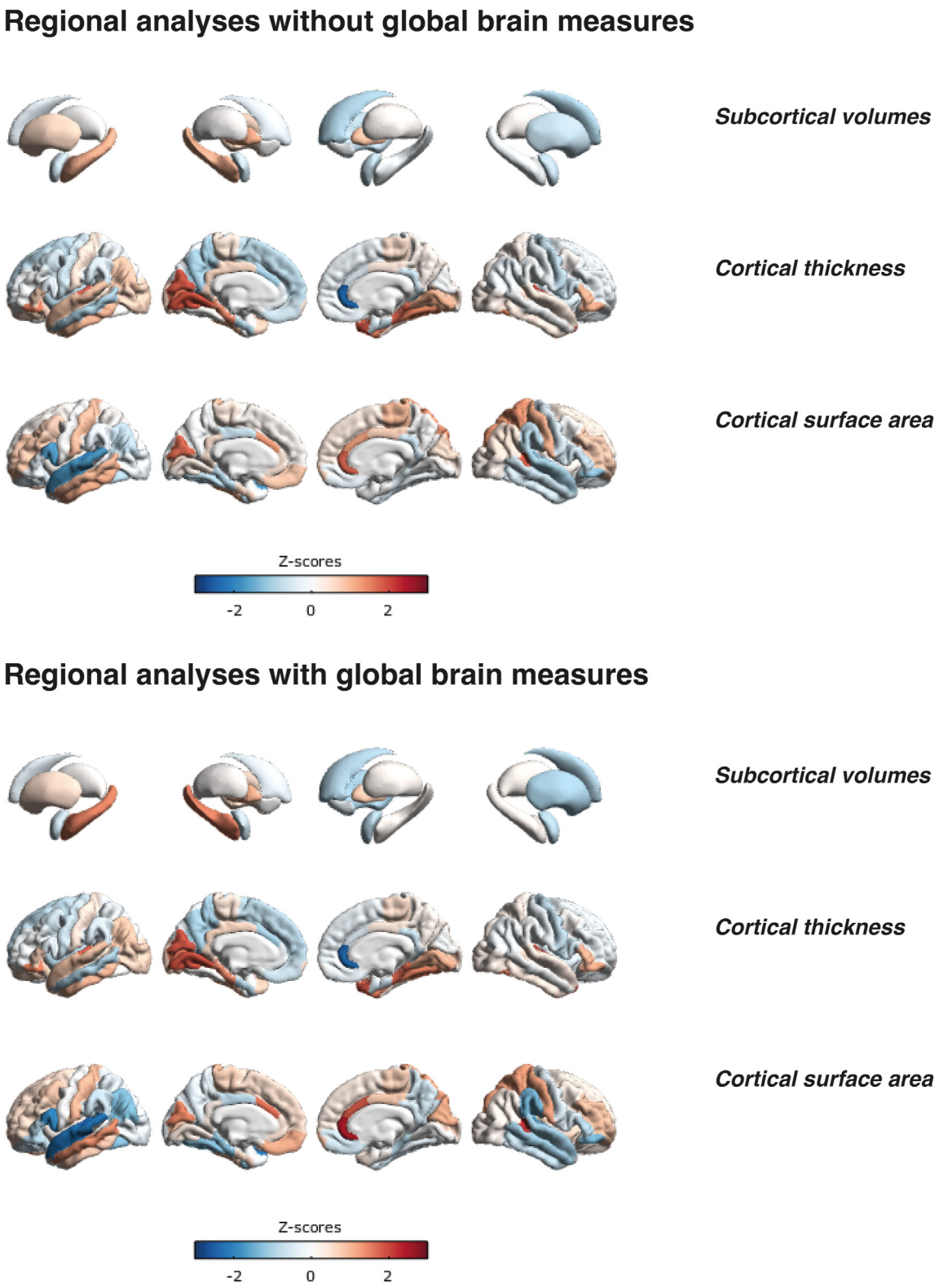
The sensitivity analyses in the second subgroup (1,879 participants for whom the level of cIT was established based on parent or teacher reports during childhood) revealed significant negative cIT by sex interactions with respect to global mean CT ($p_{\text{MC-FWER}} = .037$) and CT of the right superior parietal region ($p_{\text{MC-FWER}} = .029$), at the most stringent level of thresholding. Vertex-wise analyses revealed negative cIT by sex interactions as well, in 2 independent vertices in the left inferior parietal area ($p_{\text{MC-FWER}} = .006$ and $p_{\text{MC-FWER}} = .003$, respectively). Because most data within this subgroup originated from Generation R (n = 1,604), we repeated the analyses without Generation R data (remaining n = 275). In these analyses, the negative cIT by sex interactions with respect to mean CT and CT in the right superior parietal region ($p_{\text{MC\_FWER}} = .106$ and $p_{\text{MC\_FWER}} = .136$, respectively) did not survive, nor did we replicate the interaction effects from the vertex-wise analysis. As we considered unraveling site-specific effects out of the scope of the present preregistered work, we did not explore this interaction further. Additional findings within subgroup 2 (at M-FWER corrected level) are listed in Supplement 3, available online.

Analyses in the third subgroup (935 participants who completed self-report measures on the level of cIT) and in the fourth subgroup (3,340 participants in whom cIT was actually assessed during childhood, excluding those with retrospective measures of cIT) did not reveal significant findings when MC-FWER correction was applied. Findings from exploratory analyses are included in Supplement 3, available online.

DISCUSSION

This preregistered mega-analytic study from the ENIGMA-Anxiety Working Group explored structural brain characteristics related to childhood inhibited temperament (cIT: n = 3,803 from 17 independent samples). We used a comprehensive approach considering subcortical volumes, cortical thickness, and cortical surface area as well as multiple other subcortical and cortical indices in exploratory analyses. Contrary to expectations, cIT showed no association with brain structure, nor interactions with either sex or age in the full sample, using the preregistered statistical threshold.

FIGURE 1 *Z* Statistics for the Regional Analyses Exploring the Positive Effect of Childhood Inhibited Temperament on Brain Structure



Note: *Contrast of interest 1; none of these z statistics were statistically significant following family-wise error rate correction for multiple testing across all metrics and all contrasts (MC-FWER). Figure created with ENIGMA-Toolbox.*[185]

When considering subgroups (defined based on cIT assessment type), a significant interaction arose in the sensitivity analyses in subgroup 2. In this group, cIT level was based on assessments by parents or teachers, using questionnaires such as the Infant Behavior Questionnaire (IBQ-r; Generation R), the Behavioral Inhibition Questionnaire (BIQ; BRAINS study and VCU-JAS study), or a composite measure of cIT reported by teachers (San Raffaele sample). Individuals in this subgroup were between 8 and 16 years of age at the time of the MRI scan, and analyses revealed significant negative cIT by sex interactions with respect to global mean cortical thickness (CT) as well as when considering CT of the right superior parietal region, at the most stringent level of thresholding. These results lost significance after excluding data from Generation R, suggesting either power issues or site-specific differences rather than universal cIT relations with brain structure. However, several isolated findings from vertex-wise analyses also implicated parietal regions, namely the right superior parietal region (sensitivity analysis subgroup 1) and the left inferior parietal area (sensitivity analysis subgroup 2).

Findings in the parietal region relate to other reports in current literature. For example, a recently published Mendelian randomization study related increased volume of the left superior parietal area to the onset of anxiety disorders.[186] A study in SAD patients reported cortical thickness alterations in the right superior parietal lobule, angular gyrus, right precuneus, and inferior parietal lobule.[187] Thickening of the right parietal cortex was also found in a study comparing medication-naive patients with SAD and major depressive disorder with healthy participants,[188] suggesting that structural brain characteristics of the superior parietal region could be transdiagnostically implicated in internalizing disorders. As summarized by Sylvester *et al.*,[189] such findings could relate to the role of the superior parietal lobule in functions performed by the dorsal attention network, and multiple studies reported associations between this network and both anxiety and temperament.[190-195] A multimodal meta-analysis examining neural correlates of personality dimensions among 3,053 healthy participants further implicated these brain areas in harm avoidance, at both the structural and functional level.[196] Thus, further research in parietal structure and function as it relates to inhibited behavior appears to be warranted.

Exploratory analyses using detailed analytical techniques (analyses on amygdalar and thalamic subnuclei, hippocampal subfields, partial volume effects, volumes of additional subcortical limbic structures) are reported in Supplement 3, available online, and revealed some findings in hypothesized regions of interest. However, only 1 finding for the left hippocampus (PVE analysis; $p_{\text{MC-FWER}} = .036$, subgroup 4) survived MC-FWER correction for multiple testing. Other results at lower thresholds (Supplement 3, available online) might also be informative, as they were present in regions that we *a priori* hypothesized to be related to cIT, namely the caudate and putamen. Interestingly, volumetric differences in putamen volume were also recently reported in ENIGMA-Anxiety mega-analyses on adult patients with SAD[109] and individuals with specific phobia,[110] using a dichotomous approach comparing patients and controls. Analyses of amygdalar subnuclei and hippocampal subfields did not yield results at the MC-FWER, M-FWER, or C-FWER corrected significance level.

Previous studies in smaller samples (published before 2020; most of them included in the present mega-analysis) did report associations between the level of cIT and brain structure (Table 1). These explorations were often limited to specific regions of interest, and findings were inconsistent and lacked replication. None of these findings were replicated in this large sample in which multiple measures of cIT were combined. These null findings illustrate the value of big datasets and preregistered analyses. As outlined in a recent review paper on the advantages and challenges of big data in psychiatric neuroimaging,[197] underpowered studies are often associated with sampling variability. This could, in combination with publication bias, create the illusion of strong effect sizes, whereas real brain–behavior associations are probably smaller[197] (see also the report by Scheel *et al.* highlighting the excess of positive results in standard psychology research compared with registered reports,[198] and an editorial highlighting the value of null findings[199]).

The present findings suggest that although inhibited temperament is a partly heritable and early observable trait with a persistent character during the lifespan which is associated with the development of psychopathology,[200] this trait shows no relation to brain structure as measured using current 3T MRI scans. These findings are consistent with recent research on the structural correlates of other relevant constructs. For example, Xu *et al.* (2024) explored neuroanatomical correlates of child psychiatric problems in 11,271 children 9 to 10 years of age.[201] Using multivariate machine learning techniques, the authors reported a highly reliable and generalizable brain–behavior association between child attention problems and brain morphometry, but a robust multivariate pattern of neuroanatomy related to internalizing disorders was not demonstrated.[201] Similar results also arose from other recent studies exploring neuroanatomical data from the ABCD Study[202] and the BIS-subscale of the Carver–White BIS/BAS questionnaire.[203,204] Functional indices may be more sensitive than

structural indices to internalizing problems and their risk factors, such as cIT, at this point in the lifespan,[8] a hypothesis that seems to be supported by the evidence summarized in Table 1. Alternatively, cross-sample variation in cIT assessments (hence, phenotypic variability) could have reduced sensitivity in our analyses on associations between cIT and brain structure. This hypothesis received some support from the sensitivity analyses (although even within these subgroups, substantial variation was present in the method of cIT assessment and thus potentially in levels of cIT; see Fleece and Teglasi [2024] outlining how informant discrepancies, for example between parents' and teachers' ratings of cIT in children, emphasize the transactional nature of inhibited behavior[205]).

The strengths of the current study are the unique sample size, including data from ENIGMA-Anxiety Working Group research sites from multiple countries and continents,[80] as well as the preregistered and comprehensive standardized method to assess brain structure in relation to cIT, combined with stringent statistical correction for multiple testing. Given that there is no established way to harmonize cIT measures, when designing the study and preregistering it, we opted for an approach that would not require data to be harmonized across samples. For scanner effects within sample, we note that whereas some samples used multiple scanners, others did not; because harmonization alters the data in ways that cannot be accounted for by the NPC approach, we opted to treat the sites identically, following the classical strategy of including 1 regressor per scanner, an approach that works symmetrically for sites that use one or multiple of them.

Limitations inherent to the composed nature of the dataset deserve to be mentioned as well. For example, sample heterogeneity between datasets could not be avoided, leading to the above-mentioned cross-sample variation in cIT assessment methods, but also a large age range of included participants (cIT defined as "temperament during childhood, up to and including 12 years of age"; MRI scan acquired between 6 and 25 years of age). Combined with the cross-sectional nature of the current analysis, these limitations precluded exploring developmental patterns in more specific age groups, for example as reported by Filippi *et al.* (2020).[55] Major developmental changes occur in brain structure during childhood and adolescence, processes that are characterized by substantial interindividual variability,[206] and region-specific developmental changes in our hypothesized regions of interest in particular (see, for example, work of the ENIGMA Lifespan Working Group[114]). Hence, future work might consider hypotheses on developmental interactions between the level of cIT and structural brain characteristics. Such work could define *a priori* hypotheses on these patterns and test them in groups of participants in which MRI data were acquired within smaller age ranges. In particular, longitudinal studies with consistent assessments of cIT are recommended to disentangle such effects on brain structure at the individual level.

Next, as clinical data were not consistently available across samples, it was not possible to take levels of psychopathology (in the children and/or their parents) into account. Another methodological limitation is inherent to the use of NPC: although this method allowed a joint test that combined the various cIT measures in their native scales and yielded interpretable test statistics (z values) (Figure 1),[184] it does not produce standardized effect sizes, and unpacking interaction effects across different sites is complicated. Furthermore, information on environmental factors that might play a role in the development of internalizing psychopathology in children with high levels of cIT was lacking. For example, a recent study using longitudinal growth modeling exploring the development of social anxiety in youth 9 to 15 years of age demonstrated an interaction between infant temperament and parenting style at 36 months.[207] Another parental factor that has been identified to influence the relation between temperament of a child and the development of psychopathology is the interparental relationship quality[208]; however, a 20-year longitudinal, multi-informant, and multi-methods study on 128 children that reported an indirect association between high cIT and high loneliness during adolescence did not provide evidence for a moderating effect of infant parenting on this relationship.[209] Future longitudinal large-scale studies are needed to explore these complex relationships and interactions between innate, biological, and environmental factors that play a role in the development of anxiety in youth.

To conclude, this preregistered mega-analysis did not find consistent structural brain characteristics related to cIT in a large dataset consisting of 3,803 structural MRI scans collected from 17 independent research samples around the world that were analyzed using a single analytic pipeline and quality control. Findings from subgroup analyses and exploratory analyses point at changes in parietal regions as well as subcortical regions such as the hippocampus, putamen, and caudate in relation to cIT, but these findings did not survive statistical correction for multiple testing. Future large-scale studies, preferably using longitudinal designs and looking at brain function, are recommended to further unravel this neurobiological risk signature for internalizing psychopathology.

CRediT authorship contribution statement

Janna Marie Bas-Hoogendam: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Rachel A. Bernstein:** Writing – review & editing, Project administration, Data curation. **Brenda E. Benson:** Project administration, Data curation, Writing – review & editing. **Samuel E.C. Frank:** Writing – review & editing, Data curation. **Kristin A. Buss:** Writing – review & editing, Data curation. **Kelley E. Gunther:** Data curation, Writing – review & editing. **Koraly Pérez-Edgar:** Writing – review & editing, Data curation. **Giovanni A. Salum:** Writing – review & editing, Data curation. **Andrea Jackowski:** Data curation, Writing – review & editing. **Rodrigo A. Bressan:** Writing – review & editing, Data curation. **André Zugman:** Writing – review & editing, Data curation. **Kathryn A. Degnan:** Writing – review & editing, Data curation. **Courtney A. Filippi:** Writing – review & editing, Data curation. **Nathan A. Fox:** Writing – review & editing, Data curation, Conceptualization. **Heather A. Henderson:** Writing – review & editing, Data curation. **Alva Tang:** Writing – review & editing, Data curation. **Selin Zeytinoglu:** Data curation, Writing – review & editing. **Anita Harrewijn:** Writing – review & editing, Data curation. **Manon H.J. Hillegers:** Writing – review & editing, Data curation. **Ryan L. Muetzel:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Tonya White:** Writing – review & editing, Data curation. **Marinus H. van IJzendoorn:** Writing – review & editing, Data curation. **Carl E. Schwartz:** Writing – review & editing. **Julia Felicione:** Writing – review & editing. **Kathryn A. DeYoung:** Writing – review & editing, Data curation. **Alexander J. Shackman:** Writing – review & editing, Data curation. **Jason F. Smith:** Writing – review & editing, Data curation. **Rachael M. Tillman:** Writing – review & editing, Data curation. **Yvonne H.M. van den Berg:** Writing – review & editing, Data curation. **Antonius H.N. Cillessen:** Writing – review & editing, Data curation. **Karin Roelofs:** Writing – review & editing, Data curation. **Anna Tyborowska:** Writing – review & editing, Data curation. **Shirley Y. Hill:** Writing – review & editing, Data curation. **Marco Battaglia:** Data curation, Writing – review & editing. **Marco Tettamanti:** Writing – review & editing, Data curation. **Lea R. Dougherty:** Writing – review & editing, Data curation. **Jingwen Jin:** Writing – review & editing, Data curation. **Daniel N. Klein:** Writing – review & editing, Data curation. **Hoi-Chung Leung:** Writing – review & editing, Data curation. **Suzanne N. Avery:** Writing – review & editing, Data curation. **Jennifer Urbano Blackford:** Writing – review & editing, Conceptualization, Data curation. **Jacqueline A. Clauss:** Writing – review & editing, Data curation. **James M. Bjork:** Writing – review & editing, Data curation. **John M. Hettema:** Writing – review & editing, Data curation. **Ashlee A. Moore:** Writing – review & editing, Data curation. **Roxann Roberson-Nay:** Writing – review & editing, Data curation. **Chelsea Sawyers:** Writing – review & editing, Data curation. **Elizabeth P. Hayden:** Writing – review & editing, Data curation. **Pan Liu:** Writing – review & editing, Data curation. **Matthew R.J. Vandermeer:** Writing – review & editing, Data curation. **H. Hill Goldsmith:** Writing – review & editing, Data curation. **Elizabeth M. Planalp:** Writing – review & editing, Data curation. **Thomas E. Nichols:** Writing – review & editing, Methodology, Conceptualization. **Paul M. Thompson:** Writing – review & editing, Methodology, Conceptualization. **P. Michiel Westenberg:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Nic J.A. van der Wee:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Nynke A. Groenewold:** Methodology, Conceptualization, Writing – review & editing. **Dan J. Stein:** Writing – review & editing, Methodology, Conceptualization. **Anderson M. Winkler:** Writing – review & editing, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Daniel S. Pine:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Accepted June 27, 2025.

[a]Leiden University, Leiden, the Netherlands; [b]Leiden University Medical Center, Leiden, the Netherlands; [c]Leiden Institute for Brain and Cognition, Leiden, the Netherlands; [d]National Institute of Mental Health, Bethesda, Maryland; [e]The Pennsylvania State University, University Park, Pennsylvania; [f]University of Maryland, College Park, Maryland; [g]Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [h]National Institute of Developmental Psychiatry, São Paulo, Brazil; [i]Child Mind Institute, New York, New York; [j]Federal University of São Paulo, São Paulo, Brazil; [k]The Catholic University of America, Washington, DC; [l]New York University Grossman School of Medicine, New York, New York; [m]University of Waterloo, Waterloo, Ontario, Canada; [n]University of Texas at Dellas, Richardson, Texas; [o]Tulane University, New Orleans, Louisiana; [p]Erasmus University Rotterdam, Rotterdam, the Netherlands; [q]Erasmus University Medical Center, Rotterdam, the Netherlands; [r]UCL, London, United Kingdom; [s]Universidad San Sebastián, Santiago de Chile, Chile; [t]Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; [u]Tufts University, Medford, Massachusetts; [v]Radboud University, Behavioural Science Institute, Nijmegen, the Netherlands; [w]University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [x]University of Toronto, Toronto, Ontario, Canada; [y]Centre for Addiction and Mental Health, Toronto, Ontario, Canada; [z]University of Milan-Bicocca, Milan, Italy; [aa]The University of Hong Kong, Hong Kong; [bb]Stony Brook University, Stony Brook, New York; [cc]Vanderbilt University Medical Center, Nashville, Tennessee; [dd]Munroe-Meyer Institute, University of Nebraska Medical Center, Omaha, Nebraska; [ee]University of Maryland Medical School, Baltimore, Maryland; [ff]Virginia Commonwealth University, Richmond, Virginia; [gg]Texas A&M University, Bryan, Texas; [hh]State University of New York, Oswego, New York; [ii]University of Virginia, Charlottesville, Virginia; [jj]Western University, London, Ontario, Canada; [kk]University of Alberta, Edmonton, Alberta, Canada; [ll]University of Wisconsin-Madison, Madison, Wisconsin;

[mm]University of Oxford, Oxford, United Kingdom; [nn]University of Southern California, Marina del Rey, California; [oo]University of Cape Town, Cape Town, South Africa; [pp]The University of Texas Rio Grande Valley, Brownsville, Texas

This project is funded by a Rubicon grant from the Dutch Research Council NWO (019.201SG.022), awarded to Janna Marie Bas-Hoogendam. Rachel Bernstein, Brenda E. Benson, Samuel E. C. Frank, André Zugman, Anderson M. Winkler and Daniel S. Pine were supported by the National Institute of Mental Health Intramural Research Program Project ZIAMH002782. Kristin A. Buss received funding from the National Institute of Mental Health (R01MH094633S1). Koraly Pérez-Edgar received funding from the National Institute of Mental Health (R01MH094633 and R01MH094633S1). Kathryn A. Degnan, Nathan A. Fox and Heather A. Henderson received funding from the National Institute of Mental Health (U01MH093349). Selin Zeytinoglu received funding from the National Institute of Mental Health (F32MH127869). Anita Harrewijn received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (No 101026595). Manon H. J. Hillegers, Ryan L. Muetzel, Tonya White reported that the Generation R Study is supported by Erasmus MC, Erasmus University Rotterdam, the Rotterdam Homecare Foundation, the Municipal Health Service Rotterdam area, the Stichting Trombosedienst & Artsenlaboratorium Rijnmond, the Netherlands Organization for Health Research and Development (ZonMw), and the Ministry of Health, Welfare and Sport. Tonya White was supported by the Intramural Research Program of the National Institute of Mental Health (ZIA MH002986). Alexander J. Shackman, Kathryn A. De Young, and Jason F. Smith were supported by multiple grants from the National Institute of Mental Health (R01-AA031261, R01-AA030042, R01-MH131264, R01-MH107444 and R01-MH121409). Karin Roelofs received funding from the European Research Council (ERC_CoG-2017_772337). Shirley Y. Hill received grant support from the National Institute on Alcohol Abuse and Alcoholism (AA05909, AA08082 and AA018289). Daniel N. Klein received funding from the National Institute of Mental Health (RO1-MH069942). Jennifer Urbano Blackford received funding from the National Institute of Mental Health (K01-MH083052). Jacqueline A. Clauss received funding from the National Institute of Mental Health (F30-MH097344). James M. Bjork received funding from the National Institutes of Health (U01-DA051037). John M. Hettema received funding from the National Institutes of Health (R01MH098055) and the Brain and Behavior Research Foundation (PT110243/21984). Ashlee A. Moore was supported by the National Institute of Mental Health (F31MH111229). Roxann Roberson-Nay received funding from the National Institute of Mental Health (R01MH098055). Chelsea Sawyers reported support from the National Institutes of Health (T32MH020030). Elizabeth P. Hayden and Pan Liu declared that the BrainsCAN initiative at Western University has been funded by the Canada First Research Excellence Fund (CFREF). Matthew R. J. Vandermeer received the Canadian Institutes for Health Research (CIHR) Frederick Banting & Charles Best Canada Graduate Scholarships Doctoral Award. H. Hill Goldsmith reported support from the National Institute of Mental Health (R01MH101504 and R01MH059785), as well as infrastructure support by the Waisman Center via core support from the National Institute of Child Health and Human Development (P30 HD003352 and U54 HD090256). Elizabeth M. Planalp was supported by the National Institute of Mental Health (K01MH113710). Thomas E. Nichols was supported by the National Institutes of Health (R01MH096906).Paul M. Thompson reported grants from the National Institute of Mental Health (R01MH116147, R01MH121246, R01MH123163), support from the Milken Foundation and the Baszucki Brain Research Fund. P. Michiel Westenberg was supported by the Leiden University Research Profile Health, Prevention, and the Human Life Cycle. Nynke A. Groenewold declared that this work was made possible in part by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the author. Dan J. Stein was funded by the South African Medical Research Council.

The preregistration of this study was presented as an abstract at the 34th ECNP Congress, October 2-5 2021, hybrid attendance; the 8th biennial VNOP Conference, May 24-25, 2022, Utrecht, the Netherlands; and the Reading Emotions Conference, June 8-9, 2022, online.

Data Sharing: This project used MRI and temperament data collected at various institutions. They were brought together for mega-analysis at NIMH. With respect to data sharing, there will probably be data sharing restrictions imposed by 1) ethical review boards of the participating sites and consent documents and 2) national and trans-national data sharing laws, such as the GDPR, and 3) institutional processes, some of which require a signed data transfer agreement for limited and predefined data use. Despite these limitations, data sharing might still be possible, by means of submitting a detailed analysis plan to the ENIGMA-Anxiety Working Group. If approved, access to relevant data can be provided, dependent on data availability, local PI approval, and compliance with all supervening regulations.

Anderson M. Winkler and Thomas E. Nichols served as the statistical experts for this research.

Disclosure: Janna Marie Bas-Hoogendam received support from a Talent Acceleration grant from Medical Delta (NL) and a grant from NeurolabNL (Small projects for NWA routes 21/22; NWA.1418.22.025). Kelley E. Gunther received a National Science Foundation Graduate Research Fellowship (DGE1255832). Koraly Pérez-Edgar has reported financial support from NIDA U01DA055361-01, the National Institute of Mental Healht (R56MH126349, R01MH109692, R01MH130007) and received royalties from Springer Press, honoraria for lectures to professional audiences, and honoraria for editing the journal Developmental Psychology. Giovanni A. Salum declared that the BHRCS was supported with grants from the National Institute of Development Psychiatric for Children and Adolescent (INPD) (grants CNPq 465550/2014-2 and FAPESP 2014/50917-0). Andrea P. Jackowski has reported grant support from CNPq 427215/2018-8 and FAPESP (13/08531-5). Kathryn A. Degnan has reported grant support from a subaward from the National Institute of Mental Health (U01MH093349) via University of Maryland. Courtney A. Filippi has received support intramural Research Program of the National Institute of Mental Health (NIMH; ZIA-MH-002782), NIMH extramural funding (R00MH125878), and the NARSAD Young Investigator Grant from the Brain and Behavior Foundation (#28024). Nathan A. Fox reported grant support from the National Institute of Mental Health (U01MH093349), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Science Foundation (NSF), 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. Heather A. Henderson has reported grant support from the Social Sciences and Humanities Research Council of Canada, Canada Foundation for Innovation. Carl E. Schwartz reported grants from the National Institute of Mental Health (Career Development Award, NIH/ NIMH R01MH07484 and NIH/NIMH R01MH071467). Karin Roelofs and Anna Tyborowska stated that past MR scanning of the Nijmegen Longitudinal Study (NLS) was funded by European Research Council starting grant (ERC_StG2012_313749 awarded to KR) and a FP7-HEALTH-2013-INNOVATION grant (602805-2). Shirley Y. Hill reported grant support from the National Institute on Alcohol Abuse and Alcoholism (AA021746 and AA021746-05S1-Supplement). Lea R. Dougherty has reported grants from the National Institute of Mental Health (R01MH122487 and R01MH121385). Daniel N. Klein reported multiple grants from the National Institute of Health (R01 MH133652; NIEHS R01 ES0 13744; NIEHS R01 ES033436). Roxann Roberson-Nay reported grants from the National Institute of Mental Health (R01MH101518; 5R21MH106924) and the Brain and Behavior Research Foundation (Independent Investigator, 21976). Paul M. Thompson has received partial research support from Biogen, Inc., unrelated to this work, as well as a NIH Big Data to Knowledge (BD2K) award (U54 EB020403). Nic J. A. van der Wee reported support from the EU Innovative Medicines Initiative 'Psychiatric Ratings using Intermediate Stratified Markers' 2 (IMI PRISM 2) with grant agreement nr 101034377. Dan J. Stein received research grants and/or consultancy honoraria from Discovery Vitality, Johnson & Johnson, Kanna, Lundbeck, Orion, Sanofi, Servier and Takeda. Rodrigo A. Bresssan, Alva Tang, Marinus H. van IIzendoorn, Julia M. Felicione, Kathryn A. DeYoung, Alexander J. Shackman, Jason F. Smith, Rachael M. Tillman, Yvonne H. M. van den Berg, Antonius H. N. Cillessen, Marco Battaglia, Marco Tettamanti, Jingwen Jin, Hoi-Chung Leung, and Suzanne N. Avery 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, Wassenaarseweg 52, 2333 AK Leiden, the Netherlands; e-mail: j.m.hoogendam@fsw.leidenuniv.nl

https://doi.org/10.1016/j.jaac.2025.06.026

REFERENCES

1. Kagan J, Snidman NC. The Long Shadow of Temperament. Harvard University Press; 2009.
2. Clauss JA, Avery SN, Blackford JU. The nature of individual differences in inhibited temperament and risk for psychiatric disease: a review and meta-analysis. Prog Neurobiol. 2015;127-128:23-45. https://doi.org/10.1016/j.pneurobio.2015.03.001
3. Pérez-Edgar K, Fox NA. Temperament and anxiety disorders. Child Adolesc Psychiatr Clin N Am. 2005;14(4):681-706. https://doi.org/10.1016/j.chc.2005.05.008
4. Pérez-Edgar KE, Guyer AE. Behavioral Inhibition: temperament or prodrome? Curr Behav Neurosci Rep. Published online June 24, 2014. https://doi.org/10.1007/s40473-014-0019-9
5. Kagan J, Reznick JS, Snidman N. The physiology and psychology of behavioral inhibition in children. Child Dev. 1987;58(6):1459-1473.
6. Fox AS, Harris RA, Del Rosso L, *et al*. Infant inhibited temperament in primates predicts adult behavior, is heritable, and is associated with anxiety-relevant genetic variation. Mol Psychiatry. Published online 2021. https://doi.org/10.1038/s41380-021-01156-4
7. Shackman AJ, Fox AS, Oler JA, *et al*. Heightened extended amygdala metabolism following threat characterizes the early phenotypic risk to develop anxiety-related psychopathology. Mol Psychiatry. 2017;22(5):724-732. https://doi.org/10.1038/mp.2016.132
8. Fox AS, Kalin NH. A translational neuroscience approach to understanding the development of social anxiety disorder and its pathophysiology. Am J Psychiatry. 2014;171(11):1162-1173. https://doi.org/10.1176/appi.ajp.2014.14040449
9. Sylvester CM, Pine DS. The biological bridge between behavioral inhibition and psychopathology BT—behavioral inhibition: integrating theory, research, and clinical perspectives. In: Pérez-Edgar K, Fox NA, eds. Behavioral Inhibition. Springer; 2018:309-335. https://doi.org/10.1007/978-3-319-98077-5_14
10. Mumper EE, Klein DN. Behavioral inhibition and psychopathology in childhoodThe Handbook of Solitude: Psychological Perspectives on Social Isolation, Social Withdrawal, and Being Alone. Wiley Online Library; 2021:89-104.
11. Fox NA, Henderson HA, Marshall PJ, Nichols KE, Ghera MM. Behavioral inhibition: linking biology and behavior within a developmental framework. Annu Rev Psychol. 2004;56(1):235-262. https://doi.org/10.1146/annurev.psych.55.090902.141532
12. Bornstein MH, Putnick DL, Gartstein MA, Hahn C, Auestad N, O'Connor DL. Infant temperament: stability by age, gender, birth order, term status, and socioeconomic status. Child Dev. 2015;86(3):844-863. https://doi.org/10.1111/cdev.12367
13. Tang A, Crawford H, Morales S, Degnan KA, Pine DS, Fox NA. Infant behavioral inhibition predicts personality and social outcomes three decades later. Proc Natl Acad Sci. Published online April 17, 2020:201917376. https://doi.org/10.1073/pnas.1917376117
14. Caspi A, Roberts BW, Shiner RL. Personality development: stability and change. Annu Rev Psychol. 2004;56(1):453-484. https://doi.org/10.1146/annurev.psych.55.090902.141913
15. Clauss JA, Blackford JU. Behavioral inhibition and risk for developing social anxiety disorder: a meta-analytic study. J Am Acad Child Adolesc Psychiatry. 2012;51(10):1066-1075. https://doi.org/10.1016/j.jaac.2012.08.002
16. Sandstrom A, Uher R, Pavlova B. Prospective association between childhood behavioral inhibition and anxiety: a meta-analysis. Res Child Adolesc Psychopathol. 2020;48(1):57-66. https://doi.org/10.1007/s10802-019-00588-5
17. Tan E, Zeytinoglu S, Morales S, *et al*. Social vs non-social behavioral inhibition: differential prediction from early childhood of long-term psychosocial outcomes. Dev Sci. 2024;27(1). https://doi.org/10.1111/desc.13427
18. Schwartz CE, Snidman N, Kagan J. Adolescent social anxiety as an outcome of inhibited temperament in childhood. J Am Acad Child Adolesc Psychiatry. 1999;38(8):1008-1015. https://doi.org/10.1097/00004583-199908000-00017
19. Biederman J, Hirshfeld-Becker DR, Rosenbaum JF, *et al*. Further evidence of association between behavioral inhibition and social anxiety in children. Am J Psychiatry. 2001;158(10):1673-1679. https://doi.org/10.1176/appi.ajp.158.10.1673
20. Fox NA, Buzzell GA, Morales S, Valadez EA, Wilson M, Henderson HA. Understanding the emergence of social anxiety in children with behavioral inhibition. Biol Psychiatry. 2021;89(7):681-689. https://doi.org/10.1016/j.biopsych.2020.10.004
21. Goldsmith HH, Hilton EC, Phan JM, *et al*. Childhood inhibition predicts adolescent social anxiety: findings from a longitudinal twin study. Dev Psychopathol. Published online 2022:1-20. https://doi.org/10.1017/S0954579422000864
22. Bourdon JL, Savage JE, Verhulst B, *et al*. The genetic and environmental relationship between childhood behavioral inhibition and preadolescent anxiety. Twin Res Hum Genet. 2019;22(1):48-55. https://doi.org/10.1017/thg.2018.73
23. Bakken NR, Hannigan LJ, Shadrin A, *et al*. Childhood temperamental, emotional, and behavioral characteristics associated with mood and anxiety disorders in adolescence: a prospective study. Acta Psychiatr Scand. 2023;147(2):217-228. https://doi.org/10.1111/acps.13522
24. Bas-Hoogendam JM, Roelofs EF, Westenberg PM, van der Wee NJA. Pathogenesis of SAD. In: Simon NM, Hollander E, Rothbaum BO, Stein DJ, eds. Textbook of Anxiety, Trauma and OCD-Related Disorders. 3rd ed. American Psychiatric Association Publishing; 2020:429-444.
25. Bas-Hoogendam JM, Blackford JU, Brühl AB, Blair KS, van der Wee NJA, Westenberg PM. Neurobiological candidate endophenotypes of social anxiety disorder. Neurosci Biobehav Rev. 2016;71:362-378. https://doi.org/10.1016/j.neubiorev.2016.08.040
26. Chronis-Tuscano A, Degnan KA, Pine DS, *et al*. Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. J Am Acad Child Adolesc Psychiatry. 2009;48(9):928-935. https://doi.org/10.1097/CHI.0b013e3181ae09df
27. Essex MJ, Klein MH, Slattery MJ, Goldsmith HH, Kalin NH. Early risk factors and developmental pathways to chronic high inhibition and social anxiety disorder in adolescence. Am J Psychiatry. 2010;167(1):40-46. https://doi.org/10.1176/appi.ajp.2009.07010051
28. Morales S, Tang A, Bowers ME, *et al*. Infant temperament prospectively predicts general psychopathology in childhood. Dev Psychopathol. 2022;34(3):774-783. https://doi.org/10.1017/S0954579420001996
29. Poole KL, Tang A, Schmidt LA. The temperamentally shy child as the social adult: an exemplar of multifinality BT—behavioral inhibition: integrating theory, research, and clinical perspectives. In: Pérez-Edgar K, Fox NA, eds. Behavioral Inhibition. Springer; 2018:185-212. https://doi.org/10.1007/978-3-319-98077-5_9
30. Liu P, Pérez-Edgar KE. Developmental pathways from early behavioral inhibition to later anxiety: an integrative review of developmental psychopathology research and translational implications. Adolesc Res Rev. 2019;4(1):45-58. https://doi.org/10.1007/s40894-018-0092-5
31. Klein DN, Mumper EE. Behavioral inhibition as a precursor to psychopathology BT—behavioral inhibition: integrating theory, research, and clinical perspectives. In: Pérez-Edgar K, Fox NA, eds. Behavioral Inhibition. Springer; 2018:283-307. https://doi.org/10.1007/978-3-319-98077-5_13
32. Sylvester CM. Editorial: Several roads lead to anxiety. J Am Acad Child Adolesc Psychiatry. 2020;59(6):703-704. https://doi.org/10.1016/j.jaac.2019.09.030
33. Blackford JU, Clauss JA, Benningfield MM. The Neurobiology of behavioral inhibition as a developmental mechanism. In: Pérez-Edgar K, Fox NA, eds. Behavioral Inhibition. Springer; 2018:113-134. https://doi.org/10.1007/978-3-319-98077-5_6
34. Taber-Thomas BC, Morales S, Hillary FG, Pérez-Edgar KE. Altered topography of intrinsic functional connectivity in childhood risk for social anxiety. Depress Anxiety. 2016;33:995-1004. https://doi.org/10.1002/da.22508
35. Auday ES, Pérez-Edgar KE. Limbic and prefrontal neural volume modulate social anxiety in children at temperamental risk. Depress Anxiety. 2019;36(8):690-700. https://doi.org/10.1002/da.22941
36. Auday ES, Taber-Thomas BC, Pérez-Edgar KE. Neural correlates of attention bias to masked facial threat cues: examining children at-risk for social anxiety disorder. Neuroimage Clin. 2018;19:202-212. https://doi.org/10.1016/j.nicl.2018.04.003
37. Clauss JA, Avery SN, Vanderklok RM, *et al*. Neurocircuitry underlying risk and resilience to social anxiety disorder. Depress Anxiety. 2014;31(10):822-833. https://doi.org/10.1002/da.22265
38. Blackford JU, Clauss JA, Avery SN, Cowan RL, Benningfield MM, Vanderklok RM. Amygdala-cingulate intrinsic connectivity is associated with degree of social inhibition. Biol Psychol. 2014;99:15-25. https://doi.org/10.1016/j.biopsycho.2014.02.003
39. Blackford JU, Avery SN, Cowan RL, Shelton RC, Zald DH. Sustained amygdala response to both novel and newly familiar faces characterizes inhibited temperament. Soc Cogn Affect Neurosci. 2011;6(5):621-629. https://doi.org/10.1093/scan/nsq073
40. Blackford JU, Allen AH, Cowan RL, Avery SN. Amygdala and hippocampus fail to habituate to faces in individuals with an inhibited temperament. Soc Cogn Affect Neurosci. 2013;8(2):143-150. https://doi.org/10.1093/scan/nsr078
41. Blackford JU, Avery SN, Shelton RC, Zald DH. Amygdala temporal dynamics: temperamental differences in the timing of amygdala response to familiar and novel faces. BMC Neurosci. 2009;10:145. https://doi.org/10.1186/1471-2202-10-145
42. Clauss JA, Seay AL, Vanderklok RM, *et al*. Structural and functional bases of inhibited temperament. Soc Cogn Affect Neurosci. 2014;9(12):2049-2058. https://doi.org/10.1093/scan/nsu019
43. Schwartz CE, Wright CI, Shin LM, Kagan J, Rauch SL. Inhibited and uninhibited infants "grown up": adult amygdalar response to novelty. Science (1979). 2003;300(5627):1952-1953. https://doi.org/10.1126/science.1083703
44. Guyer AE, Nelson EE, Perez-Edgar K, *et al*. Striatal functional alteration in adolescents characterized by early childhood behavioral inhibition. J Neurosci. 2006;26(24):639LP-6405. https://doi.org/10.1523/JNEUROSCI.0666-06.2006
45. Pérez-Edgar K, Roberson-Nay R, Hardin MG, *et al*. Attention alters neural responses to evocative faces in behaviorally inhibited adolescents. Neuroimage. 2007;35(4):1538-1546. https://doi.org/10.1016/j.neuroimage.2007.02.006

46. Schwartz CE, Kunwar PS, Greve DN, *et al*. Structural differences in adult orbital and ventromedial prefrontal cortex predicted by infant temperament at 4 months of age. Arch Gen Psychiatry. 2010;67(1):78-84. https://doi.org/10.1001/archgenpsychiatry.2009.171
47. Schwartz CE, Kunwar PS, Greve DN, Kagan J, Snidman NC, Bloch RB. A phenotype of early infancy predicts reactivity of the amygdala in male adults. Mol Psychiatry. 2012;17(10):1042-1050. https://doi.org/10.1038/mp.2011.96
48. Pérez-Edgar K, Hardee JE, Guyer AE, *et al*. DRD4 and striatal modulation of the link between childhood behavioral inhibition and adolescent anxiety. Soc Cogn Affect Neurosci. 2014;9(4):445-453. https://doi.org/10.1093/scan/nst001
49. Troller-Renfree SV, Buzzell GA, Bowers ME, *et al*. Development of inhibitory control during childhood and its relations to early temperament and later social anxiety: unique insights provided by latent growth modeling and signal detection theory. J Child Psychol Psychiatry. 2019;60(6):622-629. https://doi.org/10.1111/jcpp.13025
50. Lahat A, Benson BE, Pine DS, Fox NA, Ernst M. Neural responses to reward in childhood: relations to early behavioral inhibition and social anxiety. Soc Cogn Affect Neurosci. Published online August 16, 2018;281-289. https://doi.org/10.1093/scan/nsw122
51. Hill SY, Tessner K, Wang S, Carter H, McDermott M. Temperament at 5 years of age predicts amygdala and orbitofrontal volume in the right hemisphere in adolescence. Psychiatry Res Neuroimaging. 2010;182(1):14-21. https://doi.org/10.1016/j.pscychresns.2009.11.006
52. Sylvester CM, Barch DM, Harms MP, *et al*. Early childhood behavioral inhibition predicts cortical thickness in adulthood. J Am Acad Child Adolesc Psychiatry. 2016;55(2):122-129. https://doi.org/10.1016/j.jaac.2015.11.007
53. Clarkson T, Eaton NR, Nelson EE, *et al*. Early childhood social reticence and neural response to peers in preadolescence predict social anxiety symptoms in midadolescence. Depress Anxiety. 2019;36(8):676-689. https://doi.org/10.1002/da.22910
54. Abend R, Swetlitz C, White LK, *et al*. Levels of early-childhood behavioral inhibition predict distinct neurodevelopmental pathways to pediatric anxiety. Psychol Med. 2020;50(1):96-106. https://doi.org/10.1017/S0033291718003999
55. Filippi CA, Sachs JF, Phillips D, *et al*. Infant behavioral reactivity predicts change in amygdala volume 12 years later. Dev Cogn Neurosci. 2020;42. https://doi.org/10.1016/j.dcn.2020.100776
56. Troller-Renfree SV, Buzzell GA, Pine DS, Henderson HA, Fox NA. Consequences of not planning ahead: reduced proactive control moderates longitudinal relations between behavioral inhibition and anxiety. J Am Acad Child Adolesc Psychiatry. 2019;58(8):768-775. https://doi.org/10.1016/j.jaac.2018.06.040
57. Bas-Hoogendam JM, van Steenbergen H, Tissier RLM, Houwing-Duistermaat JJ, Westenberg PM, van der Wee NJA. Subcortical brain volumes, cortical thickness and cortical surface area in families genetically enriched for social anxiety disorder-A multiplex multigenerational neuroimaging study. EBioMedicine. Published online September 26, 2018:410-428. https://doi.org/10.1016/j.ebiom.2018.08.048
58. Bas-Hoogendam JM, van Steenbergen H, van der Wee NJA, Westenberg PM. Amygdala hyperreactivity to faces conditioned with a social-evaluative meaning—a multiplex, multigenerational fMRI study on social anxiety endophenotypes. Neuroimage Clin. Published online 2020:102247. https://doi.org/10.1016/j.nicl.2020.102247
59. Bas-Hoogendam JM, Harrewijn A, Tissier RLM, *et al*. The Leiden Family Lab Study on Social Anxiety Disorder: a multiplex, multigenerational family study on neurocognitive endophenotypes. Int J Methods Psychiatr Res. 2018;27(2):e1616. https://doi.org/10.1002/mpr.1616
60. Bas-Hoogendam JM, van Steenbergen H, Blackford JU, Tissier RLM, van der Wee NJA, Westenberg PM. Impaired neural habituation to neutral faces in families genetically enriched for social anxiety disorder. Depress Anxiety. 2019;36(12):1143-1153. https://doi.org/10.1002/da.22962
61. Bas-Hoogendam JM, van Steenbergen H, Tissier RLM, van der Wee NJA, Westenberg PM. Altered neurobiological processing of unintentional social norm violations: a multiplex, multigenerational fMRI study on social anxiety endophenotypes. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5(10):981-990. https://doi.org/10.1016/j.bpsc.2019.03.003
62. Oler JA, Fox AS, Shelton SE, *et al*. Amygdalar and hippocampal substrates of anxious temperament differ in their heritability. Nature. 2010;466(7308):864-868. https://doi.org/10.1038/nature09282
63. Clinton SM, Shupe EA, Glover ME, *et al*. Modeling heritability of temperamental differences, stress reactivity, and risk for anxiety and depression: relevance to Research Domain Criteria (RDoC). Eur J Neurosci. 2022;55(9):2076-2107. https://doi.org/10.1111/ejn.15158
64. McNaughton N, Gray JA. Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. J Affect Disord. 2000;61(3):161-176. https://doi.org/10.1016/s0165-0327(00)00344-x
65. Fu X, Taber-Thomas BC, Pérez-Edgar K. Frontolimbic functioning during threat-related attention: relations to early behavioral inhibition and anxiety in children. Biol Psychol. 2017;122:98-109. https://doi.org/10.1016/j.biopsycho.2015.08.010
66. Bar-Haim Y, Fox NA, Benson B, *et al*. Neural correlates of reward processing in adolescents with a history of inhibited temperament. Psychol Sci. 2009;20(8):1009-1018. https://doi.org/10.1111/j.1467-9280.2009.02401.x
67. Guyer AE, Benson B, Choate VR, *et al*. Lasting associations between early-childhood temperament and late-adolescent reward-circuitry response to peer feedback. Dev Psychopathol. 2014;26(1):229-243. https://doi.org/10.1017/S0954579413000941
68. Hardee JE, Benson BE, Bar-Haim Y, *et al*. Patterns of neural connectivity during an attention bias task moderate associations between early childhood temperament and internalizing symptoms in young adulthood. Biol Psychiatry. 2013;74(4):273-279. https://doi.org/10.1016/j.biopsych.2013.01.036
69. Helfinstein SM, Benson B, Perez-Edgar K, *et al*. Striatal responses to negative monetary outcomes differ between temperamentally inhibited and non-inhibited adolescents. Neuropsychologia. 2011;49(3):479-485. https://doi.org/10.1016/j.neuropsychologia.2010.12.015
70. Jarcho JM, Fox NA, Pine DS, *et al*. The neural correlates of emotion-based cognitive control in adults with early childhood behavioral inhibition. Biol Psychol. 2013;92(2):306-314. https://doi.org/10.1016/j.biopsycho.2012.09.008
71. Jarcho JM, Fox NA, Pine DS, *et al*. Enduring influence of early temperament on neural mechanisms mediating attention-emotion conflict in adults. Depress Anxiety. 2014;31(1):53-62. https://doi.org/10.1002/da.22140
72. Roy AK, Benson BE, Degnan KA, *et al*. Alterations in amygdala functional connectivity reflect early temperament. Biol Psychol. 2014;103:248-254. https://doi.org/10.1016/j.biopsycho.2014.09.007
73. Shechner T, Fox NA, Mash JA, *et al*. Differences in neural response to extinction recall in young adults with or without history of behavioral inhibition. Dev Psychopathol. 2018;30(1):179-189. https://doi.org/10.1017/S0954579417000554
74. Tang A, Harrewijn A, Benson B, *et al*. Striatal activity to reward anticipation as a moderator of the association between early behavioral inhibition and changes in anxiety and depressive symptoms from adolescence to adulthood. JAMA Psychiatry. 2022;79(12):1199-1208. https://doi.org/10.1001/jamapsychiatry.2022.3483
75. Michalska KJ, Feldman JS, Ivie EJ, *et al*. Early-childhood social reticence predicts SCR-BOLD coupling during fear extinction recall in preadolescent youth. Dev Cogn Neurosci. 2019;36:100605. https://doi.org/10.1016/j.dcn.2018.12.003
76. Jarcho JM, Davis MM, Shechner T, *et al*. Early-childhood social reticence predicts brain function in preadolescent youths during distinct forms of peer evaluation. Psychol Sci. 2016;27(6):821-835. https://doi.org/10.1177/0956797616638319
77. Smith AR, White LK, Leibenluft E, *et al*. The heterogeneity of anxious phenotypes: neural responses to errors in treatment-seeking anxious and behaviorally-inhibited youth. J Am Acad Child Adolesc Psychiatry. 2020;59(6):759-769. https://doi.org/10.1016/j.jaac.2019.05.014
78. Clauss JA, Benningfield MM, Rao U, Blackford JU. Altered prefrontal cortex function marks heightened anxiety risk in children. J Am Acad Child Adolesc Psychiatry. 2016;55(9):809-816. https://doi.org/10.1016/j.jaac.2016.05.024
79. Avery SN, Blackford JU. Slow to warm up: the role of habituation in social fear. Soc Cogn Affect Neurosci. 2016;11(11):1832-1840. https://doi.org/10.1093/scan/nsw095
80. Bas-Hoogendam JM, Groenewold NA, Aghajani M, *et al*. ENIGMA-Anxiety Working Group: rationale for and organization of large-scale neuroimaging studies of anxiety disorders. Hum Brain Mapp. 2022;43(1):83-112. https://doi.org/10.1002/hbm.25100
81. Boedhoe PSW, Heymans MW, Schmaal L, *et al*. An empirical comparison of meta- and mega-analysis with data from the ENIGMA Obsessive-Compulsive Disorder Working Group. Front Neuroinform. 2019;12. https://doi.org/10.3389/fninf.2018.00102
82. Zacharek SJ, Kribakaran S, Kitt ER, Gee DG. Leveraging big data to map neurodevelopmental trajectories in pediatric anxiety. Dev Cogn Neurosci. 2021;50:100974. https://doi.org/10.1016/j.dcn.2021.100974
83. Winkler AM, Kochunov P, Blangero J, *et al*. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. Neuroimage. 2010;53(3):1135-1146. https://doi.org/10.1016/j.neuroimage.2009.12.028
84. Hirshfeld-Becker DR, Biederman J, Henin A, *et al*. Behavioral inhibition in preschool children at risk is a specific predictor of middle childhood social anxiety: a five-year follow-up. J Dev Behav Pediatr. 2007;28(3):225-233. https://doi.org/10.1097/01.DBP.0000268559.34463.d0
85. Buzzell GA, Troller-Renfree SV, Barker TV, *et al*. A neurobehavioral mechanism linking behaviorally inhibited temperament and later adolescent social anxiety. J Am Acad Child Adolesc Psychiatry. 2017;56(12):1097-1105. https://doi.org/10.1016/j.jaac.2017.10.007
86. Buzzell GA, Morales S, Bowers ME, *et al*. Inhibitory control and set shifting describe different pathways from behavioral inhibition to socially anxious behavior. Dev Sci. 2021;24(1):e13040. https://doi.org/10.1111/desc.13040
87. Vreeke LJ, Muris P, Mayer B, *et al*. The assessment of an inhibited, anxiety-prone temperament in a Dutch multi-ethnic population of preschool children. Eur Child Adolesc Psychiatry. 2012;21(11):623-633. https://doi.org/10.1007/s00787-012-0299-0

88. Marysko M, Finke P, Wiebel A, Resch F, Moehler E. Can mothers predict childhood behavioural inhibition in early infancy? Child Adolesc Ment Health. 2010;15(2):91-96. https://doi.org/10.1111/j.1475-3588.2009.00539.x
89. Muris P, Meesters C. Reactive and regulative temperament in youths: psychometric evaluation of the Early Adolescent Temperament Questionnaire—Revised. J Psychopathol Behav Assess. 2009;31(1):7-19. https://doi.org/10.1007/s10862-008-9089-x
90. Lawson KM, Atherton OE, Robins RW. The structure of adolescent temperament and associations with psychological functioning: a replication and extension of Snyder et al. (2015). J Pers Soc Psychol. Published online February 2021. https://doi.org/10.1037/pspp0000380
91. Saygin ZM, Kliemann D, Iglesias JE, *et al*. High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. Neuroimage. 2017;155:370-382. https://doi.org/10.1016/j.neuroimage.2017.04.046
92. Iglesias JE, Augustinack JC, Nguyen K, *et al*. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. Neuroimage. 2015;115:117-137. https://doi.org/10.1016/j.neuroimage.2015.04.042
93. Iglesias JE, Insausti R, Lerma-Usabiaga G, *et al*. A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. Neuroimage. 2018;183:314-326. https://doi.org/10.1016/j.neuroimage.2018.08.012
94. Abend R, Gold AL, Britton JC, *et al*. Anticipatory threat responding: associations with anxiety, development, and brain structure. Biol Psychiatry. 2020;87(10):916-925. https://doi.org/10.1016/j.biopsych.2019.11.006
95. Greve DN, Billot B, Cordero D, *et al*. A deep learning toolbox for automatic segmentation of subcortical limbic structures from MRI images. Neuroimage. 2021;244:118610. https://doi.org/10.1016/j.neuroimage.2021.118610
96. Groenewold N, Bas-Hoogendam JM, Amod AR, *et al*. F27. subcortical volumes in social anxiety disorder: preliminary results from Enigma-Anxiety. Biol Psychiatry. 2018;83(9, Suppl):S247-S248. https://doi.org/10.1016/j.biopsych.2018.02.640
97. Harrewijn A, Cardinale EM, Groenewold NA, *et al*. Cortical and subcortical brain structure in generalized anxiety disorder: findings from 28 research sites in the ENIGMA-Anxiety Working Group. Transl Psychiatry. 2021;11(1):502. https://doi.org/10.1038/s41398-021-01622-1
98. Groenewold NA, Bas-Hoogendam JM, Aghajani M, *et al*. Brain characteristics associated with anxiety disorders: an update from the ENIGMA-Anxiety Working Group. J Neural Transm. 2021;128:1767-1812. https://doi.org/10.1007/s00702-021-02422-z
99. Cha J, Greenberg T, Song I, Blair Simpson H, Posner J, Mujica-Parodi LR. Abnormal hippocampal structure and function in clinical anxiety and comorbid depression. Hippocampus. 2016;26(5):545-553. https://doi.org/10.1002/hipo.22566
100. Montag C, Reuter M, Jurkiewicz M, Markett S, Panksepp J. Imaging the structure of the human anxious brain: a review of findings from neuroscientific personality psychology. Rev Neurosci. 2013;24(2):167-190. https://doi.org/10.1515/revneuro-2012-0085
101. Moon CM, Jeong GW. Abnormalities in gray and white matter volumes associated with explicit memory dysfunction in patients with generalized anxiety disorder. Acta Radiol. 2017;58(3):353-361. https://doi.org/10.1177/0284185116649796
102. Blackford JU. Leveraging statistical methods to improve validity and reproducibility of research findings. JAMA Psychiatry. 2017;74(2):119-120. https://doi.org/10.1001/jamapsychiatry.2016.3730
103. Thompson P, Jahanshad N, Ching CRK, *et al*. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl Psychiatry. 2020;10(100). https://doi.org/10.1038/s41398-020-0705-1
104. Bearden CE, Thompson PM. Emerging global initiatives in neurogenetics: the Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. Neuron. 2017;94(2):232-236. https://doi.org/10.1016/j.neuron.2017.03.033
105. Ching CRK, Hibar DP, Gurholt TP, *et al*. What we learn about bipolar disorder from large-scale neuroimaging: findings and future directions from the ENIGMA Bipolar Disorder Working Group. Hum Brain Mapp. 2022;43:56-82. https://doi.org/10.1002/hbm.25098
106. Hoogman M, van Rooij D, Klein M, *et al*. Consortium neuroscience of attention deficit/hyperactivity disorder and autism spectrum disorder: the ENIGMA adventure. Hum Brain Mapp. 2022;43(1):37-55. https://doi.org/10.1002/hbm.25029
107. van den Heuvel OA, Boedhoe PSW, Bertolin S, *et al*. An overview of the first 5 years of the ENIGMA Obsessive–Compulsive Disorder Working Group: the power of worldwide collaboration. Hum Brain Mapp. 2022;n/a(43):23-36. https://doi.org/10.1002/hbm.24972
108. Bas-Hoogendam JM, Groenewold NA, Amod AR, *et al*. Subcortical brain volumes in social anxiety disorder: an Enigma-Anxiety international mega-analysis of 37 samples. Biol Psychiatry. 2023;93(9):S85-S86. https://doi.org/10.1016/j.biopsych.2023.02.222
109. Groenewold NA, Bas-Hoogendam JM, Amod AR, *et al*. Volume of subcortical brain regions in social anxiety disorder: mega-analytic results from 37 samples in the ENIGMA-Anxiety Working Group. Mol Psychiatry. Published online 2023. https://doi.org/10.1038/s41380-022-01933-9
110. Hilbert K, Boeken OJ, Langhammer T, *et al*. Cortical and subcortical brain alterations in specific phobia and its animal and blood-injection-injury subtypes: a mega-analysis from the ENIGMA Anxiety Working Group. Am J Psychiatry. Published online June 11, 2024:appiajp20230032. https://doi.org/10.1176/appi.ajp.20230032
111. Bruin WB, Zhutovsky P, van Wingen GA, *et al*. Brain-based classification of youth with anxiety disorders: transdiagnostic examinations within the ENIGMA-Anxiety database using machine learning. Nat Ment Health. 2024;2(1):104-118. https://doi.org/10.1038/s44220-023-00173-2
112. Novins DK, Althoff RR, Billingsley MK, *et al*. Study registration: encouraging the practice of hypothetical-deductive research in the journal. J Am Acad Child Adolesc Psychiatry. 2018;57(12):901-902. https://doi.org/10.1016/j.jaac.2018.10.004
113. Crone EA, Dahl RE. Understanding adolescence as a period of social-affective engagement and goal flexibility. Nat Rev Neurosci. 2012;13(9):636-650. https://doi.org/10.1038/nrn3313
114. Dima D, Modabbernia A, Papachristou E, *et al*. Subcortical volumes across the lifespan: data from 18,605 healthy individuals aged 3–90 years. Hum Brain Mapp. 2022;43(1):452-469. https://doi.org/10.1002/hbm.25320
115. Frangou S, Modabbernia A, Williams SCR, *et al*. Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3–90 years. Hum Brain Mapp. 2022;43:431-451. https://doi.org/10.1002/hbm.25364
116. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? Nat Rev Neurosci. 2008;9(12):947-957. https://doi.org/10.1038/nrn2513
117. Beesdo-Baum K, Knappe S, Asselmann E, *et al*. The "Early Developmental Stages of Psychopathology (EDSP) study": a 20-year review of methods and findings. Soc Psychiatry Psychiatr Epidemiol. 2015;50(6):851-866. https://doi.org/10.1007/s00127-015-1062-x
118. Morales S, Taber-Thomas BC, Pérez-Edgar KE. Patterns of attention to threat across tasks in behaviorally inhibited children at risk for anxiety. Dev Sci. Published online January 19, 2017. https://doi.org/10.1111/desc.12391
119. Wise S, Huang-Pollock C, Pérez-Edgar K. Implementation of the diffusion model on dot-probe task performance in children with behavioral inhibition. Psychol Res. Published online 2021. https://doi.org/10.1007/s00426-021-01532-3
120. Anaya B, Vallorani AM, Pérez-Edgar K. Individual dynamics of delta–beta coupling: using a multilevel framework to examine inter- and intraindividual differences in relation to social anxiety and behavioral inhibition. J Child Psychol Psychiatry. 2021;62(6):771-779. https://doi.org/10.1111/jcpp.13319
121. Anaya B, Vallorani A, Pérez-Edgar K. Dyadic behavioral synchrony between behaviorally inhibited and non-inhibited peers is associated with concordance in EEG frontal alpha asymmetry and delta-beta coupling. Biol Psychol. 2021;159:108018. https://doi.org/10.1016/j.biopsycho.2021.108018
122. Liu P, Taber-Thomas BC, Fu X, Pérez-Edgar KE. Biobehavioral markers of attention bias modification in temperamental risk for anxiety: a randomized control trial. J Am Acad Child Adolesc Psychiatry. 2018;57(2):103-110. https://doi.org/10.1016/j.jaac.2017.11.016
123. Suarez GL, Morales S, Metcalf K, Pérez-Edgar KE. Perinatal complications are associated with social anxiety: indirect effects through temperament. Infant Child Dev. 2019;28(3):e2130. https://doi.org/10.1002/icd.2130
124. Thai N, Taber-Thomas BC, Pérez-Edgar KE. Neural correlates of attention biases, behavioral inhibition, and social anxiety in children: an ERP study. Dev Cogn Neurosci. 2016;19:200-210. https://doi.org/10.1016/j.dcn.2016.03.008
125. Salum GA, Gadelha A, Pan PM, *et al*. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. Int J Methods Psychiatr Res. 2015;24(1):58-73. https://doi.org/10.1002/mpr.1459
126. Fox NA, Henderson HA, Rubin KH, Calkins SD, Schmidt LA. Continuity and discontinuity of behavioral inhibition and exuberance: psychophysiological and behavioral influences across the first four years of life. Child Dev. 2001;72(1):1-21. https://doi.org/10.1111/1467-8624.00262
127. White T, Muetzel RL, El Marroun H, *et al*. Paediatric population neuroimaging and the Generation R Study: the second wave. Eur J Epidemiol. 2018;33(1):99-125. https://doi.org/10.1007/s10654-017-0319-y
128. Jansen PW, Raat H, Mackenbach JP, *et al*. Socioeconomic inequalities in infant temperament. Soc Psychiatry Psychiatr Epidemiol. 2009;44(2):87-95. https://doi.org/10.1007/s00127-008-0416-z
129. Jaddoe VWV, van Duijn CM, van der Heijden AJ, *et al*. The Generation R Study: design and cohort update until the age of 4 years. Eur J Epidemiol. 2008;23(12):801. https://doi.org/10.1007/s10654-008-9309-4
130. Pappa I, Mileva-Seitz VR, Szekely E, *et al*. DRD4 VNTRs, observed stranger fear in preschoolers and later ADHD symptoms. Psychiatry Res. 2014;220(3):982-986. https://doi.org/10.1016/j.psychres.2014.09.004

131. Hur J, Smith JF, DeYoung KA, *et al*. Anxiety and the neurobiology of temporally uncertain threat anticipation. J Neurosci. 2020;40(41):7949LP-7964LP. https://doi.org/10.1523/JNEUROSCI.0704-20.2020
132. Hur J, DeYoung KA, Islam S, Anderson AS, Barstead MG, Shackman AJ. Social context and the real-world consequences of social anxiety. Psychol Med. 2020;50(12):1989-2000. https://doi.org/10.1017/S0033291719002022
133. Grogans SE, Hur J, Barstead MG, *et al*. Neuroticism/negative emotionality is associated with increased reactivity to uncertain threat in the bed nucleus of the stria terminalis, not the amygdala. J Neurosci. Published online July 15, 2024:e1868232024. https://doi.org/10.1523/jneurosci.1868-23.2024
134. Hur J, Kuhn M, Grogans SE, *et al*. Anxiety-related frontocortical activity is associated with dampened stressor reactivity in the real world. Psychol Sci. Published online June 3, 2022:09567976211056635. https://doi.org/10.1177/09567976211056635
135. Hur J, Tillman RM, Kim HC, *et al*. Adolescent social anxiety is associated with diminished discrimination of anticipated threat and safety in the bed nucleus of the stria terminalis. J Psychopathol Clin Sci. 2025;134(1):41-56. https://doi.org/10.1037/abn0000940
136. Tyborowska A, Volman I, Niermann HCM, *et al*. Early-life and pubertal stress differentially modulate grey matter development in human adolescents. Sci Rep. 2018;8(1):9201. https://doi.org/10.1038/s41598-018-27439-5
137. Van Bakel HJA, Riksen-Walraven JM. Stress reactivity in 15-month-old infants: links with infant temperament, cognitive competence, and attachment security. Dev Psychobiol. 2004;44(3):157-167. https://doi.org/10.1002/dev.20001
138. Van Bakel HJA, Riksen-Walraven JM. Parenting and development of one-year-olds: links with parental, contextual, and child characteristics. Child Dev. 2002;73(1):256-273. https://doi.org/10.1111/1467-8624.00404
139. Hill SY, Lowers L, Locke J, Snidman N, Kagan J. Behavioral inhibition in children from families at high risk for developing alcoholism. J Am Acad Child Adolesc Psychiatry. 1999;38(4):410-417. https://doi.org/10.1097/00004583-199904000-00013
140. Hill SY, Shen S, Lowers L, Locke-Wellman J, Matthews AG, McDermott M. Psychopathology in offspring from multiplex alcohol dependence families with and without parental alcohol dependence: a prospective study during childhood and adolescence. Psychiatry Res. 2008;160(2):155-166. https://doi.org/10.1016/j.jpsychires.2010.08.005
141. Hill SY, Tessner KD, McDermott MD. Psychopathology in offspring from families of alcohol dependent female probands: a prospective study. J Psychiatr Res. 2011;45(3):285-294. https://doi.org/10.1016/j.jpsychires.2010.08.005
142. Battaglia M, Zanoni A, Taddei M, *et al*. Cerebral responses to emotional expressions and the development of social anxiety disorder: a preliminary longitudinal study. Depress Anxiety. 2012;29(1):54-61. https://doi.org/10.1002/da.20896
143. Battaglia M, Ogliari A, Zanoni A, *et al*. Influence of the serotonin transporter promoter gene and shyness on children's cerebral responses to facial expressions. Arch Gen Psychiatry. 2005;62(1):85-94. https://doi.org/10.1001/archpsyc.62.1.85
144. Battaglia M, Ogliari A, Zanoni A, *et al*. Children's discrimination of expressions of emotions: relationship with indices of social anxiety and shyness. J Am Acad Child Adolesc Psychiatry. 2004;43(3):358-365. https://doi.org/10.1097/01.chi.0000103177.91435.1b
145. Taddei M, Tettamanti M, Zanoni A, Cappa S, Battaglia M. Brain white matter organisation in adolescence is related to childhood cerebral responses to facial expressions and harm avoidance. Neuroimage. 2012;61(4):1394-1401. https://doi.org/10.1016/j.neuroimage.2012.03.062
146. Gold AL, Steuber ER, White LK, *et al*. Cortical thickness and subcortical gray matter volume in pediatric anxiety disorders. Neuropsychopharmacology. 2017;42:2423-2433. https://doi.org/10.1038/npp.2017.83
147. Gold AL, Abend R, Britton JC, *et al*. Age differences in the neural correlates of anxiety disorders: an fMRI study of response to learned threat. Am J Psychiatry. 2020;177(5):454-463. https://doi.org/10.1176/appi.ajp.2019.19060650
148. Linke JO, Abend R, Kircanski K, *et al*. Shared and anxiety-specific pediatric psychopathology dimensions manifest distributed neural correlates. Biol Psychiatry. 2021;89(6):579-587. https://doi.org/10.1016/j.biopsych.2020.10.018
149. Dyson MW, Klein DN, Olino TM, Dougherty LR, Durbin CE. Social and non-social behavioral inhibition in preschool-age children: differential associations with parent-reports of temperament and anxiety. Child Psychiatry Hum Dev. 2011;42(4):390-405. https://doi.org/10.1007/s10578-011-0225-6
150. Dyson MW, Olino TM, Durbin CE, *et al*. The structural and rank-order stability of temperament in young children based on a laboratory-observational measure. Psychol Assess. 2015;27(4):1388-1401. https://doi.org/10.1037/pas0000104
151. Hubachek S, Botdorf M, Riggins T, Leong HC, Klein DN, Dougherty LR. Hippocampal subregion volume in high-risk offspring is associated with increases in depressive symptoms across the transition to adolescence. J Affect Disord. 2021;281:358-366. https://doi.org/10.1016/j.jad.2020.12.017
152. Klein DN, Finsaas MC. The Stony Brook Temperament Study: early antecedents and pathways to emotional disorders. Child Dev Perspect. 2017;11(4):257-263. https://doi.org/10.1111/cdep.12242
153. Kopala-Sibley DC, Cyr M, Finsaas MC, *et al*. Early childhood parenting predicts late childhood brain functional connectivity during emotion perception and reward processing. Child Dev. 2020;91(1):110-128. https://doi.org/10.1111/cdev.13126
154. Laptook RS, Klein DN, Olino TM, Dyson MW, Carlson G. Low positive affectivity and behavioral inhibition in preschool-age children: a replication and extension of previous findings. Pers Individ Diff. 2010;48(5):547-551. https://doi.org/10.1016/j.paid.2009.12.003
155. Mumper EE, Dyson MW, Finsaas MC, Olino TM, Klein DN. Life stress moderates the effects of preschool behavioral inhibition on anxiety in early adolescence. J Child Psychol Psychiatry. 2020;61(2):167-174. https://doi.org/10.1111/jcpp.13121
156. Olino TM, Klein DN, Dyson MW, Rose SA, Durbin CE. Temperamental emotionality in preschool-aged children and depressive disorders in parents: associations in a large community sample. J Abnorm Psychol. 2010;119(3):468. https://doi.org/10.1037/a0020112
157. Stumper A, Danzig AP, Dyson MW, Olino TM, Carlson GA, Klein DN. Parents' behavioral inhibition moderates association of preschoolers' BI with risk for age 9 anxiety disorders. J Affect Disord. 2017;210:35-42. https://doi.org/10.1016/j.jad.2016.12.008
158. White LK, Degnan KA, Henderson HA, *et al*. Developmental relations among behavioral inhibition, anxiety, and attention biases to threat and positive information. Child Dev. 2017;88(1):141-155. https://doi.org/10.1111/cdev.12696
159. Hane AA, Fox NA, Henderson HA, Marshall PJ. Behavioral reactivity and approach-withdrawal bias in infancy. Dev Psychol. 2008;44(5):1491. https://doi.org/10.1037/a0012855
160. Harrewijn A, Abend R, Linke J, *et al*. Combining fMRI during resting state and an attention bias task in children. Neuroimage. 2020;205:116301. https://doi.org/10.1016/j.neuroimage.2019.116301
161. Thornton-Wells TA, Avery SN, Blackford JU. Using novel control groups to dissect the amygdala's role in Williams syndrome. Dev Cogn Neurosci. 2011;1(3):295-304. https://doi.org/10.1016/j.dcn.2011.03.003
162. Carney DM, Moroney E, Machlin L, *et al*. The Twin Study of Negative Valence Emotional Constructs. Twin Res Hum Genet. 2016;19(5):456-464. https://doi.org/10.1017/thg.2016.59
163. Hettema JM, Bourdon JL, Sawyers C, *et al*. Genetic and environmental risk structure of internalizing psychopathology in youth. Depress Anxiety. 2020;37(6):540-548. https://doi.org/10.1002/da.23024
164. Lee M, Aggen SH, Carney DM, *et al*. Latent structure of negative valence measures in childhood. Depress Anxiety. 2017;34(8):742-751. https://doi.org/10.1002/da.22656
165. Liu P, Vandermeer MRJ, Mohamed Ali O, *et al*. Maternal depression, child temperament, and early-life stress predict never-depressed preadolescents' functional connectivity during a negative-mood induction. Clin Psychol Sci. Published online June 3, 2021:21677026211016420. https://doi.org/10.1177/21677026211016419
166. Liu P, Kryski KR, Smith HJ, Joanisse MF, Hayden EP. Transactional relations between early child temperament, structured parenting, and child outcomes: a three-wave longitudinal study. Dev Psychopathol. 2020;32(3):923-933. https://doi.org/10.1017/S0954579419000841
167. Liu P, Vandermeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Neural activity during self-referential processing in children at risk for depression. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5(4):429-437. https://doi.org/10.1016/j.bpsc.2019.12.012
168. Liu P, Vandemeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Depressogenic self-schemas are associated with smaller regional grey matter volume in never-depressed preadolescents. Neuroimage Clin. 2020;28:102422. https://doi.org/10.1016/j.nicl.2020.102422
169. Vandermeer MRJ, Liu P, Mohamed Ali O, *et al*. Orbitofrontal cortex grey matter volume is related to children's depressive symptoms. Neuroimage Clin. 2020;28:102395. https://doi.org/10.1016/j.nicl.2020.102395
170. Vandermeer MRJ, Sheikh HI, Singh SS, *et al*. The BDNF gene val66met polymorphism and behavioral inhibition in early childhood. Soc Dev. 2018;27(3):543-554. https://doi.org/10.1111/sode.12292
171. Schmidt NL, Lemery-Chalfant K, Goldsmith HH. Wisconsin Twin Project Overview: temperament and affective neuroscience. Twin Res Hum Genet. 2019;22(6):794-799. https://doi.org/10.1017/thg.2019.108
172. Schmidt NL, Brooker RJ, Carroll IC, *et al*. Longitudinal research at the interface of affective neuroscience, developmental psychopathology, health and behavioral genetics: findings from the Wisconsin Twin Project. Twin Res Hum Genet. 2019;22(4):233-239. https://doi.org/10.1017/thg.2019.55
173. Goldsmith HH, Lemery-Chalfant K, Schmidt NL, Arneson CL, Schmidt CK. Longitudinal analyses of affect, temperament, and childhood psychopathology. Twin Res Hum Genet. 2007;10(1):118-126. https://doi.org/10.1375/twin.10.1.118

174. Planalp EM, Goldsmith HH. Observed profiles of infant temperament: stability, heritability, and associations with parenting. Child Dev. 2020;91(3):e563-e580. https://doi.org/10.1111/cdev.13277
175. Zugman A, Harrewijn A, Cardinale EM, *et al*. Mega-analysis methods in ENIGMA: the experience of the Generalized Anxiety Disorder Working Group. Hum Brain Mapp. 2022;43(1):255-277. https://doi.org/10.1002/hbm.25096
176. Gorgolewski KJ, Auer T, Calhoun VD, *et al*. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016;3. https://doi.org/10.1038/sdata.2016.44
177. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One. 2017;12(9):e0184661. https://doi.org/10.1371/journal.pone.0184661
178. Fischl B, Salat DH, Busa E, *et al*. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33(3):341-355. https://doi.org/10.1016/s0896-6273(02)00569-x
179. Winkler AM, Sabuncu MR, Yeo BTT, *et al*. Measuring and comparing brain cortical surface area and other areal quantities. Neuroimage. 2012;61(4):1428-1443. https://doi.org/10.1016/j.neuroimage.2012.03.026
180. Pesarin F, Salmaso L. Permutation Tests for Complex Data: Theory, Applications and Software. Wiley; 2010.
181. Winkler AM, Ridgway GR, Douaud G, Nichols TE, Smith SM. Faster permutation inference in brain imaging. Neuroimage. 2016;141:502-516. https://doi.org/10.1016/j.neuroimage.2016.05.068
182. Westfall PH, Young SS. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley; 1993.
183. Desikan RS, Ségonne F, Fischl B, *et al*. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968-980. https://doi.org/10.1016/j.neuroimage.2006.01.021
184. Winkler AM, Webster MA, Brooks JC, Tracey I, Smith SM, Nichols TE. Non-parametric combination and related permutation tests for neuroimaging. Hum Brain Mapp. 2016;37(4):1486-1511. https://doi.org/10.1002/hbm.23115
185. Larivière S, Paquola C, Park B yong, *et al*. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. Nat Methods. 2021;18(7):698-700. https://doi.org/10.1038/s41592-021-01186-4
186. Teng F, Wang M, Lu Z, *et al*. Causal relationship between cortical structural changes and onset of anxiety disorder: evidence from Mendelian randomization. Cereb Cortex. 2024;34(11). https://doi.org/10.1093/cercor/bhae440
187. Brühl AB, Hänggi J, Baur V, *et al*. Increased cortical thickness in a frontoparietal network in social anxiety disorder. Hum Brain Mapp. 2014;35(7):2966-2977. https://doi.org/10.1002/hbm.22378
188. Zhao K, Liu H, Yan R, *et al*. Cortical thickness and subcortical structure volume abnormalities in patients with major depression with and without anxious symptoms. Brain Behav. 2017;7(8):e00754. https://doi.org/10.1002/brb3.754
189. Sylvester CM, Corbetta M, Raichle ME, *et al*. Functional network dysfunction in anxiety and anxiety disorders. Trends Neurosci. 2012;35(9):527-535. https://doi.org/10.1016/j.tins.2012.04.012
190. Bas-Hoogendam JM, van Steenbergen H, Cohen Kadosh K, Westenberg PM, van der Wee NJA. Intrinsic functional connectivity in families genetically enriched for social anxiety disorder—an endophenotype study. EBioMedicine. 2021;69. https://doi.org/10.1016/j.ebiom.2021.103445
191. He Y, Xu T, Zhang W, Zuo XN. Lifespan anxiety is reflected in human amygdala cortical connectivity. Hum Brain Mapp. 2016;37(3):1178-1193. https://doi.org/10.1002/hbm.23094
192. Liao W, Chen H, Feng Y, *et al*. Selective aberrant functional connectivity of resting state networks in social anxiety disorder. Neuroimage. 2010;52(4):1549-1558. https://doi.org/10.1016/j.neuroimage.2010.05.010
193. De la Peña-Arteaga V, Chavarría-Elizondo P, Juaneda-Seguí A, *et al*. Trait anxiety is associated with attentional brain networks. Eur Neuropsychopharmacol. 2024;83:19-26. https://doi.org/10.1016/j.euroneuro.2024.02.013
194. Sylvester CM, Smyser CD, Smyser T, *et al*. Cortical functional connectivity evident after birth and behavioral inhibition at age 2. Am J Psychiatry. 2018;175(2):180-187. https://doi.org/10.1176/appi.ajp.2017.17010018
195. Filippi CA, Winkler AM, Kanel D, *et al*. Neural correlates of novelty-evoked distress in 4-month-old infants: a synthetic cohort study. Biol Psychiatry Cogn Neurosci Neuroimaging. Published online April 18, 2024. https://doi.org/10.1016/j.bpsc.2024.03.008
196. Zhong S, Lin J, Zhang L, *et al*. Neural correlates of harm avoidance: a multimodal meta-analysis of brain structural and resting-state functional neuroimaging studies. Transl Psychiatry. 2024;14(1). https://doi.org/10.1038/s41398-024-03091-8
197. Jahanshad N, Lenzini P, Bijsterbosch J. Current best practices and future opportunities for reproducible findings using large-scale neuroimaging in psychiatry. Neuropsychopharmacology. Published online 2024. https://doi.org/10.1038/s41386-024-01938-8
198. Scheel AM, Schijen MRMJ, Lakens D. An excess of positive results: comparing the standard psychology literature with registered reports. Adv Methods Pract Psychol Sci. 2021;4(2). https://doi.org/10.1177/25152459211007467
199. Stein DJ, van Honk J, Thompson PM. The importance and limitations of null findings. Am J Psychiatry. 2024;181(12):1033-1035. https://doi.org/10.1176/appi.ajp.20240957
200. Bas-Hoogendam JM. Genetic vulnerability to social anxiety disorder. In: Blackford JU, Milad MR, eds. Current Topics in Behavioral Neurosciences. Springer; 2024. https://doi.org/10.1007/7854_2024_544
201. Xu B, Wang H, Dall'Aglio L, *et al*. Beyond out-of-sample: robust and generalizable multivariate neuroanatomical patterns of psychiatric problems in youth. Mol Psychiatry. Published online 2024. https://doi.org/10.1038/s41380-024-02855-4
202. Casey BJ, Cannonier T, Conley MI, *et al*. The Adolescent Brain Cognitive Development (ABCD) Study: imaging acquisition across 21 sites. Dev Cogn Neurosci. 2018;32:43-54. https://doi.org/10.1016/j.dcn.2018.03.001
203. Ide JS, Li HT, Chen Y, *et al*. Gray matter volumetric correlates of behavioral activation and inhibition system traits in children: an exploratory voxel-based morphometry study of the ABCD project data. Neuroimage. 2020;220. https://doi.org/10.1016/j.neuroimage.2020.117085
204. Hammoud RA, Ammar LA, McCall SJ, Shamseddeen W, Elbejjani M. Brain volumes, behavioral inhibition, and anxiety disorders in children: results from the Adolescent Brain Cognitive Development Study. BMC Psychiatry. 2024;24(1):257. https://doi.org/10.1186/s12888-024-05725-z
205. Fleece H, Teglasi H. Behavioral inhibition and social competence through the eyes of parent and teacher informants. Behav Sci. 2024;14(11). https://doi.org/10.3390/bs14111080
206. Mills KL, Siegmund KD, Tamnes CK, *et al*. Inter-individual variability in structural brain development from late childhood to young adulthood. Neuroimage. 2021;242:118450. https://doi.org/10.1016/j.neuroimage.2021.118450
207. Lorenzo NE, Novick DR, Seddio K, *et al*. Bidirectional and interactive effects of child temperament and parenting in early childhood on the trajectory of social anxiety in adolescence. Depress Anxiety. 2022;39(3):192-200. https://doi.org/10.1002/da.23224
208. Phillips EM, Brock RL. Interplay between interparental relationship quality and infant temperament predicts toddler transdiagnostic psychopathology. J Fam Psychol. Published online August 5, 2024. https://doi.org/10.1037/fam0001263
209. Verhagen M, Derks M, Roelofs K, Maciejewski D. Behavioral inhibition, negative parenting, and social withdrawal: longitudinal associations with loneliness during early, middle, and late adolescence. Child Dev. 2023;94(2):512-528. https://doi.org/10.1111/cdev.13874

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

Supplemental Materials Bas-Hoogendam et al. // JAACAP 2025

Contents

Supplement 1: Supplemental Information for included samples

BRAINS sample

Publications: [1–10]

In- and exclusion criteria [1]

The sample consisted of 9–12-year-olds 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) [11]. 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 [12]. 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 [1,8]

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; [13]). 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) [14], 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 [14] and community samples[15] and it offers a valuable tool to predict specific anxiety disorders in clinically-referred youths [16]. It had good internal consistency in the present study (Cronbach's alpha = .90).

Measures of temperament[8]

Behavioral inhibition (BI) was assessed using the Behavioral Inhibition Questionnaire (BIQ) [11], 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 [11], parent reports on the BIQ correlate with laboratory observations of BI in social contexts [17], and the BIQ had good internal consistency in the present study (Cronbach's alpha = .91).

In the present study, we used the total score on the BIQ as the indicator of cIT.

Brazilian High Risk Cohort (BHRC)

Publications: [18,19]

In- and exclusion criteria [18,20,21]

As summarized in Axelrud and colleagues [20], "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.[21], "mental disorders were assessed using the Brazilian Portuguese version [22] of the DAWBA [23]. 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 [23]. 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 [24]. Diagnoses are related to diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders, 4th edition."

Measure of temperament [21]

Temperament was assessed with the Brazilian-Portuguese self-report version of the early adolescent temperament questionnaire (EATQ-R)[24,25]. 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 [26,27].

In the present study, we used the sum score of the shyness items of the EATQ-R as an index of cIT.

Cohort 3/4

Publications: [28–30]

In- and exclusion criteria

As described by Shechner et al.[28], "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 [30] . 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) [31]. Anxiety levels were assessed in two ways: (1) the Beck Anxiety Inventory (BAI) [32] was used as a measure of trait anxiety, and (2) the State subscale of the State Trait Anxiety Inventory (STAI-S) [33] 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 [30]. Behavioral scores were standardized at each time point. Mothers also reported their child's social fear at 14 and 24 months using the Toddler Behavior Assessment Questionnaire [34].

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 [35].

Generation R

Publications: [36–40]

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) [38], 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 [37]. As described in White et al (2018) [38], "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 [41]. 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 [41]. 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 [38], 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 [42], 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 [40]***)*** "At the age of six months, infant temperament was assessed using an adapted version of the infant behavior questionnaire-revised (IBQ-R) [43]. A detailed description of the changes has previously been described [44]. 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 [43]."

Age 36 months [39] 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) [45].

In the present mega-analysis, we used the scores on the Lab-TAB as index of cIT (sample with behavioral observations). For participants without these scores, we used the IBQ- scores on the Fear subscale (sample with questionnaire data).

Maryland – PAX sample

Publications: [46–49]

In- and exclusion criteria

This prospective-longitudinal study focused on the emergence of anxiety disorders and depression and is described in several recent publications [46–49]. 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 [50].

Diagnostic interview: Structured Clinical Interview for DSM-5 (SCID-5-RV) [50].

Measures of temperament

Participants completed the Adult Measure of Behavioral Inhibition (AMBI) and the Retrospective Measure of Behavioral Inhibition (RMBI) [51]. 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 Maryland-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 used the total score for the RMBI as the cIT index.

Maryland – TAX sample

Publications [52]

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-11 [53]) 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-11 [53]; 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 [53] *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 [54]) 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) [65]and the BIQ-A [11]. In addition, a caregiver completed the CSRI and RSRI for their child. All five measures were highly correlated ($r > 0.55$, $p < 0.001$).

For the current mega-analysis, we used the total score for the adolescent-reported RSRI as the cIT index.

Nijmegen Longitudinal Study on Child and Infant Development

Publications [55–57]

In- and exclusion criteria at age 15 months

(As described in Van Bakel et al.[57]): "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 [58], 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 a15-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. [55]

"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. [55], "internalizing symptoms at age 17 were measured using the Child Behaviour Checklist (CBCL) [59]. 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 [57] "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; [60]) 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 [56]: "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) [61] 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 asses additional behavioral measures of infant fearfulness during the stranger/robot episode, the occurrence of three infant behaviors (adapted from Nachmias et al. [62]) 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 ($p < .05$)."

For the current project, we created a sum score (6 items) of the three behaviors described above (looking/referencing to parent, proximity seeking/maintaining physical contact with parent, and crying) within the two situations (stranger and robot), and used this as an index of cIT.

Pittsburgh sample

Publications: [63–66]

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).

The high-risk families were identified through a proband pair of alcohol dependent brothers or pair of sisters as previously described [63,65]. Both members of the proband pair were screened using an in-person structured interview (Diagnostic Interview Schedule; DIS [67]) 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)[68] 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 [63,65]. 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) [69] and CIDI-Substance Abuse Module (CIDI-SAM)[70], 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 [71]

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 [72]. 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 [72]. 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 [72,73].

Adolescent temperament measures (around time of scan) [71] As part of the longitudinal follow-up, subjects were administered the Multidimensional Personality Questionnaire (MPQ) [74]. The MPQ provides 11 personality scales and 3 higher order scales. Assessment was completed within 1 year of the scans.

San Raffaele

Publications: [75–78]

In- and exclusion criteria (from [75])

"Subjects were drawn from a sample of 49 normally developing children who had participated in an ERP study and shyness at age 8–9 [77]. The 49 ERP study participants had been drawn from a general population cohort (n = 149) assessed at age 7 for shyness [76]. 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 [79] were followed and applied rigorously." [75]

Measures of temperament

As described in the baseline-paper of this longitudinal study[76], 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 [80,81]. Three different scales were used in the questionnaire: (1) the Liebowitz Social Anxiety Scale adapted for children (LSAS) [82,83] 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 (Cooper & Eke, 1999a; Stevenson-Hinde & Glover, 1996) 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 [86], 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 [82], 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 [83] and (2) clarifying the presence of social anxiety disorder as the salient clinical feature of childhood selective mutism [82].

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 [87]. 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 [87]. 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 [85].

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 [86], is relatively stable from childhood to adult life [88], and heightens the risk of developing symptoms of anxiety/depression for people with extreme scores on the HA scale [89].

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 ± 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 used an empirical composite index of cIT 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 [77].

SDAN

Publications: [90–92]

In- and exclusion criteria (from [91,92]); 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) [79]. 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) [14].

Measures of temperament

Behavioral Inhibition Questionnaire [93].

Stony Brook Temperament Study

Publications: [17,94–98]

In- and exclusion criteria (described in [95])

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 [99]). 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. [100] for details). Negative and positive emotionality as well as behavioral inhibition were assessed via the Laboratory Temperament Assessment Battery (LabTAB) [35], which involves a standardized set of tasks designed to elicit children's bodily, vocal, and facial expressions of a range of emotions (see [101])."

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 [102]; parents and youth completed the K-SADS at ages 9, 12, and 15.

Measures of temperament

(Quoted from [101]): "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 range of temperament-relevant behaviors. Eleven episodes were from the Laboratory Temperament Assessment Battery (Lab-TAB) [35] 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 [103]. 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 [104]. 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 [35]. 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 [105], 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 [11].

For the present mega-analysis, we used a sum score (log-transformed) from three Kagan-like tasks in Goldsmith's Lab-TAB as an index of cIT.

TOTS

Publications: [106,107]

In- and exclusion criteria [106,108]

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 [30,109]). 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; 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; 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; 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 [30].

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 [16,110].

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, puppets) [111]. Furthermore, inhibited behavior to novel stimuli was coded at 24 and 36 months [30]. Behavioral scores were standardized at each time point.

For the present study, we used the 'BI-classic-24' index, composed of scores on the stranger, robot and tunnel episodes of the Laboratory Temperament Assessment Battery – preschool version [35].

Vanderbilt – children sample

Publications: [112] and unpublished

In- and exclusion criteria

Quoted from Clauss et al [112]: "Consistent with the extreme discordant phenotypes approach [113], 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) [79] 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 [114]. Handedness was assessed using the Edinburgh Handedness Inventory [115].

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) [11], a validated measure of childhood inhibited temperament, which shows convergent validity with behavioral measures and other measures of social inhibition [12,17].

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)[11]; these norms were similar to those identified in children and adolescents 4 to 15 years of age and those used in a recent similar neuroimaging study [8]."

Diagnostic interview and psychiatric symptom measures

Kiddie-Schedule for Affective Disorders and Schizophrenia–Present and Lifetime Version (KSADS-PL) [79] - 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 [110] and the Children's Depression Inventory [116].

Measures of temperament

Children completed a self-report of temperament, the Behavioral Inhibition Questionnaire-Child (BIQ-C). [12]. The total score was used as an index of cIT in the present mega-analysis.

Vanderbilt – young adults sample

Publications: [117–121]

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. [120], subjects were recruited by seeking individuals ages 18–25 who were "extremely shy or outgoing." Consistent with prior studies [118,119,122], 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) [123]. 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 [124].

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) [123], respectively. Both questionnaires have excellent reliability (Cronbach's alpha = 0.79 for the RSRI and 0.78 for the ASRI), demonstrate convergent validity [123,125] and minimize self-report bias by focusing on reports of concrete behaviors in specific situations instead of subjective feelings[126].

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 used the total scores on the RSRI as index of cIT.

Virginia Commonwealth University Juvenile Anxiety Study (VCUJAS)

Publications:[127–130]

Study design and in- and exclusion criteria

As described in Carney et al. (2016)[128], "the Twin Study of Negative Valence Emotional Constructs is a multi-site study designed to examine the relationship between a broad selection of potential measures designed to assess putative endophenotypes for negative valence systems (NVS) and early symptoms of internalizing disorders (IDs). Recruitment of twin pairs occurred through the Mid- Atlantic Twin Registry (MATR), a Virginia Commonwealth University (VCU) database comprised of twins, other higher order multiples, and their family members who were willing to consider participating in research. Twin pairs were considered ineligible for the study if either child met any of the following exclusion criteria: severe or unstable medical or neurological illness, past seizures without a clear or resolved etiology, intellectual disabilities, substance abuse, recent thoughts of suicide or homicide, episodes of psychosis, or currently taking psychotropic medications besides stimulants for ADHD. These exclusions aimed to (1) decrease the risk of exacerbation of possible medical conditions during the potentially stressful laboratory tasks and (2) minimize the likelihood that physiological responses recorded during the tasks were confounded by the potential effects of these variables. This study was approved by the VCU Institutional Review Board. Enrolled twins completed various dimensional self-report measures along with cognitive, emotional, and psychophysiological tasks designed to assess NVS function. Parents also completed surveys about their twins and themselves. In addition, a subset of the twins also participated in a neuroimaging protocol."

Diagnostic interview

Quoted from the design paper by Carney et al (2016)[128]: "A masters or doctoral-level trained clinician administered the Kiddie Schedule for Affective Disorders and Schizophrenia-Present & Lifetime Version (KSADS-PL [79]; about each twin during a face-to-face interview with the parent(s); the children were also assessed with the KSADS-PL at the NIMH as part of their standard protocol."

Measure of temperament

Parents completed the Retrospective Behavioral Inhibition Questionnaire (RBIQ) [123] about each of their twin children.

Western University sample

Publications: [131–137]

In- and exclusion criteria

As described in Vandermeer et al., 2020 [136], 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[138]. 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 [131]. 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

(CDI; [139]; $\alpha = 0.83$) and the Youth Self-Report [140] with the help of trained graduate students in clinical psychology.

Measures of temperament

A detailed description of the assessment of temperament is provided in previous publications on the study, for example Vandermeer et al.[135] and Liu et al.[131] . 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; [35]) and an age-adapted version of the Lab-TAB [103]. 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 [131,135]. 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., [96,141]).

These age 3 scores were used as an index of cIT in the present mega-analysis.

Wisconsin Twin Project

Publications: [142–145]

Study design (from [142])

"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 [146–149]. Across 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 [146,148]. 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. [142], "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, frustrative

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 Moore et al [150]. "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 [151] 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 [152]. 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 [153].

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 [150]. In **Stranger approach**, social interaction with an unfamiliar adult wearing a hat and sunglasses is investigated [152].

Post-visit observer ratings (from Moore et al.[150]). *"*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; [154]). From Moore et al[150]: "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 *r*s ranged from .27 to .53, all *p*s < .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.[152], 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 [155]. 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 [155] 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 two 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): $r = 0.30$, $p < 0.001$).

Supplement 2: 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.[156]. This concerned the following questionnaires with respect to anxiety disorders: the Hamilton Anxiety Rating Scale [157], Penn State Worry Questionnaire [158], Generalized Anxiety Disorder 7-Item questionnaire [159], State Trait Anxiety Inventory [33], Anxiety Sensitivity Index [160], Beck Anxiety Inventory [32], Liebowitz Social Anxiety Scale [161], Panic and Agoraphobia Scale [162], Agoraphobic Cognition Questionnaire [163], Panic Disorder Severity Scale [164] and the Screen Child Anxiety Related Disorders [14].

Furthermore, we asked for data on the Beck Depression Inventory II [165], the Children's Depression Inventory [166], and Hamilton Depression Rating Scale [167].

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 **Table S3a and Table S3b.**

Exploratory analyses

In exploratory analyses, we examined amygdalar subnuclei (20 in total), thalamic subnuclei (52 in total) and hippocampal subfields (44 in total), partial volume effects (PVEs; amount of gray and white matter) within 16 subcortical areas, and volumes of 12 subcortical limbic structures (hypothalamus, mammillary bodies, basal forebrain, septal nuclei, nucleus accumbens, and fornix; all bilateral). Independent variables (two models, one without and the second with global brain measures), contrasts, variance groups, and exchangeability blocks were the same as for the main analyses.

Correction for multiple testing for these exploratory analyses was done using the same structure as for the analyses described in the main manuscript, considering 20 amygdala

subnuclei, 52 thalamic subnuclei, 44 hippocampal subfields, PVE effects within 16 subcortical structures and volumes of 12 additional subcortical limbic structures.

Supplement 3: Supplemental Results

Exploratory analyses in full mega-analytic sample

Exploratory analyses on PVE within subcortical structures revealed a cIT by age interaction on the amount of white matter within the left caudate ($p_{C\text{-}FWER} = 0.043$) and a positive cIT by sex interaction on the amount of white matter within the right putamen ($p_{M\text{-}FWER} = 0.029$; both in the model without global brain measures). These interactions did not survive the most stringent correction for multiple testing.

Other exploratory analyses (amygdala subnuclei; thalamic subnuclei; hippocampal subfields; volumes of hypothalamus mammillary bodies, basal forebrain, septal nuclei, nucleus accumbens and fornix) did not reveal significant effects of cIT, nor significant interactions.

Supplemental Results: Sensitivity analyses in subgroups

Subgroup 1 - cIT-assessment based on (early-life) behavioral observations

With more liberal statistical thresholding (M-FWER correction), analyses in the first subgroup analysis revealed interaction effects in the model with global brain measures: a cIT by age interaction was found when considering CSA of the left pars orbitalis ($p_{M\text{-}FWER} = 0.049$), as well as a cIT by sex by age interaction for CSA of the left superior parietal region ($p_{M\text{-}FWER} = 0.046$).

Exploratory analyses (again in model correcting for global brain measures) on PVE within subcortical structures revealed positive cIT by sex interactions with respect to the amount of white matter within the right putamen ($p_{M\text{-}FWER} = 0.034$), left hippocampus ($p_{M\text{-}FWER} = 0.019$) and left ventral DC ($p_{M\text{-}FWER} = 0.045$).

Subgroup 2 – cIT-assessment based on parental/teacher reports during childhood

Analyses in the second subgroup revealed multiple effects at the M-FWER or C-FWER-corrected significance level. To start, a negative cIT by sex interaction was present in the left inferior parietal region ($p_{\text{M-FWER}} = 0.037$; in line with the vertex-wise analyses reported in the main manuscript), the left middle temporal area ($p_{\text{M-FWER}} = 0.019$) and the left superior parietal region ($p_{\text{M-FWER}} = 0.016$) (all CT, model without global brain measures). Furthermore, exploratory analyses on thalamic subnuclei revealed a positive effect of cIT on volume of the right lateral geniculate nucleus ($p_{\text{M-FWER}} = 0.009$) and a cIT by age interaction in the right ventromedial thalamus ($p_{\text{M-FWER}} = 0.031$; both interactions in model with global brain measures). Analyses on PVE effects within subcortical structures revealed negative effects of cIT on the amount of white matter within the right putamen ($p_{\text{C-FWER}} = 0.045$), the right accumbens ($p_{\text{C-FWER}} = 0.043$), the left and the right ventral diencephalon ($p_{\text{C-FWER}} = 0.023$, $p_{\text{C-FWER}} = 0.01$, respectively).

Subgroup 3 – cIT-assessment based on self-report measures acquired during late childhood and adolescence

Exploratory analyses in the third subgroup revealed a positive effect of cIT on gray matter amount of the left pallidum, at the M-FWER corrected level ($p_{\text{M-FWER}} = 0.033$; model with global brain measures).

Subgroup 4 – samples in which temperament was assessed during childhood

For this subgroup, exploratory analyses revealed an interaction at the most stringent significance level: PVE analyses revealed a cIT by age interaction on the amount of gray matter within the left hippocampus ($p_{\text{MC-FWER}} = 0.036$); furthermore, a cIT by age interaction was found for the amount of gray matter within the right ventral diencephalon at a more

liberal level of testing ($p_{C\text{-}FWER}$ = 0.030; both interactions in model without global brain measures). When considering the amount of white matter in PVE analyses, we found cIT by age interactions for the left hippocampus ($p_{C\text{-}FWER}$ = 0.016), the left caudate ($p_{C\text{-}FWER}$ = 0.046) and the left ventral diencephalon ($p_{C\text{-}FWER}$ = 0.041).

Supplement 4: Supplemental Tables

Table S1 **STROBE checklist case-control studies**

Table S2 **Scanner characteristics per sample**

Table S3 **Clinical characteristics per sample**

Table S1: STROBE Statement

Checklist of items that should be included in reports of *case-control studies*

	Item No	Recommendation	Location in manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Title
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Synopsis
Introduction			
Background/rationale	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			
Study design	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	Registered Report-Methods
		(*b*) For matched studies, give matching criteria and the number of controls per case	*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[a]	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	Registered Report-Methods
		(*b*) Describe any methods used to examine subgroups and interactions	Registered Report-Methods
		(*c*) Explain how missing data were addressed	Registered Report-Methods
		(*d*) If applicable, explain how matching of cases and controls was addressed	*Not applicable*
		(*e*) Describe any sensitivity analyses	Registered Report - Methods

Results			
Participants	13[a]	(a) Report numbers of individuals at each stage of study—eg 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	Registered Report -Table 3
Descriptive data	14[a]	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest	Supplemental Materials -Tables 3A and 3B
Outcome data	15[a]	Report numbers in each exposure category, or summary measures of exposure	*Not applicable*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 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	Registered Report - Results
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Registered Report - Results
Discussion			
Key results	18	Summarise key results with reference to study objectives	Registered report - Discussion
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	Registered report - Discussion
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Registered report - Discussion
Generalisability	21	Discuss the generalisability (external validity) of the study results	Registered report - Discussion
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	Registered report - Acknowledgments

[a] Give information separately for cases and controls.

Table S2: Scanner characteristics per sample

Sample	Scanner type	Field-strength	Structural MRI scan
Brains study	Siemens Trio Siemens Prismafit	3 T	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 × 256matrix, T1 = 850 ms).
Brazilian High Risk Cohort	Signa HDx (São Paulo) GE Signa HD (Porto Alegre)	1.5 T	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).
Cohort 3 / 4	GE Healthcare MR750	3 T	High-resolution, T1-weighted structural imaging sequence (MPRAGE; sagittal acquisition; 176 slices; 1 mm^3 isotropic voxels; 256 x 256 matrix; flip angle = 7°; TR = 7.7 ms; TE = 3.42 ms; TI = 425 ms).
Generation R - sample with behavioral observations and Generation R - sample with questionnaire data	GE MR750W	3 T	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 220 x 220, slice thickness 1.0 mm, in-plane resolution (mm) 1.0 mm^2
Maryland-PAX	Siemens Magnetom TIM Trio	3 T	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 = 300x320; FOV= 240x256).
Maryland-TAX	Siemens Magnetom TIM Trio	3 T	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 × 0.449 mm; matrix = 512 × 512; FOV = 230 × 230).
Nijmegen Longitudinal Study	Siemens Magnetom Trio or PRISMA	3 T	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).
Pittsburgh	Siemens Trio scanner	3 T	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 mm^2, yielding 160 transversal slices with a thickness of 1.2 mm).
San Raffaele	Philips Achieva	3 T	T1-weighted scan, 150 axial slices, resolution 1mm x 1mm x 1mm
SDAN	General Electric MR750	3 T	Whole-brain, high-resolution, T1-weighted anatomical scan (MPRAGE; 176 axial slices, 256 x 256 matrix,1 mm^3 isotropic slices; FA = 7°, FOV = 220 mm; TR = 7.7ms, TE = 3.42s)
Stony Brook Temperament Study	Siemens Trio	3 T	T1-weighted high resolution structural images with the magnetization prepared rapid 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 × 1 × 1 millimeters
TOTS	General Electric MR750	3 T	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 1x 1mm).
Vanderbilt - children	Philips	3 T	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 x 3 mm in-plane resolution.
Vanderbilt - young adults	Philips	3 T	High resolution T1-weighted anatomical images (256mm FOV, 170 slices, 1-mm slice thickness, 0-mm gap)
VCU-JAS (NIMH sample)	General Electric MR750	3 T	High-resolution anatomical scan (1-mm slices, three-dimensional spoiled gradient-echo sequence, 7° flip angle, minimum full echo time, 2563256 matrix, 25.6-cm field of view)
Western University	Siemens Trio	3 T	T1-weighted 3D magnetization prepared rapid gradient echo (MPRAGE) sequence (1 x 1 x 1 mm), voxel size, TR = 2300 ms, TE = 2.98 ms, FOV = 256 mm), 192 slices.
Wisconsin Twin Project - RDoC twin study	GE SIGNA (Discovery MR750) scanner	3 T	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◦)

Abbreviations: FA =flip angle; FOV =field of view; T =Tesla; TE =echo time; TI =inversion time; TR =repetition time.

Table S3a: Clinical characteristics per sample // psychopathology

(after preprocessing and quality control– psychopathology at (around) time of scan[a])

			Psychopathology based on clinical interview (current / lifetime but not current)												
			Anxiety disorders						**Other psychopathology**					**Psychotropic medication at time scan**	
Sample	**Clinical interview**	***n* (*n* female)**	**SAD**	**PD**	**AG**	**GAD**	**SPH**	**Other anxiety disorders**	**MDD**	**OCD**	**PTSD**	**SUD**	**Other**	***n***	**Specification medication**
Brains study	C-DISC 4 administered to parents	129 (72)	4 SAD, 3 social phobia / *na*	*na*	*na*	4 / *na*	*na*	0	0 / *na*	*na*	*na*	*na*	13 current ADHD	9	*n* = 2 SSRI/SNRI *n* = 7 medication for ADHD
Brazilian High Risk Cohort	Development and Well-being Assessment (DAWBA)	502 (233)	5 / *na*	0 / *na*	0 / *na*	16 / *na*	26 / *na*	32 current anxiety nos	23 / *na*	2 / *na*	9 / *na*	*na*	*na*	14	*n* = 1 SSRI/SNRI *n* = 4 antipsychotics *n* = 9 other medication nos
Cohort 3 / 4	Structured Clinical Interview for DSM-IV TR	88 (50)	0 / 5	0 / 1	1 / 0	3 / 0	0 / 1	1 lifetime separation anxiety	1 / 5	0 / 0	0 / 0	1 / 5	1 current ADHD, 1 current ODD, 7 lifetime ADHD, 1 lifetime ADD, 1 lifetime ODD, 1 lifetime Tourettes	*na*	*na*
Generation R - sample with behavioral observations	Diagnostic Interview Schedule for Young Children (DISC-YC) (note: at age 5 – 8)	498 (248)	4	*na*	*na*	2	8	0	0	*0*	0	*na*	*na*	*na*	*na*
Generation R - sample with questionnaire data	Diagnostic Interview Schedule for Young Children (DISC-YC) (note: at age 5 – 8)	1604 (833)	12	*na*	*na*	4	26	5	2	1	0	*na*	*na*	*na*	*na*
Maryland-PAX	Structured Clinical Interview for DSM-5 (SCID-5-RV)	139 (81)	0 / 12	0 / 1	0 / 0	0 / 0	3 / 13	18 / 18	0 / 21	*na*	0 / 2	7 / 2	8 / 28	0	*na*
Maryland-TAX	Mini-International Neuropsychiatric Interview For Children And Adolescents (MINI-KID; Sheehan et al., 1998) based on DSM-IV and ICD-10 criteria	53 (28)	24 / *na*	4 / *na*	*na*	12 / *na*	4 / *na*	*na*	8 / *na*	*na*	3 / *na*	*na*	2 current ADHD	0	*na*

			Psychopathology based on clinical interview (current / lifetime but not current)												
			Anxiety disorders						Other psychopathology					Psychotropic medication at time scan	
Sample	**Clinical interview**	***n* (*n* female)**	**SAD**	**PD**	**AG**	**GAD**	**SPH**	**Other anxiety disorders**	**MDD**	**OCD**	**PTSD**	**SUD**	**Other**	***n***	**Specification medication**
Nijmegen Longitudinal Study	*na*	68 (31)	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*
Pittsburgh	K-SADS-PL administered to parent and child	15 (3)	0 /1	0 / 1	0 / 0	0 / 0	0 / 1	*n* = 4 simple phobia (lifetime); *n* = 1 overanxious disorder (lifetime)	0 / 1	0 / 1	0 / 0	1 / 3	*na*	0	*na*
San Raffaele	K-SADS-PL adminstered to parents	20 (8)	5 / *na*	0 / *na*	0	2 / *na*	*na*	*na*	0	0	0	*na*	*na*	0	*na*
SDAN	K-SADS-PL administered to parent and child	41 (20)	12 / 2	0 / 0	0 / 0	19 / 3	7 / 2	*n* = 12 current and *n* = 5 lifetime separation anxiety disorder	0 / 0	2 / 1	0 / 0	0 / 0	*n* = 5 current, *n* =1 lifetime ADHD / ODD	*na*	*na*
Stony Brook Temperament Study	K-SADS-PL administered to parent and child	74 (31)	1 / 2	0 / 0	0 / 0	0 / 4	1 / 7	*n* = 3 lifetime (not current) separation anxiety	0 / 1	0 / 1	0 / 0	0 / 0	*n* = 10 other psychopathology (ADHD, ODD, DBD)	1	ADHD medication
TOTS	K-SADS-PL administered to parent and child	27 (15)	3 / 0	0 / 0	0 / 0	1 / 0	3 / 0	*n* = 2 lifetime (not current) separation anxiety	1 / 1	0 / 0	1 / 0	0 / 0	*n* = 2 lifetime ADHD/ADD; *n* = 4 current ADHD	1	ADHD medication
Vanderbilt-children	K-SADS-PL	55 (33)	4 / 0	0 /0	0 / 0	0 /0	1 / 2	*na*	0 / 0	0 / 0	0 / 0	0 / 0	*na*	0	*na*
Vanderbilt-young adults	Structured Clinical Interview for DSM-IV	145 (79)	15 / 1	2 / 0	1 / 0	3 / 0	7 / 1	*n* = 3 anxiety nos	5 / 0	1 /0	1 / 0	3 / 2	*n* = 2 dysthymia (current)	0	*na*
VCU-JAS	K-SADS-PL	126 (75) (diagnostic data available for *n* = 124)	4 / 2	1 / 0	*na*	6 / 0	12 / 4	*n* = 6 current and *n* = 3 lifetime (not current) separation anxiety	0 / 1	1 / 0	0 / 0	*na*	*n* = 15 current ADHD	*na*	*na*

			Psychopathology based on clinical interview (current / lifetime but not current)												
			Anxiety disorders						Other psychopathology					Psychotropic medication at time scan	
Sample	**Clinical interview**	***n* (*n* female)**	**SAD**	**PD**	**AG**	**GAD**	**SPH**	**Other anxiety disorders**	**MDD**	**OCD**	**PTSD**	**SUD**	**Other**	***n***	**Specification medication**
Western University	K-SADS-PL administered to parent and child	82 (36)	1/ 1	0 / 0	0 / 0	1 / 1	1 / 0	*na*	0 / 0	*na*	0 / 0	*na*	*n* = 5 current, *n* = 1 lifetime (ADHD, ODD)	*na*	*na*
Wisconsin Twin Project - RDoC twin study	*na*	152 (93)	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*

Footnote Table S3a

[a] Sample information following preprocessing and quality control.

Abbreviations: ADHD = AG = agoraphobia; DBD = disruptive behavior disorders; GAD = generalized anxiety disorder; MDD = major depressive disorder; na = not available; nos = not otherwise specified; OCD = obsessive compulsive disorder; ODD = Oppositional defiant disorder; PD = panic disorder; PTSD = post-traumatic stress disorder; SAD = social anxiety disorder; SPH = specific phobia; SSRI/SNRI = selective serotonin reuptake inhibitors/ serotonin and norepinephrine reuptake inhibitors; SUD = substance use disorder;

Table S3b: Clinical characteristics per sample // symptomatology
(after preprocessing and quality control at (around) time of scan[a])

	Questionnaires – scores represent mean ± SD																		
Sample	**STAI-trait**	**ASI**	**BAI**	**LSAS - total**	**PSWQ**	**BDI**	**CDI**	**SCARED-T-P**	**SCARED-T-C**	**SCARED-SC-P**	**SCARED-SC-C**	**SCARED-PN-P**	**SCARED-PN-C**	**SCARED-GD-P**	**SCARED-GD-C**	**SCARED-SH-P**	**SCARED-SH-C**	**YSR - total**	**YSR - internalizing**	**Note**
Brains study	*na*	*na*	*na*	*na*	*na*	*na*	*na*	11.4 ± 9.1	17.3 ± 11.8	4.0 ± 3.6	4.8 ± 3.6	0.9 ± 1.4	3.1 ± 3.5	3.8 ± 3.5	4.3 ± 3.3	0.4 ± 0.7	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*	*na*	*na*	*na*	24.8 ± 14.8	*na*	5.8 ± 3.6	*na*	4.6 ± 5.0	*na*	6.0 ± 4.5	*na*	1.5 ± 1.7	*na*	*na*	
Cohort 3 / 4	*na*	*na*	*na*	*na*	*na*	*na*	*na*	6.8 ± 6.5	12.0 ± 9.7	2.5 ± 2.7	3.5 ± 3.1	1.0 ± 1.5	2.2 ± 2.2	2.8 ± 2.9	4.3 ± 3.9	0.5 ± 0.9	0.9 ± 1.2	*na*	*na*	Data availability for diagnostic information: *n* = 61 participants; data availability for SCARED: range 38 - 40 participants.
Generation R - sample with behavioral observations	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	
Generation R - sample with questionnaire data	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*
Maryland-PAX	*na*	*na*	*na*	42.5 ± 21.6	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	
Maryland-TAX	45.5 ± 13.7	*na*	*na*	*na*	48.2 ± 14.6	*na*	*na*	17.1 ± 14.5	27.9 ± 20.3	6.0 ± 4.4	7.1 ± 4.9	2.9 ± 4.1	6.4 ± 6.8	5.2 ± 4.8	8.7 ± 5.5	1.3 ± 1.8	2.2 ± 2.3	*na*	*na*	Data on SCARED for 48 participants.
Nijmegen Longitudinal Study	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	
Pittsburgh	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	
San Raffaele	*na*	*na*	*na*	27.6 ± 21.3	*na*	*na*	*na*	*na*	*na*	3.9 ± 3.2	3.9 ± 2.9	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	
SDAN	34.6 ± 8.5	*na*	*na*	*na*	17.2 ± 9.4	*na*	*na*	20.7 ± 14.3	20.6 ± 14.4	5.2 ± 4.6	5.2 ± 3.8	2.6 ± 3.0	3.5 ± 4.3	6.7 ± 4.7	6.0 ± 4.8	1.4 ± 1.7	1.5 ± 1.6	*na*	*na*	Data on PSWQ for 36 participants

	Questionnaires – scores represent mean ± SD																		
Sample	**STAI-trait**	**ASI**	**BAI**	**LSAS - total**	**PSWQ**	**BDI**	**CDI**	**SCARED-T-P**	**SCARED-T-C**	**SCARED-SC-P**	**SCARED-SC-C**	**SCARED-PN-P**	**SCARED-PN-C**	**SCARED-GD-P**	**SCARED-GD-C**	**SCARED-SH-P**	**SCARED-SH-C**	**YSR - total**	**YSR - internalizing**	**Note**
Stony Brook Temperament Study	*na*	*na*	*na*	*na*	*na*	*na*	3.8 ± 3.7	8.6 ± 7.4	*na*	2.7 ± 2.9	*na*	3.0 ± 0.7	*na*	3.0 ± 3.0	*na*	0.6 ± 1.0	*na*	*na*	*na*	CDI: rated by child
TOTS	27.8 ± 4.6	*na*	*na*	*na*	*na*	*na*	*na*	11.2 ± 8.8	20.4 ± 12.1	3.8 ± 3.6	6.9 ± 3.6	2.2 ± 2.2	3.3 ± 3.7	2.9 ± 2.7	5.1 ± 3.5	0.8 ± 1.1	1.6 ± 1.4	*na*	*na*	Data availability for psychopathology: *n* = 26; for SCARED: 15 - 16 participants.
Vanderbilt - children	*na*	*na*	*na*	*na*	*na*	*na*	5.3 ± 5.5	11.5 ± 10.6	17.7 ± 13.5	4.4 ± 4.5	5.0 ± 3.6	1.0 ± 1.8	2.8 ± 3.3	3.2 ± 3.1	4.0 ± 3.4	0.3 ± 0.6	1.3 ± 1.5	*na*	*na*	Data availability for questionnaires varies from *n* = 44 to *n* = 47
Vanderbilt - young adults	31.7 ± 10.9	13.1 ± 10.8	6.2 ± 5.4	35.1 ± 22.0	45.7 ± 16.1	5.5 ± 6.5	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	Data availability for questionnaires varies from *n* = 29 *to n* = 149
VCU-JAS	*na*	*na*	*na*	*na*	*na*	*na*	*na*	13.0 ± 10.2	20.1 ± 11.4	4.0 ± 3.4	5.2 ± 3.0	1.3 ± 1.9	3.9 ± 3.7	4.5 ± 4.0	5.3 ± 3.6	0.9 ± 1.4	1.4 ± 1.3	*na*	*na*	Data availability for SCARED-P: *n* = 125; for SCARED-C: *n* = 125
Western University	*na*	*na*	*na*	*na*	*na*	*na*	5.1 ± 5.4	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	36.2 ± 24.4	13.0 ± 9.7	Data availability for psychopathology: *n = 82*
Wisconsin Twin Project - RDoC twin study	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	

Footnote Table S3b

[a] Sample information following preprocessing and quality control.

Abbreviations: ASI = anxiety sensitivity index; BAI = Beck anxiety inventor; BDI = Beck depression inventory; CDI = child depression inventory; LSAS = Liebowitz social anxiety scale; na = not available; PSWQ = Penn State worry questionnaire; SCARED = Screen for child anxiety related disorders; STAI-t = state trait anxiety inventory – trait; YSR = youth self report.

References Supplemental Materials

1. Morales S, Taber-Thomas BC, Pérez-Edgar KE. Patterns of attention to threat across tasks in behaviorally inhibited children at risk for anxiety. *Dev Sci*. Published online January 19, 2017. doi:10.1111/desc.12391
2. Taber-Thomas BC, Morales S, Hillary FG, Pérez-Edgar KE. Altered topography of intrinsic functional connectivity in childhood risk for social anxiety. *Depress Anxiety*. 2016;33:995-1004. doi:10.1002/da.22508
3. Liu P, Taber-Thomas BC, Fu X, Pérez-Edgar KE. Biobehavioral Markers of Attention Bias Modification in Temperamental Risk for Anxiety: A Randomized Control Trial. *J Am Acad Child Adolesc Psychiatry*. 2018;57(2):103-110. doi:https://doi.org/10.1016/j.jaac.2017.11.016
4. Auday ES, Pérez-Edgar KE. Limbic and prefrontal neural volume modulate social anxiety in children at temperamental risk. *Depress Anxiety*. 2019;36(8):690-700. doi:10.1002/da.22941
5. Wise S, Huang-Pollock C, Pérez-Edgar K. Implementation of the diffusion model on dot-probe task performance in children with behavioral inhibition. *Psychol Res*. Published online 2021. doi:10.1007/s00426-021-01532-3
6. Anaya B, Vallorani A, Pérez-Edgar K. Dyadic behavioral synchrony between behaviorally inhibited and non-inhibited peers is associated with concordance in EEG frontal Alpha asymmetry and Delta-Beta coupling. *Biol Psychol*. 2021;159:108018. doi:10.1016/j.biopsycho.2021.108018
7. Anaya B, Vallorani AM, Pérez-Edgar K. Individual dynamics of delta–beta coupling: using a multilevel framework to examine inter- and intraindividual differences in relation to social anxiety and behavioral inhibition. *Journal of Child Psychology and Psychiatry*. 2021;62(6):771-779. doi:10.1111/jcpp.13319
8. Fu X, Taber-Thomas BC, Pérez-Edgar K. Frontolimbic functioning during threat-related attention: Relations to early behavioral inhibition and anxiety in children. *Biol Psychol*. 2017;122:98-109. doi:10.1016/j.biopsycho.2015.08.010
9. Suarez GL, Morales S, Metcalf K, Pérez-Edgar KE. Perinatal complications are associated with social anxiety: Indirect effects through temperament. *Infant Child Dev*. 2019;28(3):e2130. doi:10.1002/icd.2130
10. Thai N, Taber-Thomas BC, Pérez-Edgar KE. Neural correlates of attention biases, behavioral inhibition, and social anxiety in children: An ERP study. *Dev Cogn Neurosci*. 2016;19:200-210. doi:10.1016/j.dcn.2016.03.008
11. Bishop G, Spence SH, Mcdonald C. Can Parents and Teachers Provide a Reliable and Valid Report of Behavioral Inhibition? *Child Dev*. 2003;74(6):1899-1917. doi:10.1046/j.1467-8624.2003.00645.x
12. Broeren S, Muris P. A Psychometric Evaluation of the Behavioral Inhibition Questionnaire in a Non-Clinical Sample of Dutch Children and Adolescents. *Child Psychiatry Hum Dev*. 2010;41(2):214-229. doi:10.1007/s10578-009-0162-9
13. Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry*. 2000;39(1):28-38.
14. Birmaher B, Brent DA, Chiapetta L, Bridge J, Monga S, Baugher M. Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study. *J Am Acad Child Adolesc Psychiatry*. 1999;38(10):1230-1236. doi:10.1097/00004583-199910000-00011

15. Hale WW, Raaijmakers Q, Muris P, Meeus W. Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in the General Adolescent Population. *J Am Acad Child Adolesc Psychiatry*. 2005;44(3):283-290. doi:10.1097/00004583-200503000-00013
16. Muris P, Dreessen L, Bögels S, Weckx M, van Melick M. A questionnaire for screening a broad range of DSM-defined anxiety disorder symptoms in clinically referred children and adolescents. *Journal of Child Psychology and Psychiatry*. 2004;45(4):813-820. doi:https://doi.org/10.1111/j.1469-7610.2004.00274.x
17. Dyson MW, Klein DN, Olino TM, Dougherty LR, Durbin CE. Social and Non-Social Behavioral Inhibition in Preschool-Age Children: Differential Associations with Parent-Reports of Temperament and Anxiety. *Child Psychiatry Hum Dev*. 2011;42(4):390-405. doi:10.1007/s10578-011-0225-6
18. Salum GA, Gadelha A, Pan PM, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. *Int J Methods Psychiatr Res*. 2015;24(1):58-73. doi:10.1002/mpr.1459
19. Moura LM, Crossley NA, Zugman A, et al. Coordinated brain development: exploring the synchrony between changes in grey and white matter during childhood maturation. *Brain Imaging Behav*. 2017;11(3):808-817. doi:10.1007/s11682-016-9555-0
20. Axelrud LK, Simioni AR, Pine DS, et al. Neuroimaging Association Scores: reliability and validity of aggregate measures of brain structural features linked to mental disorders in youth. *Eur Child Adolesc Psychiatry*. Published online 2020. doi:10.1007/s00787-020-01653-x
21. Hoffmann MS, Pan PM, Manfro GG, et al. Cross-Sectional and Longitudinal Associations of Temperament and Mental Disorders in Youth. *Child Psychiatry Hum Dev*. 2019;50(3):374-383. doi:10.1007/s10578-018-0846-0
22. Fleitlich-Bilyk B, Goodman R. Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *J Am Acad Child Adolesc Psychiatry*. 2004;43(6):727-734.
23. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: Description and Initial Validation of an Integrated Assessment of Child and Adolescent Psychopathology. *Journal of Child Psychology and Psychiatry*. 2000;41(5):645-655. doi:10.1111/j.1469-7610.2000.tb02345.x
24. Kamakura W, Mazzon JA. Socioeconomic stratification criteria and classification tools in Brazil/Criterios de estratificacao e comparacao de classificadores socioeconomicos no Brasil/Criterios de estratificación y comparación de clasificadores socioeconómicos en Brasil. *RAE*. 2016;56(1):55-71.
25. Ellis LK, Rothbart MK. Revision of the early adolescent temperament questionnaire. In: *Poster Presented at the 2001 Biennial Meeting of the Society for Research in Child Development, Minneapolis, Minnesota*. Citeseer; 2001.
26. Laceulle OM, Ormel J, Vollebergh WAM, Van Aken MAG, Nederhof E. A test of the vulnerability model: temperament and temperament change as predictors of future mental disorders–the TRAILS study. *Journal of Child Psychology and Psychiatry*. 2014;55(3):227-236.
27. Rothbart MK. Temperament, development, and personality. *Curr Dir Psychol Sci*. 2007;16(4):207-212.
28. Shechner T, Fox NA, Mash JA, et al. Differences in neural response to extinction recall in young adults with or without history of behavioral inhibition. *Dev Psychopathol*. 2018;30(1):179-189.
29. Roy AK, Benson BE, Degnan KA, et al. Alterations in amygdala functional connectivity reflect early temperament. *Biol Psychol*. 2014;103:248-254. doi:10.1016/j.biopsycho.2014.09.007

30. Fox NA, Henderson HA, Rubin KH, Calkins SD, Schmidt LA. Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Dev*. 2001;72(1):1-21. doi:10.1111/1467-8624.00262
31. First MB, Spitzer RL. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. *SCIDI/NP) New York: Biometrics Research, New York State Psychiatric Institute*. Published online 2002.
32. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev*. 1988;8(1):77-100. doi:10.1016/0272-7358(88)90050-5
33. Spielberger CD, Gorsuch RL, Lushene RE. *STAI Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press; 1970.
34. Goldsmith HH. Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Dev*. 1996;67(1):218-235.
35. Goldsmith HH, Reilly J, Lemery KS, Longley S, Prescott A. Laboratory temperament assessment battery: Preschool version. *Unpublished manuscript*. Published online 1995.
36. Jaddoe VW V, van Duijn CM, van der Heijden AJ, et al. The Generation R Study: design and cohort update until the age of 4 years. *Eur J Epidemiol*. 2008;23(12):801. doi:10.1007/s10654-008-9309-4
37. White T, Marroun H El, Nijs I, et al. Pediatric population-based neuroimaging and the Generation R Study: the intersection of developmental neuroscience and epidemiology. *Eur J Epidemiol*. 2013;28(1):99-111. doi:10.1007/s10654-013-9768-0
38. White T, Muetzel RL, El Marroun H, et al. Paediatric population neuroimaging and the Generation R Study: the second wave. *Eur J Epidemiol*. 2018;33(1):99-125. doi:10.1007/s10654-017-0319-y
39. Pappa I, Mileva-Seitz VR, Szekely E, et al. DRD4 VNTRs, observed stranger fear in preschoolers and later ADHD symptoms. *Psychiatry Res*. 2014;220(3):982-986. doi:https://doi.org/10.1016/j.psychres.2014.09.004
40. Jansen PW, Raat H, Mackenbach JP, et al. Socioeconomic inequalities in infant temperament. *Soc Psychiatry Psychiatr Epidemiol*. 2009;44(2):87-95. doi:10.1007/s00127-008-0416-z
41. Jaddoe VW V, Mackenbach JP, Moll HA, et al. The Generation R Study: design and cohort profile. *Eur J Epidemiol*. 2006;21(6):475-484.
42. Fisher P, Lucas C. Diagnostic interview schedule for children (DISC-IV)-young child. *New York: Columbia University*. Published online 2006.
43. Gartstein MA, Rothbart MK. Studying infant temperament via the revised infant behavior questionnaire. *Infant Behav Dev*. 2003;26(1):64-86.
44. Roza SJ, Van Lier PAC, Jaddoe VW V, et al. Intrauterine growth and infant temperamental difficulties: the Generation R Study. *J Am Acad Child Adolesc Psychiatry*. 2008;47(3):264-272.
45. Goldsmith HH, Reilly J, Lemery KS, Longley S, Prescott A. The laboratory assessment battery: Preschool version (LAB-TAB). *Madison: University of Wisconsin*. Published online 1999.
46. Grogans SE, Hur J, Barstead MG, et al. Neuroticism/negative emotionality is associated with increased reactivity to uncertain threat in the bed nucleus of the stria terminalis, not the amygdala. *The Journal of Neuroscience*. Published online July 15, 2024:e1868232024. doi:10.1523/jneurosci.1868-23.2024

47. Hur J, Smith JF, DeYoung KA, et al. Anxiety and the Neurobiology of Temporally Uncertain Threat Anticipation. *The Journal of Neuroscience*. 2020;40(41):7949 LP - 7964. doi:10.1523/JNEUROSCI.0704-20.2020
48. Hur J, Deyoung KA, Islam S, Anderson AS, Barstead MG, Shackman AJ. Social context and the real-world consequences of social anxiety. *Psychol Med*. 2020;50(12):1989-2000. doi:10.1017/S0033291719002022
49. Hur J, Kuhn M, Grogans SE, et al. Anxiety-Related Frontocortical Activity Is Associated With Dampened Stressor Reactivity in the Real World. *Psychol Sci*. Published online June 3, 2022:09567976211056635. doi:10.1177/09567976211056635
50. First MB, Williams JB, Karg RS, Spitzer RL. *Structured Clinical Interview for DSM-5, Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV).* American Psychiatric Association; 2015.
51. Gladstone G, Parker G. Measuring a behaviorally inhibited temperament style: Development and initial validation of new self-report measures. *Psychiatry Res*. 2005;135(2):133-143. doi:10.1016/j.psychres.2005.03.005
52. Hur J, Tillman RM, Kim HC, et al. Adolescent social anxiety is associated with diminished discrimination of anticipated threat and safety in the bed nucleus of the stria terminalis. *Journal of Psychopathology and Clinical Science*. 2025;134(1):41-56. doi:10.1037/abn0000940
53. Bunnell BE, Beidel DC, Liu L, Joseph DL, Higa-McMillan C. The SPAIC-11 and SPAICP-11: Two brief child-and parent-rated measures of social anxiety. *J Anxiety Disord*. 2015;36:103-109.
54. Sheehan D V., Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59 Suppl 2:22-33.
55. Tyborowska A, Volman I, Niermann HCM, et al. Early-life and pubertal stress differentially modulate grey matter development in human adolescents. *Sci Rep*. 2018;8(1):9201. doi:10.1038/s41598-018-27439-5
56. Van Bakel HJA, Riksen-Walraven JM. Stress reactivity in 15-month-old infants: Links with infant temperament, cognitive competence, and attachment security. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*. 2004;44(3):157-167.
57. Van Bakel HJA, Riksen-Walraven JM. Parenting and development of one-year-olds: Links with parental, contextual, and child characteristics. *Child Dev*. 2002;73(1):256-273.
58. Meyers SA. Mothering in context: Ecological determinants of parent behavior. *Merrill-Palmer Quarterly (1982-)*. Published online 1999:332-357.
59. Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatr Rev*. 2000;21(8):265-271. doi:10.1542/pir.21-8-265
60. Goldsmith HH. The toddler behavior assessment questionnaire. *Oregon Center for the Study of Emotion technical report*. 1988;(88-04).
61. Mullen M, Snidman N, Kagan J. Brief report free-play behavior in inhibited and uninhibited children. *Infant Behav Dev*. 1993;16(3):383-389.
62. Nachmias M, Gunnar M, Mangelsdorf S, Parritz RH, Buss K. Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Dev*. 1996;67(2):508-522.
63. Hill SY, Shen S, Lowers L, Locke-Wellman J, Matthews AG, McDermott M. Psychopathology in offspring from multiplex alcohol dependence families with and

without parental alcohol dependence: a prospective study during childhood and adolescence. *Psychiatry Res*. 2008;160(2):155-166. doi:10.1016/j.jpsychires.2010.08.005
64. Hill SY. Familial risk for alcohol dependence and brain morphology: the role of cortical thickness across the lifespan. *Alcohol Clin Exp Res*. 2018;42(5):841-844. doi:10.1111/acer.13621
65. Hill SY, Tessner KD, McDermott MD. Psychopathology in offspring from families of alcohol dependent female probands: A prospective study. *J Psychiatr Res*. 2011;45(3):285-294. doi:10.1016/j.jpsychires.2010.08.005
66. Hill SY, Wang S, Carter H, McDermott MD, Zezza N, Stiffler S. Amygdala volume in offspring from multiplex for alcohol dependence families: the moderating influence of childhood environment and 5-HTTLPR variation. *J Alcohol Drug Depend*. Published online 2013. doi:10.4172/2329-6488.S1-001
67. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health diagnostic interview schedule: Its history, characteristics, and validity. *Arch Gen Psychiatry*. 1981;38(4):381-389.
68. Chambers WJ, Puig-Antich J, Hirsch M, et al. The assessment of affective disorders in children and adolescents by semistructured interview: test-retest reliability of the Schedule for Affective Disorders and Schizophrenia for School-Age Children, Present Episode Version. *Arch Gen Psychiatry*. 1985;42(7):696-702.
69. Janča A, Robins LN, Cottler LB, Early TS. Clinical observation of assessment using the Composite International Diagnostic Interview (CIDI): an analysis of the CIDI field trials–wave II at the St Louis site. *The British Journal of Psychiatry*. 1992;160(6):815-818.
70. Cottler LB, Robins LN, Helzer JE. The reliability of the CIDI-SAM: a comprehensive substance abuse interview. *Br J Addict*. 1989;84(7):801-814.
71. Hill SY, Tessner K, Wang S, Carter H, McDermott M. Temperament at 5 years of age predicts amygdala and orbitofrontal volume in the right hemisphere in adolescence. *Psychiatry Res Neuroimaging*. 2010;182(1):14-21. doi:10.1016/j.pscychresns.2009.11.006
72. Hill SY, Lowers L, Locke J, Snidman N, Kagan J. Behavioral Inhibition in Children From Families at High Risk for Developing Alcoholism. *J Am Acad Child Adolesc Psychiatry*. 1999;38(4):410-417. doi:https://doi.org/10.1097/00004583-199904000-00013
73. Kagan J, Reznick JS, Snidman N, Gibbons J, Johnson MO. Childhood derivatives of inhibition and lack of inhibition to the unfamiliar. *Child Dev*. Published online 1988:1580-1589.
74. Tellegen A, Lykken DT, Bouchard TJ, Wilcox KJ, Segal NL, Rich S. Personality similarity in twins reared apart and together. *J Pers Soc Psychol*. 1988;54(6):1031.
75. Battaglia M, Zanoni A, Taddei M, et al. Cerebral responses to emotional expressions and the development of social anxiety disorder: a preliminary longitudinal study. *Depress Anxiety*. 2012;29(1):54-61. doi:10.1002/da.20896
76. Battaglia M, Ogliari A, Zanoni A, et al. Children's Discrimination of Expressions of Emotions: Relationship With Indices of Social Anxiety and Shyness. *J Am Acad Child Adolesc Psychiatry*. 2004;43(3):358-365. doi:10.1097/01.chi.0000103177.91435.1b
77. Battaglia M, Ogliari A, Zanoni A, et al. Influence of the serotonin transporter promoter gene and shyness on children's cerebral responses to facial expressions. *Arch Gen Psychiatry*. 2005;62(1):85-94. doi:10.1001/archpsyc.62.1.85
78. Taddei M, Tettamanti M, Zanoni A, Cappa S, Battaglia M. Brain white matter organisation in adolescence is related to childhood cerebral responses to facial

expressions and harm avoidance. *Neuroimage*. 2012;61(4):1394-1401. doi:10.1016/j.neuroimage.2012.03.062
79. Kaufman J, Birmaher B, Brent D, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36(7):980-988. doi:10.1097/00004583-199707000-00021
80. Biederman J, Hirshfeld-Becker DR, Rosenbaum JF, et al. Further evidence of association between behavioral inhibition and social anxiety in children. *Am J Psychiatry*. 2001;158(10):1673-1679.
81. Hirshfeld-Becker DR, Biederman J. Rationale and Principles for Early Intervention with Young Children at Risk for Anxiety Disorders. *Clin Child Fam Psychol Rev*. 2002;5(3):161-172. doi:10.1023/A:1019687531040
82. DUMMIT III ES, Klein RG, Tancer NK, Asche B, Martin J, Fairbanks JA. Systematic assessment of 50 children with selective mutism. *J Am Acad Child Adolesc Psychiatry*. 1997;36(5):653-660.
83. Masia CL, Klein RG, Storch EA, Corda B. School-based behavioral treatment for social anxiety disorder in adolescents: Results of a pilot study. *J Am Acad Child Adolesc Psychiatry*. 2001;40(7):780-786.
84. Stevenson-Hinde J, Glover A. Shy girls and boys: A new look. *Journal of Child Psychology and Psychiatry*. 1996;37(2):181-187.
85. Cooper PJ, Eke M. Childhood shyness and maternal social phobia: A community study. *British Journal of Psychiatry*. 1999;174(5):439-443. doi:10.1192/bjp.174.5.439
86. Cloninger CR, Przybeck TR, Svrakic DM, Wetzel RD. The Temperament and Character Inventory (TCI): A guide to its development and use. Published online 1994.
87. Marshall PJ, Stevenson-Hinde J. Behavioral inhibition, heart period, and respiratory sinus arrhythmia in young children. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*. 1998;33(3):283-292.
88. Švrakić NM, Švrakić DM, Cloninger CR. A general quantitative theory of personality development: fundamentals of a self-organizing psychobiological complex. *Dev Psychopathol*. 1996;8(1):247-272.
89. Battaglia M, Przybeck TR, Bellodi L, Cloninger CR. Temperament dimensions explain the comorbidity of psychiatric disorders. *Compr Psychiatry*. 1996;37(4):292-298.
90. Gold AL, Brotman MA, Adleman NE, et al. Comparing Brain Morphometry Across Multiple Childhood Psychiatric Disorders. *J Am Acad Child Adolesc Psychiatry*. 2016;55(12):1027-1037. doi:10.1016/j.jaac.2016.08.008
91. Gold AL, Abend R, Britton JC, et al. Age differences in the neural correlates of anxiety disorders: An fMRI study of response to learned threat. *American Journal of Psychiatry*. 2020;177(5):454-463.
92. Linke JO, Abend R, Kircanski K, et al. Shared and Anxiety-Specific Pediatric Psychopathology Dimensions Manifest Distributed Neural Correlates. *Biol Psychiatry*. 2021;89(6):579-587. doi:10.1016/j.biopsych.2020.10.018
93. Kim J, Klein DN, Olino TM, Dyson MW, Dougherty LR, Durbin CE. Psychometric Properties of the Behavioral Inhibition Questionnaire in Preschool Children. *J Pers Assess*. 2011;93(6):545-555. doi:10.1080/00223891.2011.608756
94. Laptook RS, Klein DN, Olino TM, Dyson MW, Carlson G. Low positive affectivity and behavioral inhibition in preschool-age children: A replication and extension of previous findings. *Pers Individ Dif*. 2010;48(5):547-551.
95. Kopala-Sibley DC, Cyr M, Finsaas MC, et al. Early Childhood Parenting Predicts Late Childhood Brain Functional Connectivity During Emotion Perception and Reward Processing. *Child Dev*. 2020;91(1):110-128. doi:https://doi.org/10.1111/cdev.13126

96. Johnson VC, Olino TM, Klein DN, et al. A longitudinal investigation of predictors of the association between age 3 and age 6 behavioural inhibition. *J Res Pers*. 2016;63:51-61. doi:10.1016/j.jrp.2016.04.008
97. Stumper A, Danzig AP, Dyson MW, Olino TM, Carlson GA, Klein DN. Parents' behavioral inhibition moderates association of preschoolers' BI with risk for age 9 anxiety disorders. *J Affect Disord*. 2017;210:35-42. doi:10.1016/j.jad.2016.12.008
98. Mumper EE, Dyson MW, Finsaas MC, Olino TM, Klein DN. Life stress moderates the effects of preschool behavioral inhibition on anxiety in early adolescence. *Journal of Child Psychology and Psychiatry*. 2020;61(2):167-174. doi:10.1111/jcpp.13121
99. Kann SJ, O'Rawe JF, Huang AS, Klein DN, Leung HC. Preschool negative emotionality predicts activity and connectivity of the fusiform face area and amygdala in later childhood. *Soc Cogn Affect Neurosci*. 2017;12(9):1511-1519.
100. Olino TM, Klein DN, Lewinsohn PM, Rohde P, Seeley JR. Latent trajectory classes of depressive and anxiety disorders from adolescence to adulthood: descriptions of classes and associations with risk factors. *Compr Psychiatry*. 2010;51(3):224-235. doi:https://doi.org/10.1016/j.comppsych.2009.07.002
101. Olino TM, Klein DN, Dyson MW, Rose SA, Durbin CE. Temperamental emotionality in preschool-aged children and depressive disorders in parents: associations in a large community sample. *J Abnorm Psychol*. 2010;119(3):468.
102. Egger HL, Ascher BH, Angold A. Preschool age psychiatric assessment (PAPA). *Durham (North Carolina): Duke University Medical Center*. Published online 1999.
103. Durbin CE, Hayden EP, Klein DN, Olino TM. Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*. 2007;7(2):388-399. doi:10.1037/1528-3542.7.2.388
104. Kagan J. Temperamental contributions to social behavior. *American psychologist*. 1989;44(4):668.
105. Durbin CE, Klein DN, Hayden EP, Buckley ME, Moerk KC. Temperamental emotionality in preschoolers and parental mood disorders. *J Abnorm Psychol*. 2005;114(1):28.
106. Hane AA, Fox NA, Henderson HA, Marshall PJ. Behavioral reactivity and approach-withdrawal bias in infancy. *Dev Psychol*. 2008;44(5):1491.
107. Buzzell GA, Troller-Renfree S V., Barker T V., et al. A Neurobehavioral Mechanism Linking Behaviorally Inhibited Temperament and Later Adolescent Social Anxiety. *J Am Acad Child Adolesc Psychiatry*. 2017;56(12):1097-1105. doi:10.1016/j.jaac.2017.10.007
108. Filippi CA, Sachs JF, Phillips D, et al. Infant behavioral reactivity predicts change in amygdala volume 12 years later. *Dev Cogn Neurosci*. 2020;42. doi:10.1016/j.dcn.2020.100776
109. Calkins SD, Fox NA, Marshall TR. Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Dev*. 1996;67(2):523-540.
110. Birmaher B, Khetarpal S, Brent D, et al. The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry*. 1997;36(4):545-553.
111. Goldsmith HH, Rothbart MK. The laboratory temperament assessment battery (Lab-TAB): Pre-locomotor Version 3.1. *Department of Psychology, University of Oregon*. Published online 1999.
112. Clauss JA, Benningfield MM, Rao U, Blackford JU. Altered Prefrontal Cortex Function Marks Heightened Anxiety Risk in Children. *J Am Acad Child Adolesc Psychiatry*. 2016;55(9):809-816. doi:10.1016/j.jaac.2016.05.024

113. Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. *Eur J Pharmacol*. 2000;410(2-3):107-120.
114. NL KASK. Kaufman Brief Intelligence Test. Manual. *K-BIT) American Guidance Service Circle Pines, MN*. Published online 1990.
115. Oldfield R. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971;9(1):97-113. doi:10.1016/0028-3932(71)90067-4
116. Smucker MR, Craighead WE, Craighead LW, Green BJ. Normative and reliability data for the Children's Depression Inventory. *J Abnorm Child Psychol*. 1986;14(1):25-39.
117. Avery SN, Blackford JU. Slow to warm up: the role of habituation in social fear. *Soc Cogn Affect Neurosci*. 2016;11(11):1832-1840. doi:10.1093/scan/nsw095
118. Blackford JU, Avery SN, Shelton RC, Zald DH. Amygdala temporal dynamics: temperamental differences in the timing of amygdala response to familiar and novel faces. *BMC Neurosci*. 2009;10:145. doi:10.1186/1471-2202-10-145
119. Blackford JU, Avery SN, Cowan RL, Shelton RC, Zald DH. Sustained amygdala response to both novel and newly familiar faces characterizes inhibited temperament. *Soc Cogn Affect Neurosci*. 2011;6(5):621-629. doi:10.1093/scan/nsq073
120. Clauss JA, Avery SN, Vanderklok RM, et al. Neurocircuitry underlying risk and resilience to Social Anxiety Disorder. *Depress Anxiety*. 2014;31(10):822-833. doi:10.1002/da.22265
121. Thornton-Wells TA, Avery SN, Blackford JU. Using novel control groups to dissect the amygdala's role in Williams syndrome. *Dev Cogn Neurosci*. 2011;1(3):295-304. doi:10.1016/j.dcn.2011.03.003
122. Blackford JU, Allen AH, Cowan RL, Avery SN. Amygdala and hippocampus fail to habituate to faces in individuals with an inhibited temperament. *Soc Cogn Affect Neurosci*. 2013;8(2):143-150. doi:10.1093/scan/nsr078
123. Reznick JS, Hegeman IM, Kaufman ER, Woods SW, Jacobs M. Retrospective and concurrent self-report of behavioral inhibition and their relation to adult mental health. *Dev Psychopathol*. 1992;4(2):301-321. doi:10.1017/S095457940000016X
124. Spitzer RL, Williams JBW, Gibbon M, First MB. The structured clinical interview for DSM-III-R (SCID): I: history, rationale, and description. *Arch Gen Psychiatry*. 1992;49(8):624-629.
125. Rohrbacher H, Hoyer J, Beesdo K, et al. Psychometric properties of the Retrospective Self Report of Inhibition (RSRI) in a representative German sample. *Int J Methods Psychiatr Res*. 2008;17(2):80-88.
126. Clauss JA, Seay AL, Vanderklok RM, et al. Structural and functional bases of inhibited temperament. *Soc Cogn Affect Neurosci*. 2014;9(12):2049-2058. doi:10.1093/scan/nsu019
127. Bourdon JL, Savage JE, Verhulst B, et al. The Genetic and Environmental Relationship Between Childhood Behavioral Inhibition and Preadolescent Anxiety. *Twin Research and Human Genetics*. 2019;22(1):48-55. doi:10.1017/thg.2018.73
128. Carney DM, Moroney E, Machlin L, et al. The Twin Study of Negative Valence Emotional Constructs. *Twin Research and Human Genetics*. 2016;19(5):456-464. doi:10.1017/thg.2016.59
129. Hettema JM, Bourdon JL, Sawyers C, et al. Genetic and environmental risk structure of internalizing psychopathology in youth. *Depress Anxiety*. 2020;37(6):540-548. doi:10.1002/da.23024
130. Lee M, Aggen SH, Carney DM, et al. Latent structure of negative valence measures in childhood. *Depress Anxiety*. 2017;34(8):742-751. doi:10.1002/da.22656
131. Liu P, Kryski KR, Smith HJ, Joanisse MF, Hayden EP. Transactional relations between early child temperament, structured parenting, and child outcomes: A three-wave

longitudinal study. *Dev Psychopathol*. 2020;32(3):923-933. doi:10.1017/S0954579419000841
132. Liu P, Vandermeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Neural Activity During Self-referential Processing in Children at Risk for Depression. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(4):429-437. doi:10.1016/j.bpsc.2019.12.012
133. Liu P, Vandemeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Depressogenic self-schemas are associated with smaller regional grey matter volume in never-depressed preadolescents. *Neuroimage Clin*. 2020;28:102422. doi:10.1016/j.nicl.2020.102422
134. Liu P, Vandermeer MRJ, Mohamed Ali O, et al. Maternal Depression, Child Temperament, and Early-Life Stress Predict Never-Depressed Preadolescents' Functional Connectivity During a Negative-Mood Induction. *Clinical Psychological Science*. Published online June 3, 2021:21677026211016420. doi:10.1177/21677026211016419
135. Vandermeer MRJ, Sheikh HI, Singh SS, et al. The BDNF gene val66met polymorphism and behavioral inhibition in early childhood. *Social Development*. 2018;27(3):543-554. doi:10.1111/sode.12292
136. Vandermeer MRJ, Liu P, Mohamed Ali O, et al. Orbitofrontal cortex grey matter volume is related to children's depressive symptoms. *Neuroimage Clin*. 2020;28:102395. doi:10.1016/j.nicl.2020.102395
137. Mohamed Ali O, Vandermeer MRJ, Sheikh HI, Joanisse MF, Hayden EP. Girls' internalizing symptoms and white matter tracts in Cortico-Limbic circuitry. *Neuroimage Clin*. 2019;21:101650. doi:10.1016/j.nicl.2018.101650
138. Dunn LM, Dunn LM, Bulheller S, Häcker H. *Peabody Picture Vocabulary Test*. American Guidance Service Circle Pines, MN; 1965.
139. Kovacs M. The Childrens' Depression Inventory: A self-rated depression scale for school-aged youngsters. *Pittsburgh: University of Pittsburgh School of Medicine*. Published online 1983.
140. Achenbach TM, Rescorla L. *Manual for the ASEBA School-Age Forms & Profiles: An Integrated System of Multi-Informant Assessment*. Aseba Burlington, VT:; 2001.
141. Rankin Williams L, Degnan KA, Perez-Edgar KE, et al. Impact of Behavioral Inhibition and Parenting Style on Internalizing and Externalizing Problems from Early Childhood through Adolescence. *J Abnorm Child Psychol*. 2009;37(8):1063-1075. doi:10.1007/s10802-009-9331-3
142. Schmidt NL, Lemery-Chalfant K, Goldsmith HH. Wisconsin Twin Project Overview: Temperament and Affective Neuroscience. *Twin Research and Human Genetics*. 2019;22(6):794-799. doi:10.1017/thg.2019.108
143. Schmidt NL, Brooker RJ, Carroll IC, et al. Longitudinal Research at the Interface of Affective Neuroscience, Developmental Psychopathology, Health and Behavioral Genetics: Findings from the Wisconsin Twin Project. *Twin Research and Human Genetics*. 2019;22(4):233-239. doi:10.1017/thg.2019.55
144. Goldsmith HH, Hilton EC, Phan JM, et al. Childhood inhibition predicts adolescent social anxiety: Findings from a longitudinal twin study. *Dev Psychopathol*. Published online 2022:1-20. doi:10.1017/S0954579422000864
145. Schmidt NL, Van Hulle CA, Brooker RJ, Meyer LR, Lemery-Chalfant K, Goldsmith HH. Wisconsin Twin Research: Early Development, Childhood Psychopathology, Autism, and Sensory Over-responsivity. *Twin Research and Human Genetics*. 2013;16(1):376-384. doi:10.1017/thg.2012.105

146. Goldsmith HH, Lemery-Chalfant K, Schmidt NL, Arneson CL, Schmidt CK. Longitudinal analyses of affect, temperament, and childhood psychopathology. *Twin Research and Human Genetics*. 2007;10(1):118-126.
147. Lemery-Chalfant K, Goldsmith HH, Schmidt NL, Arneson CL, Van Hulle CA. Wisconsin Twin Panel: current directions and findings. *Twin Research and Human Genetics*. 2006;9(6):1030-1037.
148. Schmidt NL, Van Hulle CA, Brooker RJ, Meyer LR, Lemery-Chalfant K, Goldsmith HH. Wisconsin Twin Research: Early development, childhood psychopathology, autism, and sensory over-responsivity. *Twin Research and Human Genetics*. 2013;16(1):376-384.
149. Van Hulle CA, Lemery KS, Goldsmith HH. Wisconsin twin panel. *Twin Research and Human Genetics*. 2002;5(5):502-505.
150. Moore MM, Planalp EM, Van Hulle CA, Goldsmith HH. Pediatric assessment of Research Domain Criteria positive and negative valence systems: Partial genetic mediation of links to problem behaviors. *Journal Psychopathol Clin Sci*. 2022;131(6):626-640. doi:10.1037/abn0000652
151. Goldsmith HH, Reilly J, Lemery KS, Longley S, Prescott A. Preliminary manual for the Preschool Laboratory Temperament Assessment Battery (Technical Report Version 1.0). *Madison, WI: Department of Psychology, University of Wisconsin—Madison*. Published online 1993.
152. Gagne JR, Van Hulle CA, Aksan N, Essex MJ, Goldsmith HH. Deriving childhood temperament measures from emotion-eliciting behavioral episodes: scale construction and initial validation. *Psychol Assess*. 2011;23(2):337.
153. Goldsmith HH, Lemery KS, Schmidt NL, Schmidt CK, Chow OK, Justen AA. Field administration manual for the Laboratory Temperament Assessment Battery: Middle childhood version. *Psychology Department: University of Wisconsin–Madison[Google Scholar]*. Published online 2010.
154. Armstrong JM, Goldstein LH. Manual for the MacArthur health and behavior questionnaire (HBQ 1.0). *MacArthur Foundation Research Network on Psychopathology and Development, University of Pittsburgh*. Published online 2003.
155. Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Dev*. 2001;72(5):1394-1408.
156. Bas-Hoogendam JM, Groenewold NA, Aghajani M, et al. ENIGMA-anxiety working group: Rationale for and organization of large-scale neuroimaging studies of anxiety disorders. *Hum Brain Mapp*. 2022;43(1):83-112. doi:10.1002/hbm.25100
157. Hamilton M. The assessment of anxiety states by rating. *British Journal of Medical Psychology*. 1959;32(1):50-55. doi:10.1111/j.2044-8341.1959.tb00467.x
158. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the penn state worry questionnaire. *Behaviour research and therapy*. 1990;28(6):487-495.
159. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092
160. Taylor S, Zvolensky MJ, Cox BJ, et al. Robust dimensions of anxiety sensitivity: development and initial validation of the Anxiety Sensitivity Index-3. *Psychol Assess*. 2007;19(2):176.
161. Heimberg RG, Horner KJ, Juster HR, et al. Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychol Med*. 1999;29(01):199-212. doi:10.1017/S0033291798007879

162. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia: II. The Panic and Agoraphobia Scale. *Int Clin Psychopharmacol*. 1995;10(2):73-81. doi:10.1097/00004850-199506000-00003
163. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: The Body Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. *J Consult Clin Psychol*. 1984;52(6):1090-1097. doi:10.1037/0022-006X.52.6.1090
164. Shear MK, Rucci P, Williams J, et al. Reliability and validity of the Panic Disorder Severity Scale: replication and extension. *J Psychiatr Res*. 2001;35(5):293-296.
165. Beck AT, Steer RA, Brown G. *Manual for the Beck Depression Inventory-II.* Psychological Corporation; 1996.
166. Kovacs M. The Children's Depression Inventory (CDI). *Psychopharmacol Bull*. 1985;21:995-998.
167. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56-62. doi:10.1136/jnnp.23.1.56