Age-related change in task-evoked amygdala—prefrontal circuitry: a multiverse approach with an accelerated longitudinal cohort aged 4-22 years

Paul Alexander Bloom[1*], Michelle VanTieghem[2], Laurel Gabard-Durnam[3], Dylan G. Gee[4], Jessica Flannery[5], Christina Caldera[6], Bonnie Goff[6], Eva H. Telzer[7], Kathryn L. Humphreys[8], Dominic S. Fareri[9], Mor Shapiro[10], Sameah Algharazi[11], Niall Bolger[1], Mariam Aly[1], & Nim Tottenham[1]

[1]Columbia University [2]Virta Health [3]Northeastern University [4]Yale University [5]Limbix Health [6]University of California Los Angeles [7]University of North Carolina at Chapel Hill [8]Vanderbilt University [9]Adelphi University [10]Kaiser Permanente [11]City College of New York

Please address correspondence to:
Paul Alexander Bloom
Columbia University Psychology, New York, NY
Email: paul.bloom@columbia.edu

Keywords: Amygdala, prefrontal cortex, development, longitudinal, multiverse, robustness

Data Availability

Because participants and their parents did not consent to data sharing at the time of participation, we cannot make data from the current study publicly available.

Funding

This work was supported by funding from the National Institutes of Mental Health (2R01MH091864) and the Dana Foundation for Nim Tottenham, and a National Science Foundation Graduate Research Fellowship (DGE 1644869) for Paul Alexander Bloom. The authors declare no competing interests.

Abstract

The amygdala and its connections with medial prefrontal cortex (mPFC) play central roles in the development of emotional processes. While several studies have suggested that this circuitry exhibits functional changes across the first two decades of life, findings have been mixed – perhaps resulting from differences in analytic choices across studies. Here we used multiverse analyses to examine the robustness of task-based amygdala–mPFC function findings to analytic choices within the context of an accelerated longitudinal design (4-22 years-old; N=98; 183 scans; 1-3 scans/participant). Participants, recruited from the greater Los Angeles area, completed an event-related emotional face (fear, neutral) task. Parallel analyses varying in preprocessing and modeling choices found that age-related change estimates for amygdala reactivity were more robust than task-evoked amygdala–mPFC functional connectivity to varied analytical choices. Specification curves indicated evidence for age-related decreases in amygdala reactivity to faces, though within-participant changes in amygdala reactivity could not be differentiated from between-participant differences. In contrast, amygdala—mPFC functional connectivity results varied across methods much more, and evidence for age-related change in amygdala–mPFC connectivity was not consistent. Generalized psychophysiological interaction (gPPI) measurements of connectivity were especially sensitive to whether a deconvolution step was applied. Our findings demonstrate the importance of assessing the robustness of findings to analysis choices, although the age-related changes in our current work cannot be overinterpreted given low test-retest reliability. Together, these findings highlight both the challenges in estimating developmental change in longitudinal cohorts and the value of multiverse approaches in developmental neuroimaging for assessing robustness of results. (Preprint: https://www.biorxiv.org/content/10.1101/2021.10.08.463601v1).

Key Points

- Multiverse analyses applied to fMRI data are valuable for determining the robustness of findings to varied analytical choices
- In the current study, age-related change estimates for amygdala reactivity were relatively robust to analytical decisions, though gPPI functional connectivity analyses were much more sensitive, leading some estimates to flip sign
- Both test-retest reliability and robustness to analytical choices are important considerations for developmental research

Rodent models and human neuroimaging have provided converging evidence for the importance of the amygdala and medial prefrontal cortex (mPFC) in the development of threat processing (Adolphs, 2008; Forbes et al., 2011), emotion regulation (Pozzi et al., 2020; Silvers et al., 2015; Sullivan & Perry, 2015), and affective learning (Moriceau & Sullivan, 2006; Pattwell et al., 2016). Characterizing growth trajectories of these regions may provide insight into neural constructions underlying emotional development (Meyer & Lee, 2019). To probe amygdala–mPFC circuitry across development, face stimuli are frequently employed because they effectively engage this circuitry while being child-appropriate (Hariri et al., 2002). Though a number of studies have examined age-related changes from childhood to young adulthood in amygdala responses and amygdala–mPFC functional connectivity (FC) associated with emotional face stimuli, findings have varied widely (likely due in part to differences in sample composition and task design; see sTable 1 for details). Several studies have found age-related change in amygdala reactivity, including decreases as a function of age in response to emotional faces (Gee et al., 2013; Guyer et al., 2008; Killgore et al., 2001; Passarotti et al., 2009; Swartz et al., 2014; Telzer et al., 2015) as well as other images (Decety et al., 2012; Silvers et al., 2017b; Vink et al., 2014), increases in amygdala reactivity with age (Joseph et al., 2015; Todd et al., 2011), developmental sex differences (Xu et al., 2021) or peaks during adolescence (Hare et al., 2008; Vijayakumar et al., 2019). Others have observed no age-related changes (Kujawa et al., 2016; Pfeifer et al., 2011; Pine et al., 2001; Wu et al., 2016; Yurgelun-Todd & Killgore, 2006; Zhang et al., 2019).

With task-evoked amygdala–mPFC FC, several studies have found age-related decreases from childhood to young adulthood (Gee et al., 2013; Kujawa et al., 2016; Silvers et al., 2017a; Wu et al., 2016), while others have found increases (Decety et al., 2012; Perlman & Pelphrey, 2011; Vink et al., 2014), developmental sex differences (Xu et al., 2021), or little age-

related change (Zhang et al., 2019). While some investigations have found differing age-related change for faces displaying different emotions (Killgore & Yurgelun-Todd, 2007; Swartz et al., 2014; Vijayakumar et al., 2019), even investigations of fearful faces specifically have varied in their developmental findings for both amygdala reactivity and amygdala—mPFC functional connectivity (Forbes et al., 2011; Gee et al., 2013; Killgore et al., 2001; Wu et al., 2016, 2016; Zhang et al., 2019).

While the small sample sizes examined in many studies on amygdala–mPFC development likely contribute to differences in findings (Marek et al., 2020), especially given low reliability of many amygdala—mPFC measures (Elliott et al., 2020; Herting et al., 2017; Sauder et al., 2013), important methodological differences also exist across studies. Differences in age range or sample demographics, stimuli, task (e.g. passive viewing vs. emotion labeling or matching (Lieberman et al., 2007), task design (blocked vs. event-related; Sergerie et al., 2008), or contrast used (faces > shapes vs. faces > baseline) may also contribute to discrepancies (see sTable 1). The brain regions under investigation also differ across studies; for example, prefrontal clusters with which amygdala connectivity was assessed. Interpreting discrepancies across studies without appreciation for these methodological differences would be inappropriate, and in fact, incorrect. Yet, such differences do not account for all discrepancies in findings across studies. Variation in processing pipelines is another source of differences across studies, as varying analytic decisions can produce qualitatively different findings, even between putatively identical analyses of the same dataset (Botvinik-Nezer et al., 2020). Choices including software package (Bowring et al., 2019), spatial smoothing (Jo et al., 2007), treatment of head motion (Achterberg & van der Meulen, 2019), parcellation (Bryce et al., 2021), and functional connectivity approach (Di et al., 2020) can also impact results and qualitatively change findings (Cisler et al., 2014). Additionally, the majority of developmental investigations of

amygdala–mPFC function have studied cross-sectional samples. Because cross-sectional studies are susceptible to cohort effects and cannot measure within-participant change, longitudinal work has been recommended for better charting of developmental trajectories (Crone & Elzinga, 2015; Madhyastha et al., 2017).

Here, we used multiverse analyses to examine the robustness of developmental changes to varied analytical decisions. We focused on task-related amygdala–mPFC functional development in an accelerated longitudinal sample ranging from ages 4-22 years. We selected a task that was designed to be appropriate for young ages to characterize developmental change in amygdala–mPFC responses to fear and neutral faces across childhood and adolescence, and we asked whether findings were robust to analytical choices. This accelerated longitudinal design is an extension of the sample reported on by Gee et al. (2013). We preregistered two hypotheses (https://osf.io/8nyj7/) predicting that both amygdala reactivity (1) and amygdala–mPFC connectivity (2) as measured with generalized psychophysiological interaction models (gPPI), would decrease as a function of age during viewing of fearful faces relative to baseline (see Table 1 Aims 1a & 2a).

Although we did not preregister further hypotheses, we also investigated age-related changes in within-scan differences in amygdala responses across trials and FC using beta series correlations. As previous work identified associations between amygdala–mPFC FC and separation anxiety (Carpenter et al., 2015; Gee et al., 2013), we asked whether any amygdala–mPFC measures were associated longitudinally with separation anxiety behaviors (see Table 1 Aim 3). We used ‘multiverse’ analyses and specification curves to examine the impact of analytical decisions on results. We also investigated test-retest reliability of all brain measurements across longitudinal study visits, given the importance of such reliability for interpreting individual differences or developmental change (Herting et al., 2017). Our multiverse approach allows us to thoroughly explore the robustness of different findings to

analytical choices, highlighting the importance of considering both robustness and reliability in developmental research.

Insert Figure 1 about here

Insert Table 1 about here

Methods

Before completing analyses, we preregistered methods for the current study through the Open Science Framework at https://osf.io/8nyj7/. Only analyses for age-related changes in amygdala reactivity and amygdala–mPFC gPPI were preregistered in detail (see Table 1 Aims 1a & 2a), and we did not preregister multiverse analyses. Methods detailed below include both information included in the preregistration and additional information and analyses not preregistered. Analysis code & documentation can be found at https://github.com/pab2163/amygdala_mpfc_multiverse.

Participants

Participants were recruited as part of a larger study examining brain development as a function of early life caregiving experiences. The current sample (N=98; 55 female, 43 male) included typically developing children, adolescents, and young adults covering the ages 4-22

99 years-old (M = 11.9 years old) who enrolled to participate in a study on emotional development.
100 All participants were reported to be physically and psychiatrically healthy (no medical or
101 psychiatric disorders), as indicated by a telephone screening before participation, and free of
102 MRI contraindications. All except 4 participants fell below clinical cutoffs (see sFigure 2) on the
103 Child Behavior Checklist (CBCL) Total Problems, Internalizing Problems, and Externalizing
104 Problems scales (Achenbach, 1991). The larger study also included youths with a history of
105 institutional and/or foster care outside of the United States, who are not included here.
106 Participants from the greater Los Angeles area were recruited through flyers, state birth records,
107 community events, online advertising, lab website and newsletters, psychologists' offices,
108 psychology courses at a local university (participants ages 18-22 years old only), and word-of-
109 mouth. Each participant completed up to 3 MRI scans spaced at an average interval of 18
110 months between visits. Parents provided written consent, children 7+ years old gave written
111 assent, and children under 7 years old gave verbal assent. Study protocols were approved by
112 the local university institutional review board. These data were collected between 2009 and
113 2015.

114 An accelerated longitudinal design was used such that participants' starting ages at scan
115 1 comprised the entire range of sample ages (4-22 years old), and coverage was approximately
116 balanced across the entire age range (see Figure 1B). The design was structured into 3 study
117 'waves' corresponding with recruitment efforts and visit protocols. Participants were
118 overenrolled at wave 1 to account for anticipated attrition (e.g., braces, relocation, etc) to
119 achieve the desired sample size across the three waves. While most participants were recruited
120 such that their first scan occurred at wave 1, a smaller set of participants were recruited at wave
121 2, such that some participants completed their first scan at wave 2 (see Figure 1). For these
122 participants, only 2 scans were planned.

Of the 191 participants participating in the longitudinal study, 140 completed at least one MRI scan. After exclusions for incomplete task runs (including falling asleep), computer errors resulting in missing stimulus timing files, high head motion, and failed visual QA (scanner/motion artifacts), a final sample of 98 participants (N = 183 total scans) was included for analysis (see Figure 1). Exclusion criteria were preregistered after conducting preliminary preprocessing, but before construction of group-level models and multiverse analysis plans. This sample included 40 participants with 1 scan, 31 with 2 scans, and 27 with 3 scans (one more participant than preregistered due to an initial coding error). Wave 1 data from forty-two of these participants were reported on by Gee et al. (2013).

The median annual household income for participating families was $85,001-$100,000 (for reference, median annual household income in Los Angeles County from 2015-2019 was $68,044; US Census Bureau, 2021). Epidemiological methods were not used to recruit a sample representative of the Los Angeles or United States populations (Heeringa et al., 2004), and Hispanic or Latinx participants were particularly underrepresented. Further sample demographics can be found in the supplementary materials (see sTables 2-3, sFigures 1-2).

Separation Anxiety

For each participant (except for 10 adults 18-22 years), a parent completed both the Revised Children's Anxiety and Depression Scale (RCADS-P) and the Screen for Child Anxiety Related Emotional Disorders (SCARED-P) to assess the frequency of symptoms of anxiety and low mood (Birmaher et al., 1999; Chorpita et al., 2000). Following prior work suggesting associations between task-evoked amygdala–mPFC functional connectivity and separation anxiety (Carpenter et al., 2015; Gee et al., 2013), we used the separation anxiety subscales from both the SCARED-P and RCADS-P as measures of anxiety-related behaviors in asking whether such functional connectivity may be linked to anxiety levels during childhood and

adolescence. For 11 participants who had missing items on the SCARED-P, indicating parents had skipped or forgotten to answer a question, we imputed responses using 5-Nearest Neighbor imputation using only the other items included in the SCARED-P separation anxiety subscale (Beretta & Santaniello, 2016). As expected, raw separation anxiety scores on both measures decreased as a function of age, while standardized scores (which are normed based on gender and grade level) were consistent across development with few children at or near clinical threshold (see Figure 6).

Emotion Discrimination Task

Participants completed either two (at wave 3) or three (at waves 1 and 2) runs of a modified ‘go/no-go’ task with emotional faces during fMRI scanning. Runs varied by emotional expression (fear, happy, sad), and within each run participants viewed emotional faces interspersed with neutral faces. To ensure that participants were paying attention, they were asked to press a button whenever they saw a neutral face (no response was required for any other face expression). The order of the runs was counterbalanced across participants; the stimuli within each run were pseudorandomized (Wager & Nichols, 2003) to allow for event-related estimates of the hemodynamic response, and fixed across participants. For the present analysis, only the fear run of the task was used. The other two runs, which used happy and sad faces in place of fear, are not included in the present analysis as these conditions were not present at all waves of data collection. As 50% of trials were ‘go’ trials under this paradigm, we refer to the task as an emotion discrimination task, rather than a true ‘go/no-go’ paradigm since there was no strong prepotent motor response. Stimuli within each run were presented with a jittered ITI (3-10s, Median = 4.93s]) according to a genetic algorithm with a fixation cross on the screen (Wager & Nichols, 2003). Face images were adult White female faces from the

171 Karolinska Directed Emotional Faces database (Calvo & Lundqvist, 2008), and the same face
172 stimuli were used across longitudinal study visits (Vijayakumar et al., 2019). Each run (130 TRs,
173 duration of 4:20) consisted of 48 trials (24 neutral faces, 24 fearful faces), each presented for
174 350ms. All fMRI analyses of this task used event-related designs.

175

176 *MRI Acquisition*

177 Participants under 18-years-old completed a mock scanning session before the MRI
178 scan to acclimate to the scanner environment and practice lying still for data collection. Waves 1
179 and 2 were collected on a Siemens 3T TIM Trio MRI scanner using a standard radiofrequency
180 head coil. A 2D spin echo image (TR, 4000 ms; TE, 40 ms; matrix size, 256 x 256; 4 mm thick;
181 0mm gap) was acquired in the oblique plane to guide slice configuration in both structural and
182 functional scans. A whole-brain high-resolution T1-weighted anatomical scan (MPRAGE; 256 x
183 256 in-plane resolution; 256mm FOV; 192 x 1 mm sagittal slices) was acquired for each
184 participant for registration of functional data. The task was presented through MR-compatible
185 goggles during scanning. T2*-weighted echoplanar images (interleaved slice acquisition) were
186 collected at an oblique angle of ~30 degrees (130 volumes/run; TR=2000ms; TE=30 ms; flip
187 angle=90°; matrix size=64 x 64; FOV=192 mm; 34 slices; 4 mm slice thickness; skip=0 mm).
188 Wave 3 was collected on a Siemens 3T TIM Trio MRI scanner at a different location using
189 identical acquisition parameters.

190

191 *Behavioral Analyses*

192 We used multilevel logistic regression models to estimate age-related changes in
193 several task performance metrics. We fit separate models for the d' performance metric, overall
194 accuracy (probability of a correct response on any trial), hit rate (on neutral face trials), and false

alarm rate (on fear face trials) as the respective outcomes, and included nested random effects for task sessions within participants (models were not nested for d' as this analysis used only 1 metric per session rather than trial-wise outcomes, but still included random effects for participants). Additionally, to model age-related change in reaction times during correct hit trials, we fit linear, quadratic, cubic, and inverse age (1/age; Luna et al., 2004, 2021) regressions with identical random effects structures. Model equations and results for all behavioral analyses can be found in the supplement (see supplemental methods p.12-14, sFigures 3-4).

Preregistered fMRI Pipeline

3dskullstrip from the Analysis of Functional NeuroImages (AFNI, v20.1.16) software package (Cox, 1996) was first run on all MPRAGE scans. Next, experimenters checked the quality of the skull stripping. If there were outstanding issues with a particular scan run (areas of brain tissue cut off, or significant areas of skull left in, 30/195 scans), FSL's brain extraction tool (BET; Jenkinson et al., 2012) was used instead. We used robust brain center estimation, and modified the fractional intensity values between 0.5-0.7 to optimize quality. Slice-time correction was not used. Timeseries of the 6 motion parameters were calculated and subsequent spatial realignment of BOLD volumes was completed using MCFLIRT in FSL (Jenkinson et al., 2002). Scans over a threshold of >40 volumes with > .9mm framewise displacement (framewise displacement calculated as the sum of absolute frame-to-frame differences between head realignment estimates; Power et al., 2012) were excluded from analysis (12 out of an initial 195, or 6.2%). After this exclusion, an average of 96.7% (range = [70.1-100%]) of stimulus-coincident volumes in each scan were below the 0.9mm framewise displacement threshold. The mean age of participants with excluded scans was 7.16 and 8/12 were male. Registration matrices were calculated for registration of functional images to high-resolution structural T1 images using FSL's FLIRT with boundary-based registration. Registration matrices for standard MNI space

were also calculated using both FLIRT (linear registration) and FNIRT (nonlinear registration) with 12 DOF and a warp resolution of 10mm. Data were high-pass filtered at .01Hz and smoothed with an isotropic Gaussian kernel with FWHM of 6mm before running general linear models (GLMs), and 4d volumes were grand mean scaled such that the average intensity value was 10000.

Following preprocessing, we ran scan-level GLMs using FSL's FEAT (v6.00). We included event-related regressors for fear and neutral faces (convolved with a double-gamma HRF), their temporal derivatives (Pernet, 2014), and 24 head motion nuisance regressors (the 6 head realignment parameters, their temporal derivatives, and their squares (Power et al., 2012). Volumes with FD > .9mm were downweighed to 0 in the GLM. Pre-whitening was used to estimate and remove temporal autocorrelation (Woolrich et al., 2001). For each scan, we calculated fear > baseline, neutral > baseline, and fear > neutral contrasts. We used native-space bilateral amygdala masks generated using Freesurfer (v6.0; Fischl, 2012) by VanTieghem et al. (2021).

Multiverse Analyses and Specification Curves

In addition to the preregistered pipelines, we conducted multiverse analyses to address all aims in Table 1 and constructed sets of separate specification curves for each aim (see Table 2). In general, multiverse analyses aim to probe the consistency of results across all 'reasonable' possible combinations of analysis decisions (i.e. simultaneously taking all possible 'forking paths')(Steegen et al., 2016). Because analyzing fMRI data using all reasonable specifications was infeasible (i.e., possibilities are virtually infinite), we took the approach of 'sampling' from the many reasonable or commonly-used analysis choices for each multiverse. Despite not being completely comprehensive, this approach still allowed for thorough investigation into the robustness of results. For all multiverse analyses, we constructed

specification curves by ranking models by their beta estimates (ascending) for parameters of interest for interpretation and visualization (Cosme & Lopez, 2020; Klapwijk et al., 2019; Orben & Przybylski, 2019; Simonsohn et al., 2015, 2020). Because specification choices were not preregistered, we did not conduct formal null hypothesis testing of specification curves. Instead, as continuous measures of evidence, we report the proportion of specifications resulting in an estimate of the same sign, as well as the proportion of specifications resulting in 95% posterior intervals excluding 0 in the same direction. In addition, to analyze in more detail the impact of specific choices, we submitted point estimates for parameters of interest across all specifications to multiple regression models. From these models, we examined the conditional effects of each analysis decision point on the parameter of interest (see supplemental methods p.30-31, sFigures 11-13, 41-43, & 55-57).

Multiverse amygdala reactivity analyses

For amygdala reactivity analyses, we examined the robustness of age-related change estimates to a variety of analytical decisions. In addition to the preregistered FSL-based pipeline, we preprocessed data using C-PAC software (v1.4.1; Craddock et al., 2013). We used C-PAC to take advantage of features supporting running multiple pipeline ‘forks’ in parallel (for example multiple nuisance regression forks using the same registration). No spatial smoothing was used in C-PAC pipelines (see supplemental methods p.14). Following C-PAC and FSL preprocessing, we examined the impact of different sets of commonly-used analysis methods on age-related change in amygdala reactivity. We varied analysis choices of GLM software, hemodynamic response function, nuisance regressors, first-level GLM estimates, amygdala ROI, exclusion criteria (exclude vs. include scans analyzed by Gee et al., 2013), group-level model outlier treatment, and group-level model covariates (see Table 2 & supplemental methods p.14-17). Multiverse analyses of amygdala reactivity included a total of 2808 model

specifications (156 ways of defining participant-level amygdala reactivity x 18 group-level model specifications) for each contrast. We analyzed all specifications in parallel. In addition, we examined nonlinear age-related changes using quadratic and inverse age models (see sFigures 14-17) and ran a smaller set of analyses (sFigure 19) to ask whether we could differentiate within-participant change over time from between-participant differences through alternative model parametrization (see supplemental methods p.19).

For all specifications, individual-level amygdala reactivity estimates were submitted to a group-level multilevel regression model for estimation of age-related changes. All models allowed intercepts to vary by participant, and some specifications also allowed for varying slopes (see supplemental methods p.15 for model syntax). All models also included a scan-level covariate for head motion (mean framewise displacement [FD]; Power et al., 2012; Satterthwaite et al., 2012, 2013). Consistent with prior work, head motion was higher on average in younger children, and decreased with age (see sFigure 5), though head motion was not associated with amygdala reactivity estimates for most specifications (see sFigure 26). Age-related change findings examined for the preregistered pipeline also remained consistent under more stringent exclusion thresholds based on mean framewise displacement (see sFigure 27). Across preprocessing specifications, we also examined within-scan similarity of amygdala and whole-brain voxelwise reactivity patterns (see sFigures 20-21) and between-scan correlations of average amygdala reactivity estimates (sFigures 22-24).

Change in amygdala reactivity across trials

To probe whether amygdala reactivity exhibited within-scan change in an age-dependent manner, we modeled reactivity to each face trial using a Least Squares Separate method (LSS; Abdulrahman & Henson, 2016). After preprocessing, we used FEAT to fit 48 separate GLMs corresponding to each trial in each scan. A given trial was modeled with its own regressor and

295 the remaining 47 trials were modeled with a single regressor. Each GLM also included 24 head
296 motion nuisance regressors and had TRs with framewise displacement > .9mm downweighted
297 to 0. BOLD data were high-pass filtered at .01Hz before the GLM. From each of the 48 GLMs,
298 we extracted the mean amygdala beta estimates corresponding to a contrast for each single
299 trial > baseline.

300 We constructed separate multiverse analyses using three different methods for
301 measuring change in amygdala reactivity across trials. For method 1 (*slopes*), we measured
302 rank-order correlations between trial number and trial-wise amygdala betas. For method 2 (*trial*
303 *halves*), we split trials into the first half (trials 1-12) and second half (trials 13-24), and modeled
304 age-related change in each half. For method 3 (*single-trial models*), we constructed larger
305 multilevel models with individual trials as the unit of observation. We conducted several analysis
306 specifications for each method (see Table 2 & supplemental methods p.21-23), and generated
307 corresponding specification curves.

308

309 *Multiverse amygdala–mPFC functional connectivity (FC) analyses*

310 We applied multiverse analysis techniques towards examining age-related changes in
311 amygdala–mPFC FC using gPPI and beta-series correlation (BSC) methods. Briefly, gPPI
312 estimates functional connectivity by constructing an interaction term between the timecourse in
313 a seed region of interest and a stimulus (task) regressor. Voxels whose activity is well fit by this
314 interaction term (a psychological-physiological interaction, or PPI) are assumed to be
315 "functionally coupled" with the seed region in a way that depends on the behavioral task
316 (McLaren et al., 2012; O'Reilly et al., 2012). BSC offers a different way of estimating functioning
317 connectivity, by constructing 'timeseries' of beta values (i.e., a beta series) in a condition of
318 interest for two regions of interest, and calculating the product-moment correlation between
319 those beta series.

320 We constructed separate specification curves for age-related change in gPPI and BSC
321 for each contrast. Across gPPI specifications, we varied whether to use a deconvolution step in
322 creating interaction regressors (Di & Biswal, 2017; Gitelman et al., 2003), as well as several
323 other analysis decision points (see Table 2 & supplemental methods p.24-25). The
324 deconvolution step applies to the preprocessed BOLD data from the seed timecourse: these
325 data are first deconvolved to estimate the "underlying neural activity" that produced the BOLD
326 signal (Gitelman et al., 2003), then these deconvolved signals are multiplied with the task
327 regressor (e.g., for fear faces). Finally, this new interaction term is convolved with a
328 hemodynamic response to produce the BOLD functional connectivity regressor of interest.
329 Given recent work indicating that centering the task regressor before creation of the interaction
330 term can mitigate spurious effects (Di et al., 2017), we also compared pipelines in which we
331 centered the task regressor before deconvolution (pipelines including deconvolution in main
332 analyses did not include this step; see sFigure 44).
333 We preregistered constructing an mPFC ROI containing 120 voxels centered at the peak
334 coordinates reported by Gee at al. (2013) for age-related change in fear > baseline gPPI
335 (Talairach 2,32,8; or MNI 3,35,8). However, after preregistration we discovered that these peak
336 coordinates were not at the center of the ROI reported by Gee at al. (2013), and were quite
337 close to the corpus callosum. The 120-voxel ROI we created that was centered at this peak
338 coordinate would have contained a high proportion of white matter voxels relative to cortical
339 voxels (though this was not true for the mPFC ROI identified by Gee et al. (2013). To address
340 this issue, we instead constructed three spherical ROIs with 5mm radii; the first centered at the
341 above peak coordinates, the second shifted slightly anterior, and the third shifted slightly ventral
342 relative to the second (see Figure 4). Lastly, to examine amygdala functional connectivity with a
343 more broadly-defined mPFC, we also used a 'large vmPFC' mask encompassing many of the

areas within the ventromedial prefrontal cortex derived from Mackey & Petrides (Mackey & Petrides, 2014).

For BSC analyses, we used beta estimates from the LSS GLMs described above for analyses of within-scan change in amygdala reactivity. Across BSC specifications we varied analyses across several decision points (see Table 2 & supplemental methods p.26), including whether to include a correction for global signal (post-hoc distribution centering(Fox et al., 2009)). We extracted mean beta estimates for amygdala and mPFC ROIs for each trial, then calculated product-moment correlations between the timeseries of betas across trials (neutral and fear separately) for both regions (Di et al., 2020). These correlation coefficients were transformed to z-scores, then submitted to group-level models.

Age-related changes in gPPI and BSC were estimated using multilevel regression models as described for the amygdala reactivity analyses. We focused primarily on linear age-related change, but also examined quadratic and inverse age associations (see sFigures 45-48 & 58-61). We separately examined group mean gPPI and BSC for each contrast (see sFigures 38 & 52), as well as associations between mean framewise displacement and both FC measures across specifications (see sFigures 49 & 62). Additionally, we examined mean estimates and age-related change in 'task-independent' FC as measured by beta weight of the 'physio' term from the seed amygdala timeseries within the gPPI model (representing baseline amygdala–mPFC functional connectivity controlling for task-induced variance; sFigures 50-51).

Multiverse analyses of associations between amygdala–mPFC circuitry and separation anxiety behaviors.

We used further multiverse analyses to ask whether amygdala reactivity, change in amygdala reactivity over the course of the task, or amygdala–mPFC FC were associated with separation anxiety behaviors. Separate specification curves were created for each brain

measure type (amygdala reactivity, amygdala reactivity change across trials, amygdala–mPFC FC). All analyses used multilevel regression models with covariates for age, and specification curves included both RCADS-P and SCARED-P separation anxiety subscales as outcomes (see Table 2 & supplemental methods p.29-30). Because we did not have parent-reported RCADS-P or SCARED-P scores for 10 adult participants, these analyses had an N=173.

Reliability analyses

To better understand the proportion of variance in each measure explained by the grouping of observations within repeated measurements of the same participants over time, we computed Bayesian intraclass correlation (ICC) estimates through variance decomposition of the posterior predictive distributions of the multilevel regression models previously described. We implemented these through the 'performance' R package (Lüdecke et al., 2021; Nakagawa et al., 2017). Negative ICC estimates under this method are possible, and indicate that the posterior predictive distribution has higher variance when not conditioning on random effects than when conditioning on them (likely indicating the posterior predictive variance is large, and random effects explain very little of this variance).

Model-fitting

All statistical models fit at the group level were run in the R (v 3.6.1) computing environment. In order to most accurately model age-related changes in each of our measures, we attempted to take into account both between-participants information and repeated measurements within participants over time. Unless otherwise indicated, models were estimated

using Hamiltonian Markov chain Monte Carlo sampling as implemented in the Stan programming language through the brms package in R (Bürkner, 2019; Gelman et al., 2015). Unless otherwise indicated, all models used package default weakly-informative priors (student-t distributions with mean 0, scale parameter of 10 standardized units, and 3 degrees of freedom for all fixed effects), and were run with 4 chains of 2000 sampling iterations (1000 warmup) each (see supplemental methods p.18-19 and p.30 for syntax).

Interactive visualizations

Because static plots visualizing the model predictions for all models in each multiverse would require far more page space than available, we created web-based interactive visualization tools for exploring different model specifications and viewing the corresponding raw (participant-level) data and fitted model predictions using R and Shiny (Beeley, 2013). These visualizations can be found at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/

Deviations from preregistration

Although we largely completed the preregistered analyses, the current study includes many analyses beyond those proposed in the initial preregistration. Because the additional analyses (i.e., all multiverses) conducted here give us substantial analytical flexibility over that initially indicated by preregistration, we consider all results here to be at least in part exploratory (rather than completely confirmatory), despite the preregistered hypotheses. Additionally, we note that BSC analyses, analyses of change in amygdala reactivity across trials, and analyses of associations between all brain measures and separation anxiety were exploratory, and conducted after we had seen the results of the preregistered reactivity and gPPI analyses. In addition, to avoid possible selection bias introduced by the analytical flexibility inherent in running many parallel analyses, we consider all analysis specifications simultaneously,

416 emphasizing that without further methodological work, we consider all such choices in tandem
417 as providing equal evidential value. While reliability analyses were not preregistered, they too
418 provide key information for interpreting the current analyses.
419
420 **Results**
421
422 *Age-related change in amygdala reactivity*
423 We used multilevel regression models and specification curve analyses to examine age-
424 related changes in amygdala reactivity to faces in an accelerated longitudinal sample ranging
425 from ages 4-22 years (Figure 2). Across specifications, we found relatively consistent evidence
426 for negative age-related change in anatomically-defined (Harvard-Oxford atlas and Freesurfer-
427 defined) amygdala reactivity to fear faces > baseline, such that the vast majority of analysis
428 specifications (99.6%) estimated linear slopes at the group level that were negative in sign, and
429 the majority (60.0%) of 95% posterior intervals about these slopes excluded 0 (Figure 2A;
430 interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/). Thus, over half
431 of models indicated that on average, increases in age were associated with decreases in
432 amygdala reactivity to fear faces > baseline. Because the timepoint 1 data in the current study
433 included the 42 scans used by Gee et al. (2013) to age-related changes in amygdala—mPFC
434 circuitry for the fear > baseline contrast, results including these scans may have been more
435 likely to find similar change (particularly for fear > baseline, see sFigures 11-13, & 25).
436 Estimated age-related change was on average weaker, though still largely negative (98.1%
437 negative, 25.3% of posterior intervals excluding 0) when 42 previously analyzed scans (ages 4-
438 17 years) were excluded to provide stricter independence from previously analyzed data (see
439 sFigure 11, Gee et al., 2013). Estimated average age-related change for the fear > baseline
440 contrast was somewhat stronger when using a right amygdala ROI compared to the left

amygdala, and when using t-stats extracted from scan-level GLMs rather than beta estimates for group-level models (see sFigure 11).

Parallel multiverse analyses found similarly consistent age-related decreases in neutral faces > baseline amygdala reactivity (see Figure 2C for an example pipeline & sFigure 9 for specification curve), but no consistent evidence for age-related change for the fear > neutral contrast (see sFigure 10). However, there was consistent evidence for higher reactivity for fear faces > neutral on average as well as each emotion compared to baseline (sFigures 6-8), indicating that while the amygdala responses were robust and generally stronger for fear faces compared to neutral, such fear > neutral differences did not change with age. Across contrasts, varying the inclusion of block order or scanner covariates, inclusion of random intercepts, and use of robust regression models had little impact on age-related change estimates (see Figure 2B sFigures 6-8).

While group-level estimates of average age-related change were relatively consistent across specifications, the estimated age terms in these models could be influenced by both within-participant change and between-participant differences (King et al., 2018; Madhyastha et al., 2018). A smaller separate specification curve indicated that when models were parametrized to differentiate within-participant change and between-participant differences, average within-participant change was not consistent across specifications and could not be estimated with precision (Figure 2D). In contrast, estimates of between-participant differences largely indicated negative age-related change in concurrence with our initial model parametrization. At the same time, within-participant versus between-participant terms were not reliably different from one another, indicating that models could not distinguish them despite higher precision for estimating between-participant differences (see sFigure 19). We did not find consistent evidence for quadratic age-related changes in amygdala reactivity (see sFigures 14-17). Inverse age models (i.e. amygdala reactivity modeled as a function of 1/age) indicated results similar to

466 those of linear and quadratic models with most specifications for the fear > baseline and neutral
467 > baseline (though less consistent) contrasts indicating age-related decreases (see sFigure 18).
468
469
470 _____
471 Insert Figure 2 about here
472 _____
473
474 *Age-related differences in within-scan amygdala reactivity change*
475 To ask whether age-related changes in amygdala reactivity could be due to
476 developmental changes in patterns of amygdala reactivity across face trials (within a run), we
477 examined whether within-scan change in amygdala reactivity varied with age (see Table 1 Aim
478 1b). Analyses included 42 specifications (3 amygdala regions of interest [ROIs] x 2 global signal
479 correction options x 7 group-level models). Across both fear and neutral trials, linear slopes of
480 amygdala reactivity were negative on average, indicating higher amygdala reactivity at the
481 beginning of the run (Figure 3A, sFigure 30). Across specifications, for both fear (100% of
482 estimates had the same sign, 95.2% of posterior intervals excluding 0 in the same direction) and
483 neutral trials (100% of estimates in the same direction, 38.1% of posterior intervals excluding 0),
484 there was evidence that these within-scan slopes were steeper (i.e., more negative) at younger
485 ages, though evidence was relatively weaker for neutral trials (Figure 3D-E). Specifications with
486 a global signal subtraction step also tended to find stronger age-related change.
487 Similarly, when splitting trials into the first half (trials 1-12) versus second half (trials 13-
488 24), there was consistent evidence (100% of estimates had the same sign, 69.2% with posterior
489 interval excluding 0) for an interaction between age and trial half, such that average reactivity to
490 fear faces > baseline in the first half of trials decreased as a function of age more so than did

average reactivity during the second half of trials (see Figure 3B, sFigure 32). This interaction was in the same direction for neutral trials across most specifications (88.5% of estimates), but was typically not as strong (3.8% of posterior intervals excluding 0). Single-trial models indicated similar age-related change in within-scan amygdala dynamics (see Figure 3C, sFigures 33-34). Mean group-level amygdala reactivity was higher for the first half of trials for fear faces > baseline across several specifications, though there were not consistent differences between trial halves for mean amygdala reactivity to neutral faces (sFigure 31).

Insert Figure 3 about here

Age-related change in task-evoked amygdala–mPFC functional connectivity

We used multilevel regression modelling and specification curve analyses to examine age-related change in task-evoked amygdala–mPFC functional connectivity within the accelerated longitudinal cohort (see Table 1 Aims 2a-b). For the fear > baseline contrast, a specification curve with 288 total specifications (4 definitions of participant-level gPPI estimates x 4 mPFC ROIs x 18 group-level models) of amygdala–mPFC gPPI did not find consistent evidence of age-related change: while 59.0% of models found point estimates in the positive direction, only 23% of posterior intervals excluded 0 (Figure 4C-D, interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/). Specification curve analyses found that the sign of the estimated age-related change depended almost entirely on deconvolution, such that most specifications including a deconvolution step resulted in negative age-related change estimates never distinguishable from 0 (78.5% of point estimates negative, 0% of posterior intervals excluding 0), and most specifications not including a deconvolution step

516 resulted in positive age-related change estimates (96.5% of point estimates positive, 47.9% of
517 posterior intervals excluding 0). A visualization of the effects of the deconvolution step on
518 amygdala FC with each of four mPFC ROIs is presented in Figure 4B. While mPFC ROI
519 definition and other analysis decision points also influenced estimates of age-related change in
520 gPPI (Figure 4D), follow-up regression models indicated that the effect of including the
521 deconvolution step was several times larger for the fear > baseline contrast (see sFigures 41-
522 43).

523 Through equivalent multiverse analyses we also found no evidence of consistent linear
524 age-related change in amygdala–mPFC gPPI for the neutral > baseline and fear > neutral
525 contrasts (see sFigures 39-40), or nonlinear change for any contrast (see sFigures 45-48). In
526 addition, we did not see consistent evidence for group average amygdala–mPFC gPPI for any
527 contrast, though such results often differed as a function of whether a deconvolution step was
528 included (see sFigure 38). Though we included gPPI analysis specifications excluding the 42
529 scans at timepoint 1 studied by Gee et al. (Gee et al., 2013), exclusion of these scans had little
530 impact on age-related change results (see Figure 4D).

531

532 _____

533 Insert Figure 4 about here

534 _____

535

536 In addition to gPPI analyses, we used beta series correlation (BSC) analyses to examine
537 age-related changes in task-evoked amygdala–mPFC connectivity (see Table 1 Aim 2b). As
538 with gPPI, multiverse analyses of amygdala–mPFC BSC (168 total specifications; 3 amygdala
539 ROI definitions x 4 mPFC ROI definitions x 2 global signal options x 7 group-level models) for
540 fear trials (vs baseline) did not yield strong evidence of age-related change across pipelines

541 (84.5% of point estimates in the same direction, 24.4% of posterior intervals excluding 0; Figure
542 5A, interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse). Unlike gPPI
543 analyses, however, choice of mPFC ROI (as well as amygdala ROI, though this was not
544 examined for gPPI) most impacted age-related change in BSC estimates, rather than
545 preprocessing or modeling analytical choices (Figure 5B, sFigures 55-57). Accordingly, while
546 global signal subtraction resulted in weaker amygdala–mPFC BSC on average (see sFigure
547 52), inclusion of this step did not consistently affect age-related change estimates (Figure 4C).
548 We did not find consistent evidence for age-related change in amygdala–mPFC BSC for neutral
549 trials (vs baseline), or for fear > neutral trials (sFigures 53-54). We did not find consistent
550 evidence for nonlinear age-related change for any contrast (sFigures 58-61).

551 Additionally, we constructed a correlation matrix using rank-order correlations of scan-
552 level BSC and gPPI estimates for the fear (vs baseline) condition. Across scans, there was little
553 evidence of correspondence between BSC and gPPI metrics for amygdala–mPFC connectivity
554 (Figure 5D, sFigures 63-66). Further, FC estimates tended to be positively correlated within a
555 method type (BSC, gPPI) across mPFC ROIs, though less strongly for gPPI estimates with
556 versus without a deconvolution step.

557 In addition to gPPI and BSC methods for functional connectivity, we also explored
558 between-scan associations between amygdala reactivity and mPFC reactivity (sFigures 28-29).
559 Multilevel models indicated that amygdala reactivity for fear faces > baseline was positively
560 associated with mPFC reactivity for fear faces > baseline for all mPFC ROIs, though we did not
561 find consistent evidence for age-related changes in associations between amygdala and mPFC
562 reactivity to fear faces > baseline (see sFigure 29).

563

564 _____

565 Insert Figure 5 about here

566 _____

567

568

569 *Amygdala–mPFC Measures & Separation Anxiety*

570 We conducted multiverse analyses of associations between several amygdala–mPFC
571 measures (amygdala reactivity, amygdala–mPFC FC, within-scan changes in amygdala
572 reactivity) and separation anxiety behaviors (see Table 1 Aim 3). Separation anxiety behaviors
573 on average decreased with age, as indicated by the RCADS-P and SCARED-P raw scores
574 (Figure 6A-C). Neither specification curves for amygdala reactivity (18 total specifications, 56%
575 of point estimates in the same direction as median estimate, 0% of posterior intervals excluding
576 0), amygdala–mPFC gPPI FC (90 total specifications, 72% of point estimates in the same
577 direction as median estimate, 1% of posterior intervals excluding 0), amygdala–mPFC BSC FC
578 (18 total specifications, 83% of point estimates in the same direction as median estimate, 0% of
579 posterior intervals excluding 0), nor slope of amygdala responses across trials (12 total
580 specifications, 75% of point estimates in the same direction as median estimate, 17% of
581 posterior intervals excluding 0), found consistent evidence for associations between brain
582 measures and separation anxiety. Similar specification curves found little consistent evidence
583 for associations between brain measures and generalized anxiety, social anxiety, or total
584 anxiety behaviors (see sFigure 67). All specifications controlled for age (see supplemental
585 methods p.30).

586 To more specifically follow up on previous work reporting associations between
587 separation anxiety behaviors and amygdala–mPFC gPPI for fear > baseline specifically (Gee et
588 al., 2013), we plotted model predictions for such models from the above multiverse analysis for
589 each of the four mPFC ROIs, across all three separation anxiety outcome measures, and both
590 with and without a deconvolution step (Figure 6E). We did not find consistent evidence for

associations with separation anxiety, and results showed high sensitivity to the deconvolution step, mPFC ROI, and outcome measure used.

Insert Figure 6 about here

Reliability:

To examine test-retest reliability estimates of amygdala—mPFC measures across longitudinal visits, we computed Bayesian ICC estimates using a variance decomposition method (Lüdecke et al., 2021). Because such models can accommodate missing data, all observations (98 participants, 183 total scans) were used, including participants with only 1 visit. All amygdala reactivity (Figure 7A) and amygdala—mPFC functional connectivity (Figure 7C) measures, as well as slopes of amygdala reactivity estimates across trials (Figure 7B), demonstrated poor reliability (ICC < 0.4; Cicchetti & Sparrow, 1981; Elliott et al., 2020). Separation anxiety measures demonstrated somewhat higher, though still largely poor reliability (point estimates ~0.4, 95% CIs included values below 0.4; Figure 7D). Head motion in the scanner (mean framewise displacement) showed the highest reliability (ICC = 0.52, 95% CI [0.29, 0.68]),

---

Insert Figure 7 about here

Discussion

Measures that are both robust to researcher decisions and reliable across measurement instances are critical for studies of the human brain (Botvinik-Nezer et al., 2020; Bowring et al., 2019; Elliott et al., 2020; T. Xu et al., 2022). The accelerated longitudinal design

and multiverse analysis approach used in the current study allowed a rare opportunity to examine both reliability and robustness of amygdala—mPFC measures using a rapid event-related face task from early childhood through young adulthood. Overall, estimates for age-related change in amygdala reactivity were relatively robust to a variety of analytical decision points, while age-related change estimates for amygdala—mPFC connectivity were more sensitive to researcher choices. gPPI analyses were particularly sensitive to whether a deconvolution step was applied. Yet, in concurrence with previous work (Elliott et al., 2020; Haller et al., 2022; Herting et al., 2017; Infantolino et al., 2018; Kennedy et al., 2021; Nord et al., 2017; Sauder et al., 2013), amygdala—mPFC measures displayed consistently poor test-retest reliability across many analytical specifications. While low reliability estimates in the present study may be due in part to the long (~18 months) test-retest interval (Elliott et al., 2020) and potential true developmental change (Herting et al., 2017), low reliability nevertheless imposes a major caveat towards interpretation of the current developmental findings.

The present findings are valuable from a methodological standpoint in evaluating the robustness of analytical tools used. A measurement can have high test-retest reliability yet low robustness (high sensitivity) to analytical decisions, or vice versa (Li et al., 2021). Because neither robustness nor reliability guarantee the other, current findings on the impacts of analytic choices will likely be informative in guiding future studies. Thus, we discuss each of the main analyses below, with particular emphasis on how findings are impacted by analytic choices.

Amygdala reactivity

While there were differences across model specifications, the majority of pipelines supported our hypothesis that amygdala reactivity to fearful faces decreases with age from

early childhood through early adulthood (see Table 1 Aim 1a). Across specifications, we found relatively robust evidence for age-related decreases in amygdala reactivity to both fearful and neutral faces (Figure 2A). Yet, findings also varied considerably across specifications. For example, only 60% of pipelines produced results that would be individually labeled as 'significant' (under α = .05), indicating that multiple investigations of this dataset could likely lead to qualitatively different conclusions. While over half of analyses found evidence consistent with studies indicating greater amygdala reactivity to fear faces > baseline in younger children (Forbes et al., 2011; Gee et al., 2013; Guyer et al., 2008; Swartz et al., 2014), the other 40% of specifications would have been consistent with investigations that found little age-related change (NB: there were also differences in samples, age ranges, task parameters, and behavioral demands across these studies; Kujawa et al., 2016; Wu et al., 2016; Zhang et al., 2019). We also found that different specifications resulted in somewhat different nonlinear trajectories (see sFigures 14-18). Not only did inverse age and quadratic age models find different trajectories (as would be expected), but quadratic trajectories themselves also displayed considerable analytic variability, with some specifications finding "convex" and others finding "concave" fits (see sFigure 17). Although estimating nonlinear age-related change was not a primary goal of the present study, future work should use model comparisons for better differentiating nonlinear patterns (Curran et al., 2010; Luna et al., 2021).

Models also found evidence for between-participant differences, but could neither identify within-participant change (Figure 2D) nor differentiate between-participant from within-participant estimates. As such, interpretation of the age-related change reported here is subject to many of the same limitations that apply to cross-sectional designs (Glenn, 2003), where age-related changes may not necessarily indicate 'true' developmental growth. High

uncertainty in estimating average within-participant change could be driven by several factors, including true heterogeneity in individual trajectories, low measurement reliability, scanner differences across longitudinal timepoints, or unmodeled variables impacting amygdala reactivity. Additionally, the within-participants terms represent a smaller age range (a maximum of 4 years for any given participant), relative to the broader age range assessed by the between-participants terms (18 years), which may have placed additional limits on identifying reliable within-participant change.

Age-related change in amygdala responses to fear faces over baseline seemed largely the result of earlier trials in the task (see sFigures 32-34). While differences in task design and contrast across studies have been highlighted as potential sources of discrepant findings on the development of amygdala function (Killgore & Yurgelun-Todd, 2007; Lieberman et al., 2007; Swartz et al., 2014), this result indicates that attention to trial structure and task duration may also be necessary in comparing studies. Because the paradigm used in the current study involved a task requiring participants to press for one face ('neutral') and not press for 'fear' faces, findings specific to fear faces over baseline under the current paradigm may also be driven by behavioral task demands.

Amygdala–mPFC Functional Connectivity

We did not find evidence for our second hypothesis, as neither gPPI nor BSC analyses indicated consistent evidence of age-related change in amygdala–mPFC functional connectivity (see Table 1 Aims 2a-2b, Figures 4-5). Thus, the age-related changes in task-evoked amygdala–mPFC connectivity identified in prior work (Gee et al., 2013; Kujawa et al., 2016; Wu et al., 2016) was not identified here, consistent with (Zhang et al., 2019). Crucially, however, our specification curves did not find strong evidence *against* such age-related change, as we did not observe precise and consistent 'null' estimates across specifications.

Additionally, quadratic and inverse age models did not find consistent evidence for nonlinear age-related change (see sFigures 45-48 & 58-61).

gPPI results were sensitive to whether a deconvolution step had been included in the preprocessing pipeline, such that we mostly found age-related decreases in amygdala–mPFC connectivity with a deconvolution step included, and age-related increases without it (although most pipelines would not have been 'statistically significant' on their own, see Figure 4B). While deconvolution has been argued to be a necessary step for event-related PPI analyses (Gitelman et al., 2003), recent work has shifted guidelines on its use, and it may not be recommended for block designs (Di et al., 2020; Di & Biswal, 2017). Because the true 'neuronal' signal underlying the BOLD timeseries within a given ROI cannot be directly measured, deconvolution algorithms are difficult to validate. Further, deconvolution may cause PPI results to be driven by baseline connectivity if task regressors are not centered (Di et al., 2017), although such centering did not have a major influence on age-related change results in the present analyses (see sFigure 44). Within the current study, small tweaks to AFNI's 3dTfitter algorithm for deconvolution resulted in vastly different regressors (see sFigure 36), suggesting the potential for high analytic variability even between gPPI analyses ostensibly using deconvolution. While the present study does not provide evidence that can inform whether or not deconvolution is recommended, further work is needed to optimize and validate applications of gPPI methods and selection of appropriate task designs. gPPI may be better equipped for block-designs and particularly ill-posed for rapid event-related tasks due to both difficulties in resolving which times within the BOLD timeseries reflect functional connectivity evoked by rapid (350ms) events and low statistical power in estimating such task-evoked connectivity (see sFigures 35-37; O'Reilly et al., 2012). Concurrent with previous work, beta series correlation analyses may have higher statistical power for identifying task-related connectivity signal than gPPI within event-related designs more generally (Cisler et al., 2014).

716 Age-related change estimates for amygdala—mPFC BSC showed somewhat higher
717 robustness to analytic decisions compared to gPPI. For BSC analyses, choice of mPFC ROI
718 contributed most to variability in age-related change estimates (see Figure 5B, sFigures 55-
719 57). While a global signal correction (post-hoc distribution centering) greatly decreased
720 *average* amygdala—mPFC BSC connectivity (see Figure 5D, sFigure 52) for both fear and
721 neutral faces, this analytical step did not impact age-related change estimates as heavily
722 (sFigures 55-57). The fact that global signal correction so dramatically decreased average
723 estimated amygdala—mPFC BSC may indicate that, like with resting-state fMRI analyses,
724 positive functional connectivity values are due in part to motion and physiology-related
725 confounds (Gratton et al., 2020; Power et al., 2019). Supporting this, BSC estimates were
726 correlated with mean framewise displacement across scans for the fear > baseline and neutral >
727 baseline contrasts only when a global signal correction was not applied (see sFigure 62). In
728 addition, while test-retest reliability for all BSC measures was poor, BSC estimates from
729 pipelines including a global signal correction step mostly demonstrated somewhat higher ICC
730 (Figure 6). While these results are consistent with prior work indicating that correcting for the
731 global signal can mitigate artifacts (Ciric et al., 2017; Satterthwaite et al., 2012), other work
732 indicates that such corrections also remove meaningful biological signals (Belloy et al., 2018;
733 Glasser et al., 2018; Yousefi et al., 2018).

734

735 *Amygdala–mPFC circuitry and separation anxiety*

736 We did not find associations between any task-related amygdala–mPFC measures
737 (reactivity or functional connectivity) and separation anxiety behaviors (see Table 1 Aim 3;
738 Figure 6). This finding stands in contrast to associations between amygdala–mPFC
739 connectivity and anxiety identified in previous developmental work (Gee et al., 2013;

740 Jalbrzikowski et al., 2017; Kujawa et al., 2016; Qin et al., 2014). However, given that analyses
741 of brain-behavior associations may require imaging cohorts much larger than the current
742 sample (especially considering the low reliability of the measures used; Grady et al., 2020;
743 Marek et al., 2020), the absence of relationships here may not be strong evidence against the
744 existence of potential associations between amygdala–mPFC circuitry and developing
745 anxiety-related behaviors.
746
747 *Advantages and pitfalls of the multiverse approach*
748 Our findings contribute to a body of work demonstrating that preprocessing and
749 modeling choices can meaningfully influence results (Botvinik-Nezer et al., 2020). Indeed,
750 most studies involving many analytical decision points could benefit from multiverse analyses.
751 Such specification curves can help to examine the stability of findings in both exploratory and
752 confirmatory research (Flournoy et al., 2020). Particularly when methodological 'gold
753 standards' have not been determined, specification curves may be informative for examining
754 the impacts of potential analysis decisions (Bridgeford et al., 2020; Dafflon et al., 2020).
755 Further, wider use of specification curves might help to resolve discrepancies between study
756 findings stemming from different analysis pipelines.
757 While specification curve analyses may benefit much future research, we also note
758 that multiverses are only as comprehensive as the included specifications (Steegen et al.,
759 2016), and such analyses alone do not solve problems related to unmodeled confounds,
760 design flaws, inadequate statistical power, circular analyses, or non-representative sampling.
761 Further, unless all specifications are decided *a priori*, analyses are vulnerable to problems of
762 analytic flexibility (Gelman & Loken, 2014), and inclusion of less justified specifications can
763 bias results (Del Giudice & Gangestad, 2021). Because specification curves can include
764 hundreds or thousands of individual analyses, rigorous evaluation of individual models can be

difficult. To this end, we created interactive visualizations for visual exploration of individual analysis specifications.

Computational resources are a relevant concern when conducting multiverse analyses as well. In the current study, preprocessing (registration in particular) was the most computationally intensive step, taking an estimated 4 hours of compute time per scan per pipeline using 4 cores on a Linux-based institutional research computing cluster. However, specification curve analyses themselves were relatively less intensive, with all group-level models of amygdala reactivity completing in a total of 48 hours using 4 cores on a laboratory Linux-based server. Specification curves using maximum likelihood models (lme4 in R; Bates & Bolker, 2011) were even more efficient, with thousands of models running within minutes using a 2019 MacBook Pro (2.8 GHz Intel Core i7).

Limitations

The present study is subject to several limitations that may be addressed in future investigations. Perhaps most crucially, our conclusions (along with those of many developmental fMRI studies) are limited by the poor test-retest reliability of the fMRI data. Because amygdala—mPFC measures showed low reliability across study visits, the statistical power of our analyses of age-related changes is likely low (Elliott et al., 2020; Zuo et al., 2019). Low-powered studies can yield increased rates of both false positive and false negative results (as well as errors of the sign and magnitude of estimates; Button et al., 2013; Gelman & Carlin, 2014); therefore we caution against interpretation of our developmental findings (and brain-behavior associations) beyond the cohort studied in the present investigation. In particular, the low statistical power of our rapid event-related task design may be a major contributor to the low test-retest reliability and variance in outcomes across analysis specifications. That being said, achieving high-powered studies presents a challenge for

studying populations that cannot tolerate lengthy fMRI sessions. Both findings that were more robust to analytical decisions (amygdala reactivity) and findings that were less so (amygdala—mPFC connectivity, associations with separation anxiety) may be most valuable in meta-analytic contexts where greater aggregate statistical power can be achieved. In particular, future work on amygdala—mPFC development will benefit from optimization of measures both on robustness to analytic variability (Li et al., 2021) and reliability (Kragel et al., 2021).

Present findings are also limited by the number of participants studied (Bossier et al., 2020; Marek et al., 2020), the number of longitudinal study sessions per participant (King et al., 2018), and the duration of the task (Nee, 2019). Work with larger sample sizes, more study sessions per participant, and more task data collected per session will be necessary for charting functional amygdala–mPFC development and examining heterogeneity across individuals (although collecting task-based fMRI will continue to be challenging for studies including younger children). The generalizability of the current findings may also be limited by the fact that this cohort was skewed towards high incomes and not racially or ethnically representative of the Los Angeles or United States population.

Findings are also somewhat limited by the fact that the present study is not wholly confirmatory, despite preregistration. Because our multiverse analysis approaches expanded significantly beyond the methods we preregistered, most of the present analyses, while hypothesis-driven, must be considered exploratory (Flournoy et al., 2020). The fact that some specifications used data included in previous similar analyses of the same cohort(Gee et al., 2013) also limits the confirmatory power of the present study (Kriegeskorte et al., 2009). This may be especially true because longitudinal models could not identify within-person change as distinct from between-participant differences (see Figure 2D), indicating that our age-related change estimates may be influenced by cross-sectional information similar to that investigated by Gee et al. (Gee et al., 2013).

Though the current study aimed to estimate longitudinal age-related changes in amygdala–mPFC functional circuitry evoked by fear and neutral faces, the current findings may not be specific to these stimuli (Hariri et al., 2002). Because our task did not include non-face foils or probe specific emotion-related processes, results may be driven by attention, learning, or visual processing, rather than affective or face processing. In particular, because participants were instructed to press a button for neutral faces and withhold a button press for fear faces, observed amygdala—mPFC responses may in part reflect response inhibition (for fear faces; Menon et al., 2001) and target detection processes (for neutral faces; Jonkman et al., 2003). Findings for the fear faces > baseline and neutral > baseline contrasts also may not be valence-specific in the absence of a different emotional face as part of the contrast. Further, because all faces were adult White women, the current results may not generalize to faces more broadly (Richeson et al., 2008; Telzer et al., 2012). Additionally, because face stimuli were the same across study visits, exposure effects across sessions may confound longitudinal findings (although exposure effects may be possible any time a task is repeated, even if stimuli are unique), particularly age-related decreases in amygdala responses (Telzer et al., 2018). While within-session amygdala habituation effects have been shown across several paradigms (Geissberger et al., 2020; Hare et al., 2008; Hein et al., 2018), between-session habituation effects are unlikely beyond 2-3 weeks (Geissberger et al., 2020; Johnstone et al., 2005; Plichta et al., 2014; Spohrs et al., 2018).

Finally, our findings on age-related change in amygdala and mPFC function may be biased or confounded by age-related differences in head motion (Ciric et al., 2017), anatomical image quality and alignment (Gilmore et al., 2020; Rorden et al., 2012), signal dropout, and physiological artifacts (Boubela et al., 2015; Fair et al., 2020; Gratton et al., 2020). While our multiverse analyses included preprocessing and group-level modeling specifications designed

to minimize some of such potential issues, future work is still needed to optimize discrimination of developmental changes of interest from such potential confounds.

Despite these limitations, the present study concords with prior investigations in demonstrating the value of multiverse approaches to quantify sensitivity to researcher decisions. The results highlight key analytic considerations for future studies of age-related changes in amygdala—mPFC function, as well as for studies of human brain development more broadly.

CRediT Author Statement:

Paul Alexander Bloom: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft, Visualization **Michelle VanTieghem:** Methodology, Writing – Review & Editing **Laurel Gabard-Durnam:** Investigation, Methodology, Writing – Review & Editing, **Dylan Gee:** Investigation, Methodology, Writing – Review & Editing **Jessica Flannery:** Investigation**,** Writing – Review & Editing **Christina Caldera:** Investigation**,** Writing – Review & Editing **Bonnie Goff:** Investigation**,** Writing – Review & Editing **Eva Telzer:** Investigation**,** Writing – Review & Editing **Kathryn L. Humphreys:** Investigation**,** Writing – Review & Editing **Dominic Fareri:** Investigation**,** Writing – Review & Editing **Mor Shapiro:** Investigation**,** Writing – Review & Editing **Sameah Algharazi:** Validation**,** Writing – Review & Editing **Niall Bolger:** Methodology, Formal Analysis, Writing – Review & Editing **Mariam Aly:** Methodology, Formal Analysis, Supervision, Writing – Review & Editing **Nim Tottenham:** Conceptualization, Methodology, Investigation, Resources, Data Curation, Writing - Original Draft, Supervision, Funding Acquisition

Code Availability

Code for preprocessing, analysis, and data visualizations for this manuscript is available at https://github.com/pab2163/amygdala_mpfc_multiverse. While unfortunately this code cannot be run as written without data, we have attempted to document analysis steps clearly. In addition, we provide publicly available simulated data structured similarly to the study data on amygdala reactivity, such that interested readers can view multiverse analysis walkthroughs (https://pab2163.github.io/amygdala_mpfc_multiverse) and experiment with analysis code. Additional materials, including MNI space masks and preregistration documentation, are available at https://osf.io/hvdmx/

881 **References**

882

883 Abdulrahman, H., & Henson, R. N. (2016). Effect of trial-to-trial variability on optimal event-
884 related fMRI design: Implications for Beta-series correlation and multi-voxel pattern
885 analysis. *NeuroImage*, *125*, 756–766. https://doi.org/10.1016/j.neuroimage.2015.11.009
886 Achenbach, T. M. (1991). *Integrative Guide for the 1991 CBCL/4-18, Ysr, and Trf Profiles* (1st
887 US-1st Printing edition). Univ Vermont/Dept Psychiatry.
888 Achterberg, M., & van der Meulen, M. (2019). Genetic and environmental influences on MRI
889 scan quantity and quality. *Developmental Cognitive Neuroscience*, *38*, 100667.
890 https://doi.org/10.1016/j.dcn.2019.100667
891 Adolphs, R. (2008). Fear, Faces, and the Human Amygdala. *Current Opinion in Neurobiology*,
892 *18*(2), 166–172. https://doi.org/10.1016/j.conb.2008.06.006
893 Bates, D., & Bolker, M. M. and B. (2011). *lme4: Linear mixed-effects models using S4 classes*
894 (0.999375-39) [Computer software].
895 http://www.idg.pl/mirrors/CRAN/web/packages/lme4/
896 Beeley, C. (2013). *Web Application Development with R using Shiny*. Packt Publishing Ltd.
897 Belloy, M. E., Naeyaert, M., Abbas, A., Shah, D., Vanreusel, V., van Audekerke, J., Keilholz, S.
898 D., Keliris, G. A., Van der Linden, A., & Verhoye, M. (2018). Dynamic resting state fMRI
899 analysis in mice reveals a set of Quasi-Periodic Patterns and illustrates their relationship
900 with the global signal. *NeuroImage*, *180*(Pt B), 463–484.
901 https://doi.org/10.1016/j.neuroimage.2018.01.075
902 Beretta, L., & Santaniello, A. (2016). Nearest Neighbor Imputation Algorithms: A Critical
903 Evaluation. *BMC Medical Informatics and Decision Making*, *16*(Suppl 3).
904 https://doi.org/10.1186/s12911-016-0318-z

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(10), 1230–1236. https://doi.org/10.1097/00004583-199910000-00011

Bossier, H., Roels, S. P., Seurinck, R., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Quinlan, E. B., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Artiges, E., Nees, F., Orfanos, D. P., Poustka, L., … Moerkerke, B. (2020). The empirical replicability of task-based fMRI as a function of sample size. *NeuroImage*, *212*, 116601. https://doi.org/10.1016/j.neuroimage.2020.116601

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Boubela, R. N., Kalcher, K., Huf, W., Seidel, E.-M., Derntl, B., Pezawas, L., Našel, C., & Moser, E. (2015). FMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. *Scientific Reports*, *5*(1), 10499. https://doi.org/10.1038/srep10499

Bowring, A., Maumet, C., & Nichols, T. E. (2019). Exploring the impact of analysis software on task fMRI results. *Human Brain Mapping*, *40*(11), 3362–3384. https://doi.org/10.1002/hbm.24603

Bridgeford, E. W., Wang, S., Yang, Z., Wang, Z., Xu, T., Craddock, C., Dey, J., Kiar, G., Gray-Roncal, W., Priebe, C. E., Caffo, B., Milham, M., Zuo, X.-N., Reproduciblity, C. for R.

and, & Vogelstein, J. T. (2020). Big Data Reproducibility: Applications in Brain Imaging. *BioRxiv*, 802629. https://doi.org/10.1101/802629

Bryce, N. V., Flournoy, J. C., Guassi Moreira, J. F., Rosen, M. L., Sambook, K. A., Mair, P., & McLaughlin, K. A. (2021). Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *NeuroImage*, *243*, 118487. https://doi.org/10.1016/j.neuroimage.2021.118487

Bürkner, P.-C. (2019). *brms: Bayesian Regression Models using "Stan"* (2.10.0) [Computer software]. https://CRAN.R-project.org/package=brms

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Carpenter, K. L. H., Angold, A., Chen, N.-K., Copeland, W. E., Gaur, P., Pelphrey, K., Song, A. W., & Egger, H. L. (2015). Preschool Anxiety Disorders Predict Different Patterns of Amygdala-Prefrontal Connectivity at School-Age. *PLOS ONE*, *10*(1), e0116854. https://doi.org/10.1371/journal.pone.0116854

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of Symptoms of DSM-IV Anxiety and Depression in Children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy*, *38*(8), 835–855. https://doi.org/10.1016/S0005-7967(99)00130-8

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–137.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Cisler, J. M., Bush, K., & Steele, J. S. (2014). A Comparison of Statistical Methods for Detecting Context-Modulated Functional Connectivity in fMRI. *NeuroImage*, *84*, 1042–1052. https://doi.org/10.1016/j.neuroimage.2013.09.018

Cosme, D., & Lopez, R. (2020). *Neural indicators of food cue reactivity, regulation, and valuation and their associations with body composition and daily eating behavior*. PsyArXiv. https://doi.org/10.31234/osf.io/23mu5

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.

Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S. S., Yan, C.-G., Li, Q., Lurie, D., Vogelstein, J., Burns, R., Colcombe, S., Mennes, M., Kelly, C., Di Martino, A., Castellanos, F. X., & Milham, M. (2013). *Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC)*. Neuroinformatics, Stockholm, Sweden. https://doi.org/10.3389/conf.fninf.2013.09.00042

Crone, E. A., & Elzinga, B. M. (2015). Changing brains: How longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. *WIREs Cognitive Science*, *6*(1), 53–63. https://doi.org/10.1002/wcs.1327

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development*, *11*(2), 121–136. https://doi.org/10.1080/15248371003699969

Dafflon, J., Costa, P. F. D., Váša, F., Monti, R. P., Bzdok, D., Hellyer, P. J., Turkheimer, F., Smallwood, J., Jones, E., & Leech, R. (2020). *Neuroimaging: Into the Multiverse* (p. 2020.10.29.359778). https://doi.org/10.1101/2020.10.29.359778

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The Contribution of Emotion and Cognition to Moral Sensitivity: A Neurodevelopmental Study. *Cerebral Cortex*, *22*(1), 209–220. https://doi.org/10.1093/cercor/bhr111

Del Giudice, M., & Gangestad, S. W. (2021). A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920954925. https://doi.org/10.1177/2515245920954925

Di, X., & Biswal, B. B. (2017). Psychophysiological Interactions in a Visual Checkerboard Task: Reproducibility, Reliability, and the Effects of Deconvolution. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00573

Di, X., Reynolds, R. C., & Biswal, B. B. (2017). Imperfect (de)Convolution May Introduce Spurious Psychophysiological Interactions and How to Avoid It. *Human Brain Mapping*, *38*(4), 1723–1740. https://doi.org/10.1002/hbm.23413

Di, X., Zhang, Z., & Biswal, B. B. (2020). Understanding psychophysiological interaction and its relations to beta series correlation. *BioRxiv*, 322073. https://doi.org/10.1101/322073

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis: *Psychological Science*. https://doi.org/10.1177/0956797620916786

Fair, D. A., Miranda-Dominguez, O., Snyder, A. Z., Perrone, A., Earl, E. A., Van, A. N., Koller, J. M., Feczko, E., Tisdall, M. D., van der Kouwe, A., Klein, R. L., Mirro, A. E., Hampton, J. M., Adeyemo, B., Laumann, T. O., Gratton, C., Greene, D. J., Schlaggar, B. L., Hagler, D. J., … Dosenbach, N. U. F. (2020). Correction of respiratory artifacts in MRI head motion estimates. *NeuroImage*, *208*, 116400. https://doi.org/10.1016/j.neuroimage.2019.116400

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Flournoy, J. C., Vijayakumar, N., Cheng, T. W., Cosme, D., Flannery, J. E., & Pfeifer, J. H. (2020). Improving practices and inferences in developmental cognitive neuroscience. *Developmental Cognitive Neuroscience*, *45*, 100807. https://doi.org/10.1016/j.dcn.2020.100807

Forbes, E. E., Phillips, M. L., Silk, J. S., Ryan, N. D., & Dahl, R. E. (2011). Neural Systems of Threat Processing in Adolescents: Role of Pubertal Maturation and Relation to Measures of Negative Affect. *Developmental Neuropsychology*, *36*(4), 429–452. https://doi.org/10.1080/87565641.2010.550178

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009). The Global Signal and Observed Anticorrelated Resting State Brain Networks. *Journal of Neurophysiology*, *101*(6), 3270–3283. https://doi.org/10.1152/jn.90777.2008

Gee, D. G., Humphreys, K. L., Flannery, J., Goff, B., Telzer, E. H., Shapiro, M., Hare, T. A., Bookheimer, S. Y., & Tottenham, N. (2013). A Developmental Shift from Positive to Negative Connectivity in Human Amygdala–Prefrontal Circuitry. *Journal of Neuroscience*, *33*(10), 4584–4593. https://doi.org/10.1523/JNEUROSCI.3446-12.2013

Geissberger, N., Tik, M., Sladky, R., Woletz, M., Schuler, A.-L., Willinger, D., & Windischberger, C. (2020). Reproducibility of amygdala activation in facial emotion processing at 7T. *NeuroImage*, *211*, 116585. https://doi.org/10.1016/j.neuroimage.2020.116585

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization. *Journal of Educational and Behavioral Statistics*, *40*(5), 530–543. https://doi.org/10.3102/1076998615606113

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis--a 'garden of forking paths'--explains why many statistically significant comparisons don't hold up. *American Scientist*, *102*(6), 460–466.

Gilmore, A., Buser, N., & Hanson, J. L. (2020). Variations in Structural MRI Quality Significantly Impact Commonly-Used Measures of Brain Anatomy. *BioRxiv*, 581876. https://doi.org/10.1101/581876

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *NeuroImage*, *19*(1), 200–207. https://doi.org/10.1016/S1053-8119(03)00058-2

Glasser, M. F., Coalson, T. S., Bijsterbosch, J. D., Harrison, S. J., Harms, M. P., Anticevic, A., Van Essen, D. C., & Smith, S. M. (2018). Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *NeuroImage*, *181*, 692–717. https://doi.org/10.1016/j.neuroimage.2018.04.076

Glenn, N. D. (2003). Distinguishing Age, Period, and Cohort Effects. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the Life Course* (pp. 465–476). Springer US. https://doi.org/10.1007/978-0-306-48247-2_21

Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M., & Kennedy, K. M. (2020). Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. *Human Brain Mapping*. https://doi.org/10.1002/hbm.25217

Gratton, C., Dworetsky, A., Coalson, R. S., Adeyemo, B., Laumann, T. O., Wig, G. S., Kong, T. S., Gratton, G., Fabiani, M., Barch, D. M., Tranel, D., Miranda-Dominguez, O., Fair, D. A., Dosenbach, N. U. F., Snyder, A. Z., Perlmutter, J. S., Petersen, S. E., & Campbell, M. C. (2020). Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *BioRxiv*, 837161. https://doi.org/10.1101/837161

Guyer, A. E., Monk, C. S., McClure-Tone, E. B., Nelson, E. E., Roberson-Nay, R., Adler, A. D., Fromm, S. J., Leibenluft, E., Pine, D. S., & Ernst, M. (2008). A Developmental Examination of Amygdala Response to Facial Expressions. *Journal of Cognitive Neuroscience*, *20*(9), 1565–1582. https://doi.org/10.1162/jocn.2008.20114

Haller, S. P., Chen, G., Kitt, E. R., Smith, A. R., Stoddard, J., Abend, R., Cardenas, S. I., Revzina, O., Coppersmith, D., Leibenluft, E., Brotman, M. A., Pine, D. S., & Pagliaccio, D. (2022). Reliability of task-evoked neural activation during face-emotion paradigms: Effects of scanner and psychological processes. *Human Brain Mapping*, *n/a*(n/a). https://doi.org/10.1002/hbm.25723

Hare, T. A., Tottenham, N., Galvan, A., Voss, H. U., Glover, G. H., & Casey, B. J. (2008). Biological Substrates of Emotional Reactivity and Regulation in Adolescence During an Emotional Go-Nogo Task. *Biological Psychiatry*, *63*(10), 927–934. https://doi.org/10.1016/j.biopsych.2008.03.015

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The Amygdala Response to Emotional Stimuli: A Comparison of Faces and Scenes. *NeuroImage*, *17*(1), 317–323. https://doi.org/10.1006/nimg.2002.1179

Heeringa, S. G., Wagner, J., Torres, M., Duan, N., Adams, T., & Berglund, P. (2004). Sample designs and sampling methods for the Collaborative Psychiatric Epidemiology Studies (CPES). *International Journal of Methods in Psychiatric Research*, *13*(4), 221–240. https://doi.org/10.1002/mpr.179

Hein, T. C., Mattson, W. I., Dotterer, H. L., Mitchell, C., Lopez-Duran, N., Thomason, M. E., Peltier, S. J., Welsh, R. C., Hyde, L. W., & Monk, C. S. (2018). Amygdala habituation and uncinate fasciculus connectivity in adolescence: A multi-modal approach. *NeuroImage*, *183*, 617–626. https://doi.org/10.1016/j.neuroimage.2018.08.058

Herting, M. M., Gautam, P., Chen, Z., Mezher, A., & Vetter, N. C. (2017). Test-Retest Reliability of Longitudinal Task-Based fMRI: Implications for Developmental Studies. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2017.07.001

Infantolino, Z. P., Luking, K. R., Sauder, C. L., Curtin, J. J., & Hajcak, G. (2018). Robust Is Not Necessarily Reliable: From within-Subjects fMRI Contrasts to between-Subjects Comparisons. *NeuroImage*, *173*, 146–152. https://doi.org/10.1016/j.neuroimage.2018.02.024

Jalbrzikowski, M., Larsen, B., Hallquist, M. N., Foran, W., Calabro, F., & Luna, B. (2017). Development of White Matter Microstructure and Intrinsic Functional Connectivity Between the Amygdala and Ventromedial Prefrontal Cortex: Associations With Anxiety and Depression. *Biological Psychiatry*, *82*(7), 511–521. https://doi.org/10.1016/j.biopsych.2017.01.008

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jo, H. J., Lee, J.-M., Kim, J.-H., Shin, Y.-W., Kim, I.-Y., Kwon, J. S., & Kim, S. I. (2007). Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques. *NeuroImage*, *34*(2), 550–564. https://doi.org/10.1016/j.neuroimage.2006.09.047

Johnstone, T., Somerville, L. H., Alexander, A. L., Oakes, T. R., Davidson, R. J., Kalin, N. H., & Whalen, P. J. (2005). Stability of amygdala BOLD response to fearful faces over multiple scan sessions. *NeuroImage*, *25*(4), 1112–1123. https://doi.org/10.1016/j.neuroimage.2004.12.016

Jonkman, L. M., Lansbergen, M., & Stauder, J. E. A. (2003). Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. *Psychophysiology*, *40*(5), 752–761. https://doi.org/10.1111/1469-8986.00075

Joseph, J. E., Zhu, X., Gundran, A., Davies, F., Clark, J. D., Ruble, L., Glaser, P., & Bhatt, R. S. (2015). Typical and Atypical Neurodevelopment for Face Specialization: An fMRI Study. *Journal of Autism and Developmental Disorders*, *45*(6), 1725–1741. https://doi.org/10.1007/s10803-014-2330-4

Kennedy, J. T., Harms, M. P., Korucuoglu, O., Astafiev, S. V., Barch, D. M., Thompson, W. K., Bjork, J. M., & Anokhin, A. P. (2021). *Reliability and Stability Challenges in ABCD Task fMRI Data* (p. 2021.10.08.463750). https://doi.org/10.1101/2021.10.08.463750

Killgore, W. D. S., Oki, M., & Yurgelun-Todd, D. A. (2001). Sex-specific developmental changes in amygdala responses to affective faces. *NeuroReport*, *12*(2), 427.

Killgore, W. D. S., & Yurgelun-Todd, D. A. (2007). Unconscious Processing of Facial Affect in Children and Adolescents. *Social Neuroscience*, *2*(1), 28–47. https://doi.org/10.1080/17470910701214186

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. *Developmental Cognitive Neuroscience*, *33*, 54–72. https://doi.org/10.1016/j.dcn.2017.11.009

Klapwijk, E., van den Bos, W., Tamnes, C. K., Mills, K. L., & Raschle, N. (2019). *Opportunities for Increased Reproducibility and Replicability of Developmental Cognitive Neuroscience* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fxjzt

Kragel, P. A., Han, X., Kraynak, T. E., Gianaros, P. J., & Wager, T. D. (2021). Functional MRI Can Be Highly Reliable, but It Depends on What You Measure: A Commentary on Elliott et al. (2020). *Psychological Science*, *32*(4), 622–626. https://doi.org/10.1177/0956797621989730

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Kujawa, A., Wu, M., Klumpp, H., Pine, D. S., Swain, J. E., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. (2016). Altered Development of Amygdala-Anterior Cingulate Cortex Connectivity in Anxious Youth and Young Adults. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(4), 345–352. https://doi.org/10.1016/j.bpsc.2016.01.006

Li, X., Ai, L., Giavasis, S., Jin, H., Feczko, E., Xu, T., Clucas, J., Franco, A., Heinsfeld, A. S., Adebimpe, A., Vogelstein, J. T., Yan, C.-G., Esteban, O., Poldrack, R. A., Craddock, C., Fair, D., Satterthwaite, T., Kiar, G., & Milham, M. P. (2021). *Moving Beyond Processing and Analysis-Related Variation in Neuroscience* (p. 2021.12.01.470790). bioRxiv. https://doi.org/10.1101/2021.12.01.470790

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science*, *18*(5), 421–428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of Cognitive Processes From Late Childhood to Adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

Luna, B., Tervo-Clemmens, B., & Calabro, F. J. (2021). Considerations When Characterizing Adolescent Neurocognitive Development. *Biological Psychiatry*, *89*(2), 96–98. https://doi.org/10.1016/j.biopsych.2020.04.026

Mackey, S., & Petrides, M. (2014). Architecture and Morphology of the Human Ventromedial Prefrontal Cortex. *The European Journal of Neuroscience*, *40*(5), 2777–2796. https://doi.org/10.1111/ejn.12654

Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2017). Current Methods and Limitations for Longitudinal fMRI Analysis across Development. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2017.11.006

Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*, *33*, 118–128. https://doi.org/10.1016/j.dcn.2017.11.006

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., Moore, L. A., Conan, G., Uriarte, J., … Dosenbach, N. U. F. (2020). Towards Reproducible Brain-Wide Association Studies. *BioRxiv*, 2020.08.21.257758. https://doi.org/10.1101/2020.08.21.257758

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, *61*(4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068

Menon, V., Adleman, N. E., White, C. d., Glover, G. h., & Reiss, A. l. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, *12*(3), 131–143. https://doi.org/10.1002/1097-0193(200103)12:3<131::AID-HBM1010>3.0.CO;2-C

Meyer, H. C., & Lee, F. S. (2019). Translating Developmental Neuroscience to Understand Risk for Psychiatric Disorders. *American Journal of Psychiatry*, *176*(3), 179–185. https://doi.org/10.1176/appi.ajp.2019.19010091

Moriceau, S., & Sullivan, R. M. (2006). Maternal presence serves as a switch between learning fear and attraction in infancy. *Nature Neuroscience; New York*, *9*(8), 1004–1006. http://dx.doi.org.ezproxy.cul.columbia.edu/10.1038/nn1733

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nee, D. E. (2019). FMRI replicability depends upon sufficient individual-level data. *Communications Biology*, *2*(1), 1–4. https://doi.org/10.1038/s42003-019-0378-6

Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of Putative fMRI Biomarkers during Emotional Face Processing. *NeuroImage*, *156*, 119–127. https://doi.org/10.1016/j.neuroimage.2017.05.024

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the Trade: Psychophysiological Interactions and Functional Connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604–609. https://doi.org/10.1093/scan/nss055

Passarotti, A. M., Sweeney, J. A., & Pavuluri, M. N. (2009). Neural Correlates of Incidental and Directed Facial Emotion Processing in Adolescents and Adults. *Social Cognitive and Affective Neuroscience*, *4*(4), 387–398. https://doi.org/10.1093/scan/nsp029

Pattwell, S. S., Liston, C., Jing, D., Ninan, I., Yang, R. R., Witztum, J., Murdock, M. H., Dincheva, I., Bath, K. G., Casey, B. J., Deisseroth, K., & Lee, F. S. (2016). Dynamic Changes in Neural Circuitry during Adolescence Are Associated with Persistent Attenuation of Fear Memories. *Nature Communications*, *7*, 11475. https://doi.org/10.1038/ncomms11475

Perlman, S. B., & Pelphrey, K. A. (2011). Developing Connections for Affective Regulation: Age-Related Changes in Emotional Brain Connectivity. *Journal of Experimental Child Psychology*, *108*(3), 607–620. https://doi.org/10.1016/j.jecp.2010.08.006

Pernet, C. R. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: A tutorial for junior neuro-imagers. *Frontiers in Neuroscience*, *8*, 1. https://doi.org/10.3389/fnins.2014.00001

Pfeifer, J. H., Masten, C. L., Moore, W. E., Oswald, T. M., Mazziotta, J. C., Iacoboni, M., & Dapretto, M. (2011). Entering Adolescence: Resistance to Peer Influence, Risky Behavior, and Neural Changes in Emotion Reactivity. *Neuron*, *69*(5). https://doi.org/10.1016/j.neuron.2011.02.019

Pine, D. S., Grun, J., Zarahn, E., Fyer, A., Koda, V., Li, W., Szeszko, P. R., Ardekani, B., & Bilder, R. M. (2001). Cortical Brain Regions Engaged by Masked Emotional Faces in Adolescents and Adults: An fMRI Study. *Emotion*, *1*(2), 137–147. https://doi.org/10.1037/1528-3542.1.2.137

Plichta, M. M., Grimm, O., Morgen, K., Mier, D., Sauer, C., Haddad, L., Tost, H., Esslinger, C., Kirsch, P., Schwarz, A. J., & Meyer-Lindenberg, A. (2014). Amygdala habituation: A reliable fMRI phenotype. *NeuroImage*, *103*, 383–390. https://doi.org/10.1016/j.neuroimage.2014.09.059

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Lynch, C. J., Silver, B. M., Dubin, M. J., Martin, A., & Jones, R. M. (2019). Distinctions among real and apparent respiratory motions in human fMRI data. *NeuroImage*, *201*, 116041. https://doi.org/10.1016/j.neuroimage.2019.116041

Pozzi, E., Vijayakumar, N., Rakesh, D., & Whittle, S. (2020). Neural correlates of emotion regulation in adolescents and emerging adults: A meta-analytic study. *Biological Psychiatry*, *0*(0). https://doi.org/10.1016/j.biopsych.2020.08.006

Qin, S., Young, C. B., Duan, X., Chen, T., Supekar, K., & Menon, V. (2014). Amygdala Subregional Structure and Intrinsic Functional Connectivity Predicts Individual Differences in Anxiety During Early Childhood. *Biological Psychiatry*, *75*(11), 892–900. https://doi.org/10.1016/j.biopsych.2013.10.006

Richeson, J. A., Todd, A. R., Trawalter, S., & Baird, A. A. (2008). Eye-Gaze Direction Modulates Race-Related Amygdala Activity. *Group Processes & Intergroup Relations*, *11*(2), 233–246. https://doi.org/10.1177/1368430207088040

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*, *60*(1), 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., Gennatas, E. D., Jackson, C., Prabhakaran, K., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. E., & Gur, R. C. (2013). Heterogeneous Impact of Motion on Fundamental Patterns of Developmental Changes in Functional Connectivity during Youth. *NeuroImage*, *83*, 45–57. https://doi.org/10.1016/j.neuroimage.2013.06.045

Sauder, C. L., Hajcak, G., Angstadt, M., & Phan, K. L. (2013). Test-retest reliability of amygdala response to emotional faces. *Psychophysiology*, *50*(11), 1147–1156. https://doi.org/10.1111/psyp.12129

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *32*(4), 811–830. https://doi.org/10.1016/j.neubiorev.2007.12.002

Silvers, J. A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R. E., Weber, J., Mischel, W., Casey, B. J., & Ochsner, K. N. (2017a). VlPFC–vmPFC–Amygdala Interactions Underlie Age-Related Differences in Cognitive Regulation of Emotion. *Cerebral Cortex (New York, NY)*, *27*(7), 3502–3514. https://doi.org/10.1093/cercor/bhw073

Silvers, J. A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R., Weber, J., Mischel, W., Casey, B. J., & Ochsner, K. N. (2017b). The Transition from Childhood to Adolescence Is Marked by a General Decrease in Amygdala Reactivity and an Affect-Specific Ventral-to-Dorsal Shift in Medial Prefrontal Recruitment. *Developmental Cognitive Neuroscience*, *25*, 128–137. https://doi.org/10.1016/j.dcn.2016.06.005

Silvers, J. A., Shu, J., Hubbard, A. D., Weber, J., & Ochsner, K. N. (2015). Concurrent and Lasting Effects of Emotion Regulation on Amygdala Response in Adolescence and Young Adulthood. *Developmental Science*, *18*(5), 771–784. https://doi.org/10.1111/desc.12260

Simonsohn, U., Simmons, J., & Nelson, L. (2015). Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. *Marketing Papers*. https://doi.org/10.2139/ssrn.2694998

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Spohrs, J., Bosch, J. E., Dommes, L., Beschoner, P., Stingl, J. C., Geiser, F., Schneider, K., Breitfeld, J., & Viviani, R. (2018). Repeated fMRI in measuring the activation of the

amygdala without habituation when viewing faces displaying negative emotions. *PLOS ONE*, *13*(6), e0198244. https://doi.org/10.1371/journal.pone.0198244

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Sullivan, R. M., & Perry, R. E. (2015). Mechanisms and Functional Implications of Social Buffering in Infants: Lessons from Animal Models. *Social Neuroscience*, *10*(5), 500–511. https://doi.org/10.1080/17470919.2015.1087425

Swartz, J. R., Carrasco, M., Wiggins, J. L., Thomason, M. E., & Monk, C. S. (2014). Age-Related Changes in the Structure and Function of Prefrontal Cortex–Amygdala Circuitry in Children and Adolescents: A Multi-Modal Imaging Approach. *NeuroImage*, *86*, 212–220. https://doi.org/10.1016/j.neuroimage.2013.08.018

Telzer, E. H., Flannery, J., Humphreys, K. L., Goff, B., Gabard-Durman, L., Gee, D. G., & Tottenham, N. (2015). “The Cooties Effect”: Amygdala Reactivity to Opposite-versus Same-Sex Faces Declines from Childhood to Adolescence. *Journal of Cognitive Neuroscience*, *27*(9), 1685–1696. https://doi.org/10.1162/jocn_a_00813

Telzer, E. H., Humphreys, K. L., Shapiro, M., & Tottenham, N. (2012). Amygdala Sensitivity to Race Is Not Present in Childhood but Emerges over Adolescence. *Journal of Cognitive Neuroscience*, *25*(2), 234–244. https://doi.org/10.1162/jocn_a_00311

Telzer, E. H., McCormick, E. M., Peters, S., Cosme, D., Pfeifer, J. H., & van Duijvenvoorde, A. C. K. (2018). Methodological Considerations for Developmental Longitudinal fMRI Research. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2018.02.004

Todd, R. M., Evans, J. W., Morris, D., Lewis, M. D., & Taylor, M. J. (2011). The Changing Face of Emotion: Age-Related Patterns of Amygdala Activation to Salient Faces. *Social Cognitive and Affective Neuroscience*, *6*(1), 12–23. https://doi.org/10.1093/scan/nsq007

US Census Bureau. (2021). *U.S. Census Bureau QuickFacts: Los Angeles County, California*. https://www.census.gov/quickfacts/losangelescountycalifornia

VanTieghem, M., Korom, M., Flannery, J., Choy, T., Caldera, C., Humphreys, K. L., Gabard-Durnam, L., Goff, B., Gee, D. G., Telzer, E. H., Shapiro, M., Louie, J. Y., Fareri, D. S., Bolger, N., & Tottenham, N. (2021). Longitudinal changes in amygdala, hippocampus and cortisol development following early caregiving adversity. *Developmental Cognitive Neuroscience*, *48*, 100916. https://doi.org/10.1016/j.dcn.2021.100916

Vijayakumar, N., Pfeifer, J. H., Flournoy, J. C., Hernandez, L. M., & Dapretto, M. (2019). Affective reactivity during adolescence: Associations with age, puberty and testosterone. *Cortex*, *117*, 336–350. https://doi.org/10.1016/j.cortex.2019.04.024

Vink, M., Derks, J. M., Hoogendam, J. M., Hillegers, M., & Kahn, R. S. (2014). Functional Differences in Emotion Processing during Adolescence and Early Adulthood. *NeuroImage*, *91*, 70–76. https://doi.org/10.1016/j.neuroimage.2014.01.035

Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: A general framework using a genetic algorithm. *NeuroImage*, *18*(2), 293–309.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wu, M., Kujawa, A., Lu, L. H., Fitzgerald, D. A., Klumpp, H., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. (2016). Age-related changes in amygdala–frontal connectivity during emotional face processing from childhood into young adulthood. *Human Brain Mapping*, *37*(5), 1684–1695. https://doi.org/10.1002/hbm.23129

Xu, J., Hao, L., Chen, M., He, Y., Jiang, M., Tian, T., Wang, H., Wang, Y., Wang, D., Han, Z. R., Tan, S., Men, W., Gao, J., He, Y., Tao, S., Dong, Q., & Qin, S. (2021). Developmental Sex Differences in Negative Emotion Decision-Making Dynamics: Computational Evidence and Amygdala-Prefrontal Pathways. *Cerebral Cortex*, bhab359. https://doi.org/10.1093/cercor/bhab359

Xu, T., Cho, J. W., Kiar, G., Bridgeford, E. W., Vogelstein, J. T., & Milham, M. P. (2022). *A Guide for Quantifying and Optimizing Measurement Reliability for the Study of Individual Differences* (p. 2022.01.27.478100). bioRxiv. https://doi.org/10.1101/2022.01.27.478100

Yousefi, B., Shin, J., Schumacher, E. H., & Keilholz, S. D. (2018). Quasi-periodic patterns of intrinsic brain activity in individuals and their relationship to global signal. *NeuroImage*, *167*, 297–308. https://doi.org/10.1016/j.neuroimage.2017.11.043

Yurgelun-Todd, D. A., & Killgore, W. D. S. (2006). Fear-related activity in the prefrontal cortex increases with age during adolescence: A preliminary fMRI study. *Neuroscience Letters*, *406*(3), 194–199. https://doi.org/10.1016/j.neulet.2006.07.046

Zhang, Y., Padmanabhan, A., Gross, J. J., & Menon, V. (2019). Development of human emotion circuits investigated using a Big-Data analytic approach: Stability, reliability, and robustness. *Journal of Neuroscience*, 0220–19. https://doi.org/10.1523/JNEUROSCI.0220-19.2019

Zuo, X.-N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour*, *3*(8), 768–771. https://doi.org/10.1038/s41562-019-0655-x

1369
1370 Table 1: Summary of main aims, hypotheses, methods, and findings

Aim	Preregistered Hypothesis	Analysis Methodology	Key Findings
1a. Age-related change in amygdala reactivity to fear faces	Amygdala reactivity to fearful faces will decrease with age, such that younger children will have more positive amygdala reactivity (higher BOLD response to fear faces relative to implicit baseline) than older youth.	Multiverse amygdala ROI (anatomically-defined) analysis using multilevel linear regression at the group level. *Multiverse decision points:* Preprocessing software, GLM software, GLM nuisance regressors, amygdala ROI definition, contrast estimate type (t-stat vs. beta estimate), HRF shape, group-level model covariates, exclusion of previously analyzed scans	• Across decision points, weak but consistent negative age-related change in amygdala reactivity to fear > baseline and neutral > baseline contrasts • No consistent evidence for age-related change in fear > neutral contrast • Longitudinal models could identify consistent between-participant differences but not within-participant age-related change
1b. Age-related change in patterns of amygdala responses across task trials	None	Multiverse analysis of slopes of amygdala reactivity across trials, and amygdala reactivity in each half of trials using multilevel linear regression at the group level, single trial models *Multiverse decision points:* Global signal subtraction, amygdala ROI definition, group-level model covariates	• On average, amygdala reactivity decreased across trials (for both fear and neutral faces) • Amygdala reactivity for earlier trials was higher at younger ages • Age-related change in amygdala reactivity to fear faces in the first half of trials, but not the second half • Similar, but somewhat weaker age-related change for neutral faces
2a. Age-related change in amygdala–mPFC functional connectivity (FC) to fear faces, as measured by generalized psychophysiological interaction (gPPI)	Amygdala–mPFC FC will decrease as a function of age such that as age increases, the valence of FC will shift from positive to negative.	Multiverse gPPI analysis with anatomically defined bilateral amygdala seed and mPFC target ROIs using multilevel linear regression at the group level. *Multiverse decision points:* Deconvolution step, mPFC ROI definition, contrast estimate type (t-stat vs. beta estimate), group-level model covariates	• No consistent evidence for age-related change in gPPI for any contrast • gPPI estimates extremely sensitive to deconvolution step in creation of regressors
2b. Age-related change in amygdala–mPFC functional connectivity to fear faces, as measured by beta-series correlation (BSC)	None for BSC specifically	Multiverse BSC analysis between amygdala and mPFC using multilevel linear regression at the group level. *Multiverse decision points:* Global signal subtraction, amygdala ROI definition, mPFC ROI definition, group-level model covariates	• No consistent evidence for age-related change in BSC for any condition • Amygdala–mPFC BSC was most sensitive to selection of mPFC ROI • Global signal subtraction reduced average amygdala–mPFC BSC, but impacts on age-related changes were small • BSC estimates were not strongly associated with gPPI estimates
3. Associations of amygdala reactivity, change in amygdala reactivity across trials, or amygdala–mPFC FC with separation anxiety	None	Multiverse multilevel linear regressions with brain measures as predictors for separation anxiety behaviors, controlling for age *Multiverse decision points:* Separation anxiety measure, FC measure, mPFC ROI (FC only), amygdala ROI, contrast, deconvolution step (gPPI only)	• No evidence that amygdala reactivity, amygdala–mPFC connectivity, or change in amygdala reactivity across trials were associated with separation anxiety behaviors

1371
1372
1373

1374 Table 2: Summary of forking pipelines used in analyses for each aim[1]

Aim/Analysis	Decision Point	Choices
1a. Age-related change in amygdala reactivity to fear faces > baseline	Preprocessing Software	**FSL FEAT**, C-PAC
	GLM Software	**FSL FEAT**, AFNI 3dDeconvolve
	Hemodynamic Response Function	**Double Gamma**, Single Gamma
	Nuisance Regressors	**24 motion regressors**, 6 motion regressors, 18 motion regressors + WM + CSF
	Low-frequency artifact removal	**High-pass filter (.01Hz)**, Quadratic drift regressor
	First-level GLM Estimates	**Beta Estimates**, T-statistics
	Native vs. Standard MNI Space	**Native Space (Freesurfer)**, Harvard-Oxford Atlas in MNI
	Amygdala ROI	**Bilateral**, Left, Right, High Signal, Low Signal
	Inclusion of 45 previously analyzed scans	**Include**, Exclude
	Outlier treatment	**Exclude +/-3SD from mean**, Exclude +/-3SD from mean + robust regression
	Group-level model covariates	**Mean FD**, Mean FD + run, Mean FD + scanner, Mean FD + run + scanner
	Group-level model quadratic term	Yes, **No**
	Group-level model random slopes	**Yes**, No
1b. Age-related change in patterns of amygdala responses across task trials *FSL preproc & GLM, high-pass filter, 24 motion regressors, 2G HRF, beta estimates, included previously analyzed scans, and robust group-level regression*	Method of quantifying within-scan change	Slopes across trials, trials split into halves, single-trial models
	Global Signal Subtraction	Yes, No
	Amygdala ROI (all MNI space)	Bilateral, Left, Right
	Group-level model covariates	Mean FD**,** Mean FD + run, Mean FD + scanner, Mean FD + run + scanner
	Group-level model quadratic term	Yes, No
	Group-level model random slopes	Yes**,** No
2a Age-related change in amygdala–mPFC functional connectivity (FC) to fear faces > baseline, as measured by (gPPI) *FSL preproc & GLM, high-pass filter, 24 motion regressors, 2G HRF, bilateral amygdala ROI in MNI space*	Deconvolution step	Yes, **No**
	mPFC ROI (all MNI space)	**3 different 5mm spheres**, large vmPFC mask
	Outlier treatment	**Exclude +/-3SD from mean**, Exclude +/-3SD from mean + robust regression
	Inclusion of 45 previously analyzed scans	**Include**, Exclude
	Group-level model covariates	**Mean FD**, Mean FD + run, Mean FD + scanner, Mean FD + run + scanner
	Group-level model quadratic term	Yes, **No**
	Group-level model random slopes	**Yes,** No
2b. Age-related change in amygdala–mPFC functional connectivity to fear faces > baseline, as measured by (BSC) *FSL preproc & GLM, high-pass filter, 24 motion regressors, 2G HRF, beta estimates, robust group-level regression, included previously analyzed scans*	Amygdala ROI (all MNI space)	Bilateral, Left, Right
	mPFC ROI (all MNI space)	3 different 5mm spheres, large vmPFC mask
	Global Signal Subtraction	Yes, No
	Group-level model covariates	Mean FD, Mean FD + run, Mean FD + scanner, Mean FD + run + scanner
	Group-level model quadratic term	Yes, No
	Group-level model random slopes	Yes, No
3. Associations of amygdala reactivity, change in amygdala reactivity across trials, or amygdala–mPFC FC with separation anxiety *See supplemental methods p. 29 for details on included pipelines.*	Brain measure	Amygdala reactivity, amygdala reactivity slopes, amygdala–mPFC gPPI, amygdala–mPFC BSC
	Global Signal Subtraction (amygdala reactivity slopes & BSC only)	Yes, No
	Deconvolution step (gPPI only)	Yes, No
	mPFC ROI (all MNI space, gPPI & BSC only)	3 different 5mm spheres, large vmPFC mask
	Separation anxiety outcome variable	RCADS, SCARED raw scores, SCARED t-scores

1375 [1]**Bolded** choices indicate those most closely matching preregistered pipelines.

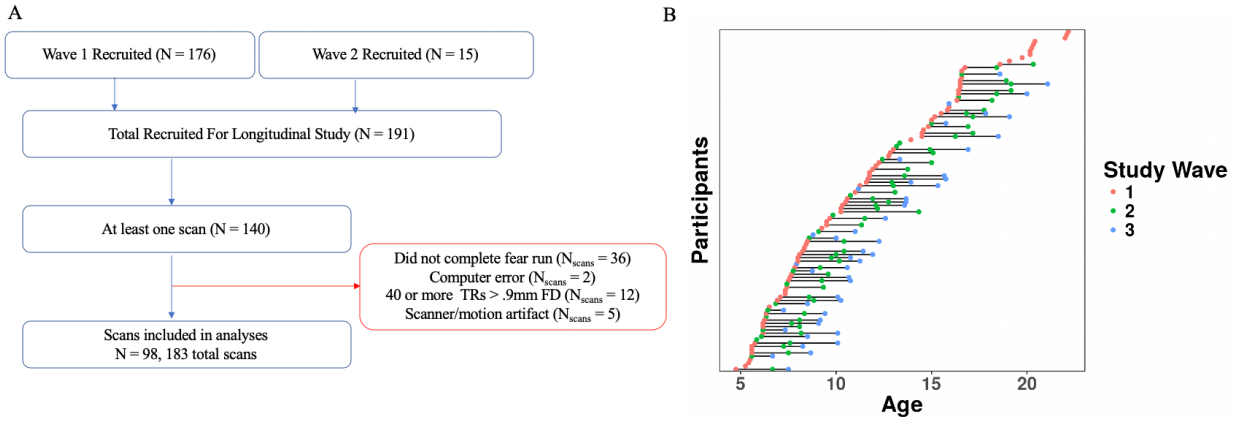

Figure 1. A. Schematic showing study inclusion criteria. **B.** Included scans at each study wave, with each dot representing one scan, and horizontal lines connecting participants across study waves.

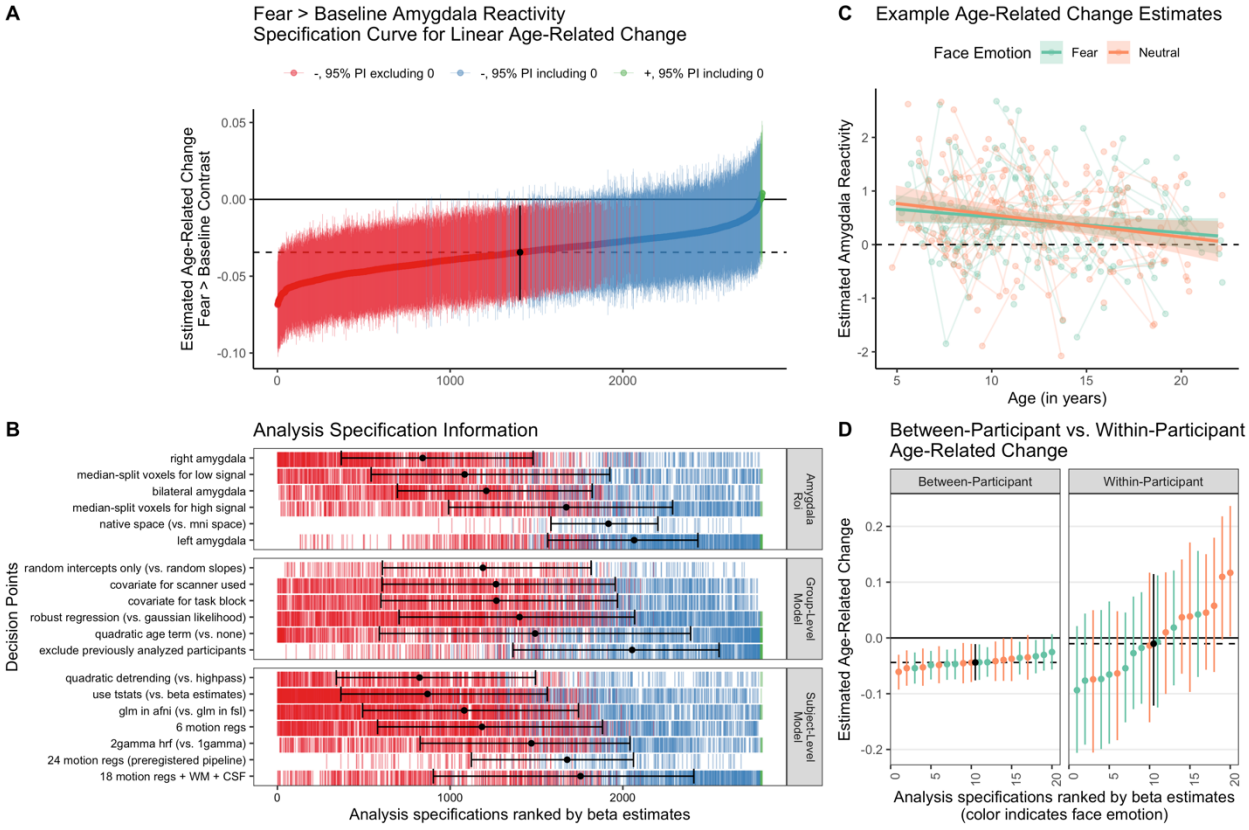

Figure 2. Multiverse analyses of age-related change in amygdala reactivity. **A.** Specification curve of age-related change in fear > baseline amygdala reactivity. Points represent estimated linear age-related change and lines are corresponding 95% posterior intervals (PIs). Models are ordered by age-related change estimates, with the dotted line representing the median estimate across all specifications. Color indicates sign of beta estimates and whether respective posterior intervals include 0 (red = negative excluding 0; blue = negative including 0, green = positive including 0, black = median across all specifications). **B.** Model specification information corresponding to each model in A. Variables on the y-axis represent analysis choices, corresponding color-coded marks indicate that a choice was made, and blank space indicates that the choice was not made in a given analysis. Within each category panel (amygdala ROI, Group-Level Model, and Participant-Level Model), decision points are ordered from top to bottom by the median model rank when the corresponding choice is made (i.e. choices at the top of each panel tend to have more negative age-related change estimates). Black points with error bars represent the median and IQR ranks of specifications making the choice indicated on the corresponding line. **C.** Example participant-level data and model predictions for age-related related change in amygdala reactivity for both the fear > baseline (green) and neutral-baseline (orange) contrasts. Data are shown for a preregistered pipeline using a native space bilateral amygdala mask, 24 motion regressors, t-statistics, high-pass filtering, and participant-level GLMs in FSL. Points represent participant-level estimates, light lines connect estimates from participants with multiple study visits, and dark lines with shaded area represent model predictions and 95% posterior intervals. **D.** Specification curves for a subset of models separately parametrizing within-participant (right) vs. between-participant (left) age-related change for both the fear > baseline (green) and neutral > baseline (orange) contrasts, as well as the median across specifications (black). See https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/ for interactive visualizations.

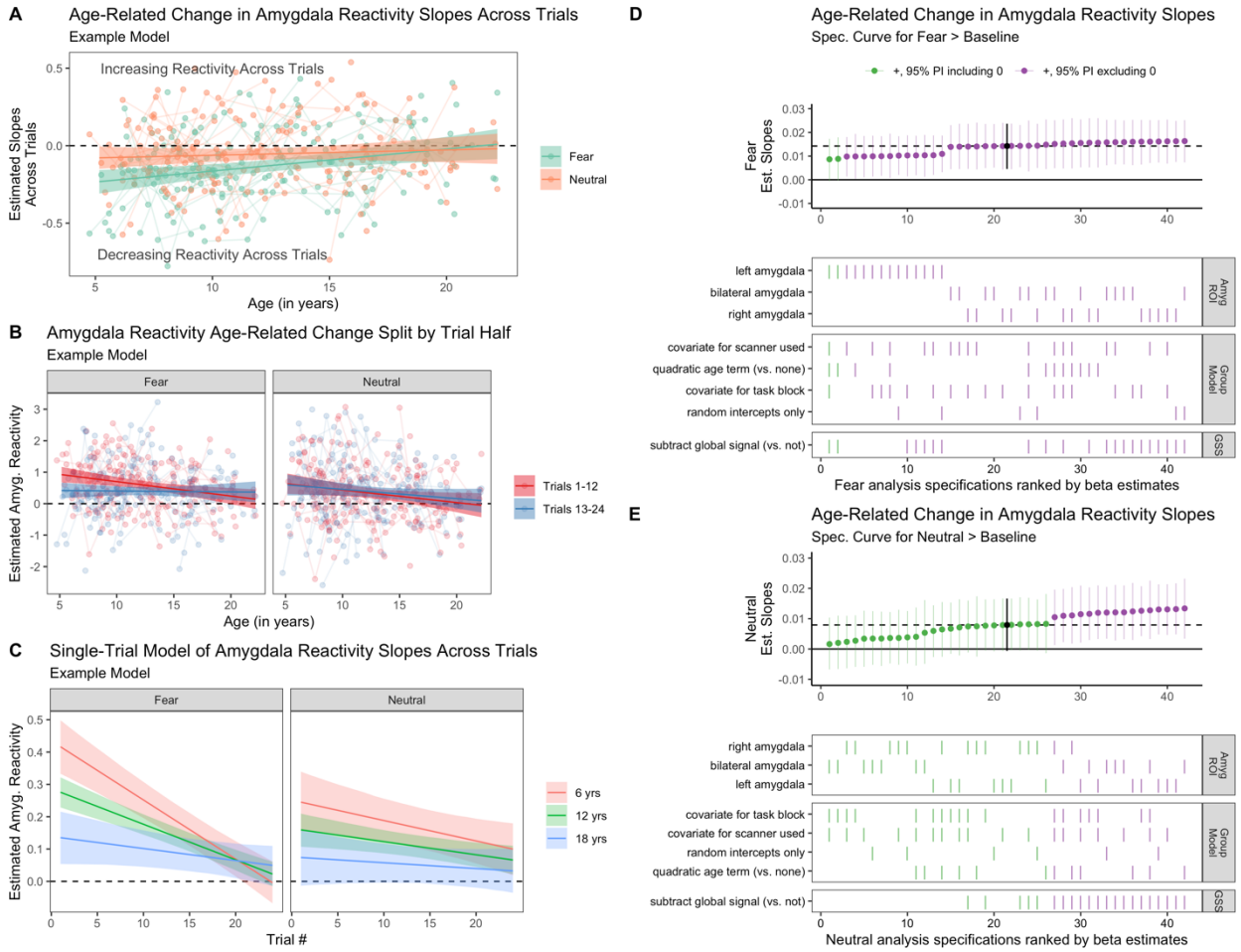

Figure 3. Age-related change in amygdala reactivity across trials. **A.** An example model of estimated age-related change in slopes of beta estimates across both fear (green) and neutral (orange) trials. Negative slopes represent higher amygdala activity in earlier trials relative to later trials. **B.** Example models of estimated age-related change in amygdala reactivity for the fear > baseline (left) and neutral > baseline (right) contrasts for both the first (red) and second (blue) halves of trials. In both A and B, points represent participant-level estimates, light lines connect estimates from participants with multiple study visits, and dark lines with shaded area represent model predictions and 95% posterior intervals. **C.** Example single-trial model predictions of estimated amygdala reactivity for fear (left) and neutral (right) faces as a function of age and trial number. Age was modeled as a continuous variable, and average predictions for participants of age 6 (red), 12 (green) and 18 (blue) years are shown for visualization purposes. All estimates in A-C shown are from an example analysis pipeline using bilateral amygdala estimates and without global signal correction. **D.** Specification curve for age-related change in slopes across fear trials (i.e., many parallel analyses for the fear trials in subplot B). **E.** Specification curve for age-related change in slopes across neutral trials (i.e., neutral trials in plot B). GSS = global signal correction using post-hoc mean centering. For both D and E, color indicates sign of beta estimates and whether respective posterior intervals include 0 (green = positive including 0, purple = positive excluding 0, black = median across all specifications), and horizontal dotted lines represent median estimates across all analysis decisions. Variables on the y-axis represent analysis choices, corresponding color-coded marks indicate that a choice was made, and blank space indicates that the choice was not made in a given analysis.

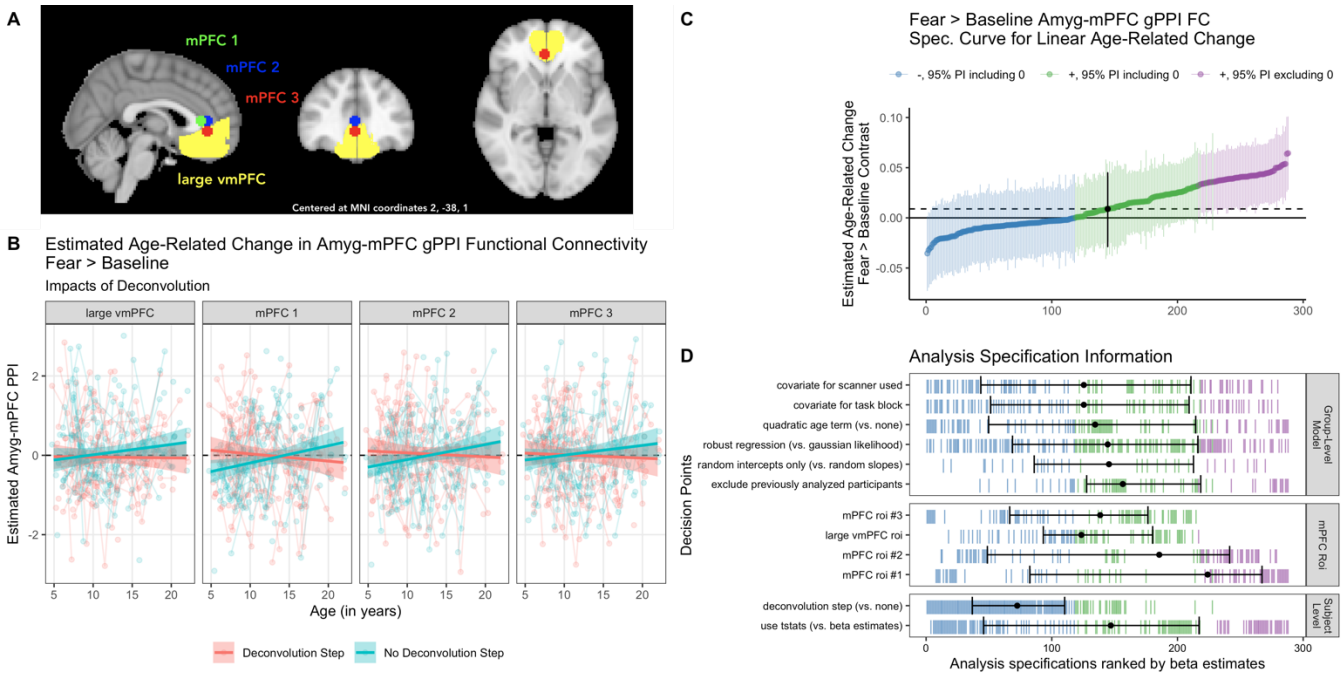

Figure 4. Multiverse analyses of age-related change in amygdala–mPFC connectivity using gPPI methods. **A.** MNI space mPFC ROIs used in connectivity analyses. **B.** Example participant-level data and model predictions for age-related related change in amygdala–mPFC gPPI for analysis pipelines with a deconvolution step (red), or without (blue) for each of the four regions shown in A. Although deconvolution changed the sign of age-related change estimates, the estimates are not 'statistically significant' for each pipeline alone, except for mPFC ROIs 1 & 2 without deconvolution. **C.** Specification curve of age-related change in fear > baseline amygdala–mPFC gPPI. Points represent estimated linear age-related change and lines are corresponding 95% posterior intervals. Models are ordered by age-related change estimates, and the dotted line represents the median estimate across all specifications. Color indicates sign of beta estimates and whether respective posterior intervals include 0 (blue = negative including 0, green = positive including 0, purple = positive excluding 0, black = median across all specifications). Black points with error bars represent the median and IQR ranks of specifications making the choice indicated on the corresponding line. **D.** Model specification information corresponding to each model in C. Variables on the y-axis represent analysis choices, corresponding color-coded marks indicate that a choice was made, and blank space indicates that the choice was not made in a given analysis. Within each category (Group-Level Model, mPFC ROI, and Participant-Level Model) respectively, decision points are ordered from top to bottom by the median model rank when the corresponding choice is made (i.e., choices at the top of each panel tend to have more negative age-related change estimates). See https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/ for interactive visualizations.

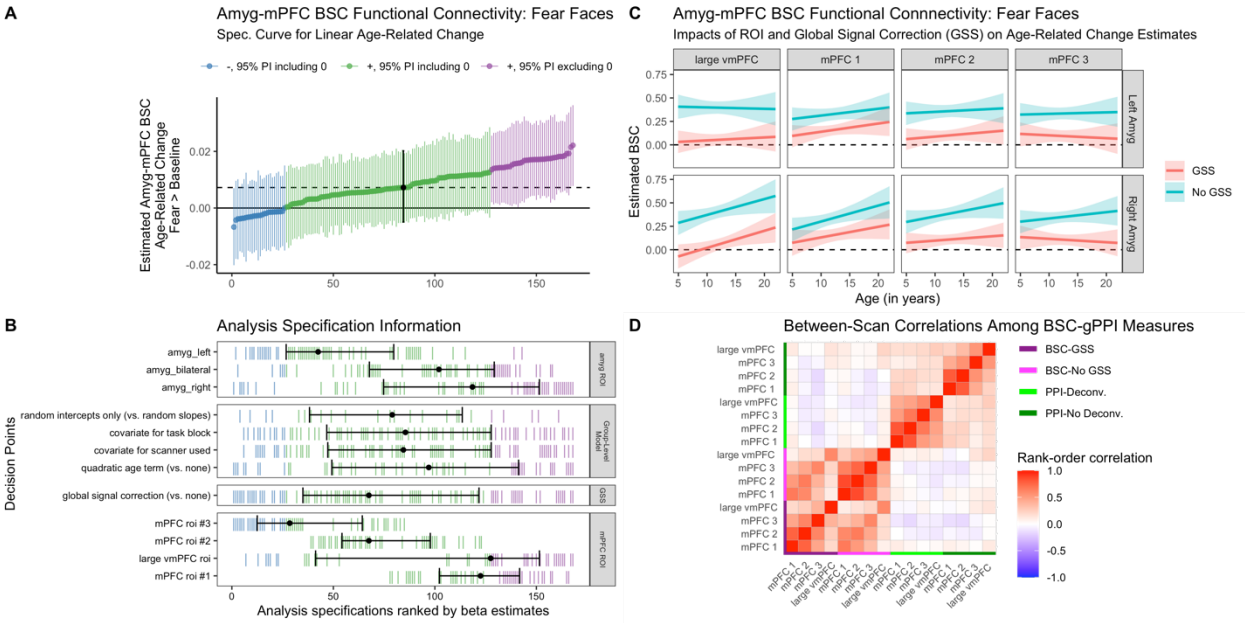

Figure 5. Multiverse analyses of age-related change in amygdala–mPFC connectivity using beta-series correlation (BSC) methods. **A.** Specification curve of age-related change in amygdala–mPFC BSC for fear trials. Points represent estimated linear age-related change and lines are corresponding 95% posterior intervals. Models are ordered by age-related change estimates, and the dotted line represents the median estimate across all specifications. Color indicates sign of beta estimates and whether respective posterior intervals include 0 (blue = negative including 0, green = positive including 0, purple = positive excluding 0, black = median across all specifications). **B.** Model specification information corresponding to each model in A. Variables on the y-axis represent analysis choices, corresponding color-coded marks indicate that a choice was made, and blank space indicates that the choice was not made in a given analysis. Within each category (amygdala ROI, group-level model, global signal subtraction, and mPFC ROI) respectively, decision points are ordered from top to bottom by the median model rank when the corresponding choice is made (i.e., choices at the top of each panel tend to have more negative age-related change estimates). Black points with error bars represent the median and IQR ranks of specifications making the choice indicated on the corresponding line. GSS = global signal correction using post-hoc mean centering. **C.** Example model predictions for age-related change in amygdala–mPFC BSC for fear trials for analysis pipelines with a global signal subtraction (GSS, post-hoc mean centering) step (red), or without (blue) for each of the four mPFC regions (see Figure 4A) with the left and right amygdala. Pipelines shown have random slopes, no covariates for task block or scanner, and no quadratic age term. **D.** Between-scan rank-order correlations between amygdala–mPFC connectivity measures. All gPPI measures are for the fear > baseline contrast, and BSC measures are for fear trials. See https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/ for interactive visualizations.

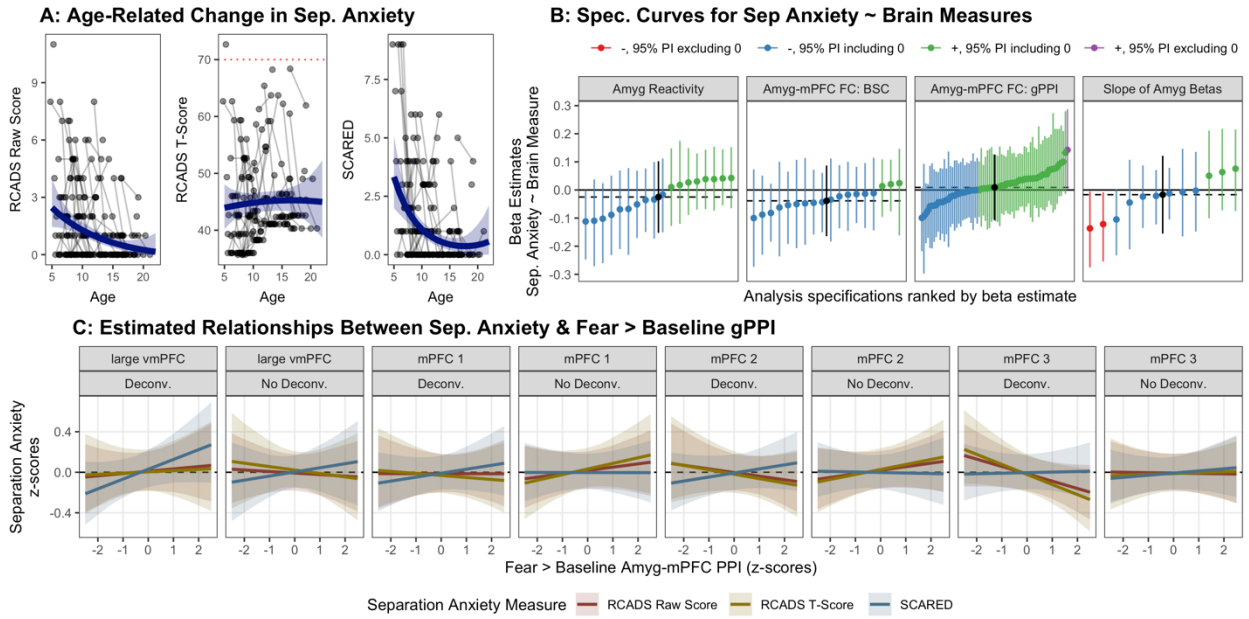

Figure 6. Multiverse analyses of associations between amygdala–mPFC circuitry and separation anxiety. **A.** Age-related change in SCARED and RCADS raw and t-scores for parent-reported separation anxiety subscales. The red dotted line in the middle panel represents the clinical threshold for the standardized RCADS measure (because this T-score measure is standardized based on age and gender, no age-related change is expected). **B.** Separate specification curves for associations of amygdala reactivity (left), amygdala–mPFC connectivity (both gPPI and BSC; center two panels), and amygdala reactivity slopes across trials (right) with the three separation anxiety outcomes shown in A. Points represent estimated associations between brain measures and separation anxiety (controlling for mean framewise displacement and age) and lines are corresponding 95% posterior intervals. Models are ordered by beta estimates, and the dotted line represents the median estimate across all specifications. Color indicates sign of beta estimates and whether respective posterior intervals include 0 (red = negative excluding 0, blue = negative including 0, green = positive including 0). Scores on each separation anxiety outcome were z-scored for comparison. **C.** Example model predictions for associations between fear > baseline amygdala–mPFC gPPI and each separation anxiety measure. Predictions and 95% posterior intervals are plotted for each separation anxiety measure separately for each mPFC region, and for gPPI pipelines with and without a deconvolution step. Pipelines shown use robust regression, have random slopes, no covariates for task block or scanner, and no quadratic age term.

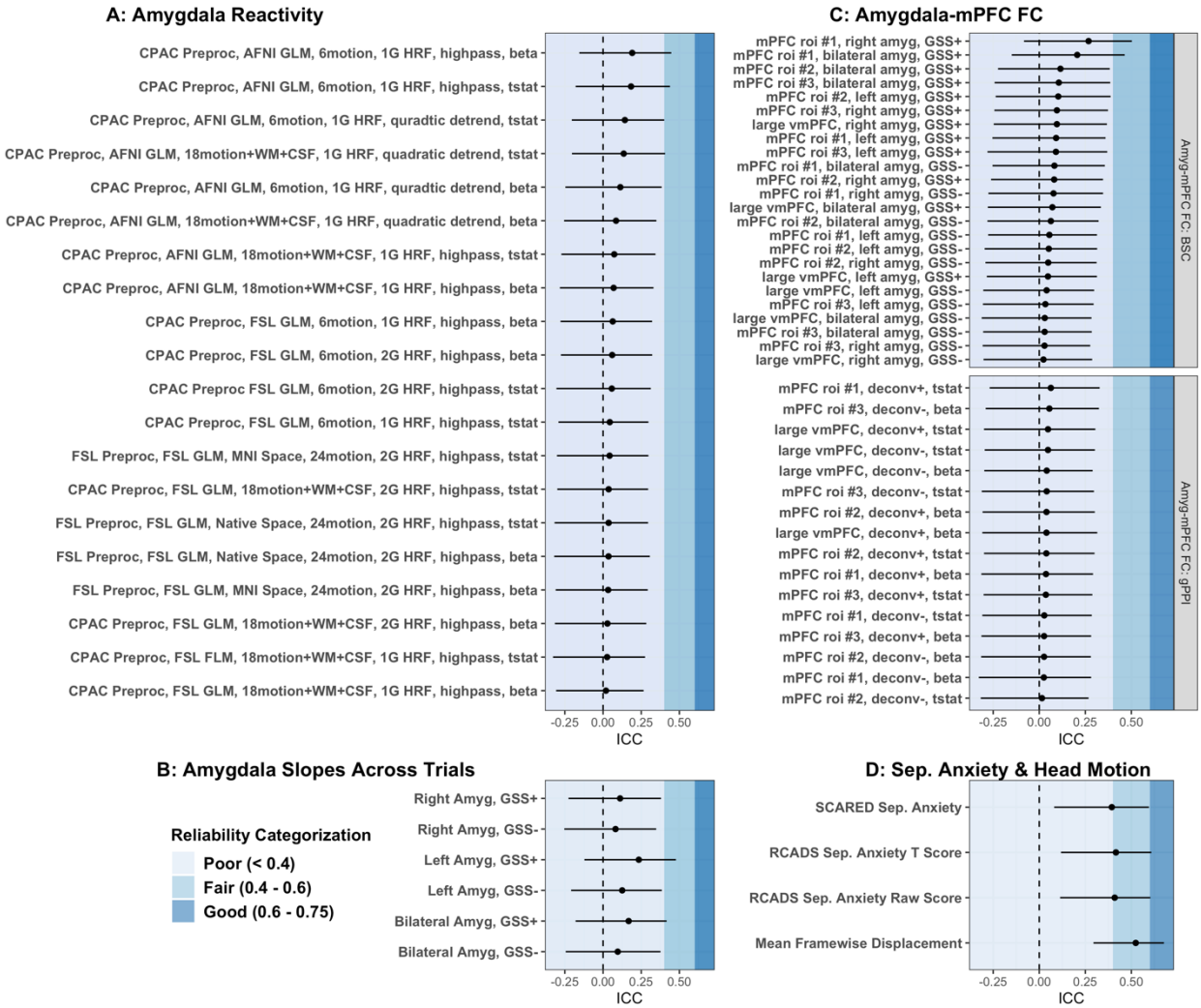

Figure 7. Longitudinal test-retest Bayesian ICC estimates. ICC values are shown for amygdala reactivity (A), slopes of amygdala reactivity betas across trials (B), amygdala—mPFC functional connectivity using both gPPI and BSC methods (C), and separation anxiety and in-scanner head motion measurements (D). Shaded background colors depict whether ICC estimates are categorized as poor (< 0.4), fair (0.4 - 0.6), or good (0.6 – 0.75) reliability. No ICC estimates met the threshold for excellent reliability (>0.75). Bayesian ICC estimates were calculated through a variance decomposition based on posterior predictive distributions. Negative values indicate higher posterior predictive variances not conditioned on random effect terms than conditioned on random effects terms.