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]

Affiliations:
[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

Abstract
There has been considerable interest in the development of the amygdala and its connections with medial prefrontal cortex (mPFC) given the central role of these brain regions in emotional processes. While several studies have suggested that this circuitry exhibits functional changes across the first two decades of life, they have typically employed cross-sectional designs, and findings have been mixed. Additionally, analytic choices may contribute to discrepancies across studies. Here we used an accelerated longitudinal design to examine task-evoked changes in amygdala–mPFC circuitry from 4-22 years of age (N=98; 183 total scans; 1-3 scans per participant). Participants were recruited from the greater Los Angeles area, and completed an event-related emotional face (fear, neutral) task designed to be appropriate for the wide age range. 'Multiverse' analyses examined the robustness of our findings to fMRI analysis choices. 2808 parallel analyses varying in preprocessing and modeling choices found evidence for average age-related decreases in amygdala reactivity to faces. Greater amygdala reactivity at younger ages was attributable to elevated responses during the first few trials relative to later trials. Within-participant changes in amygdala reactivity with age could not be differentiated from between-participant differences, however. Across analysis decision points, we did not find consistent evidence of age-related change in amygdala–mPFC connectivity through generalized psychophysiological interaction (gPPI) or beta-series correlation (BSC) methods. We also did not find evidence for associations between separation anxiety behaviors and amygdala reactivity or amygdala-mPFC connectivity. Within the context of this faces task and age range, age-related changes in amygdala reactivity were more robust to processing pipeline than were task-evoked functional connectivity measures, particularly those using gPPI. 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.

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[1,2], emotion regulation[3–5], and affective learning[6,7]. Characterizing growth trajectories of these regions may provide insight into neural constructions underlying emotional development[8]. To probe amygdala–mPFC circuitry across development, face stimuli are frequently employed because they effectively engage this circuitry while being child-appropriate[9]. 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. Several studies have found age-related change in amygdala reactivity, including decreases as a function of age in response to emotional faces[10–15] (as well as other images)[16–18], increases in amygdala reactivity with age[19,20], or peaks during adolescence[21,22]. Others have observed no age-related changes[23–28]. With task-evoked amygdala–mPFC FC, several studies have found age-related decreases from childhood to young adulthood[10,23,26,29], while others have found increases[16,18,30], or little age-related change[28].

While the small sample sizes examined in many studies on amygdala–mPFC development likely contribute to differences in findings[31], 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[32]), task design (block vs. event-related[33]), or contrast used (faces > shapes vs. faces > baseline) may also contribute to discrepancies. 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[34]. Choices including software package[35], spatial smoothing[36], treatment of head motion[37], parcellation[38], and functional connectivity approach[39] can also impact results and qualitatively change findings[40]. 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[41,42].

Here, we examined 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), and that gPPI would show a shift from positive to negative valence to fear faces during early adolescence (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[10,43], 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 whether results were robust to many analytical decisions.

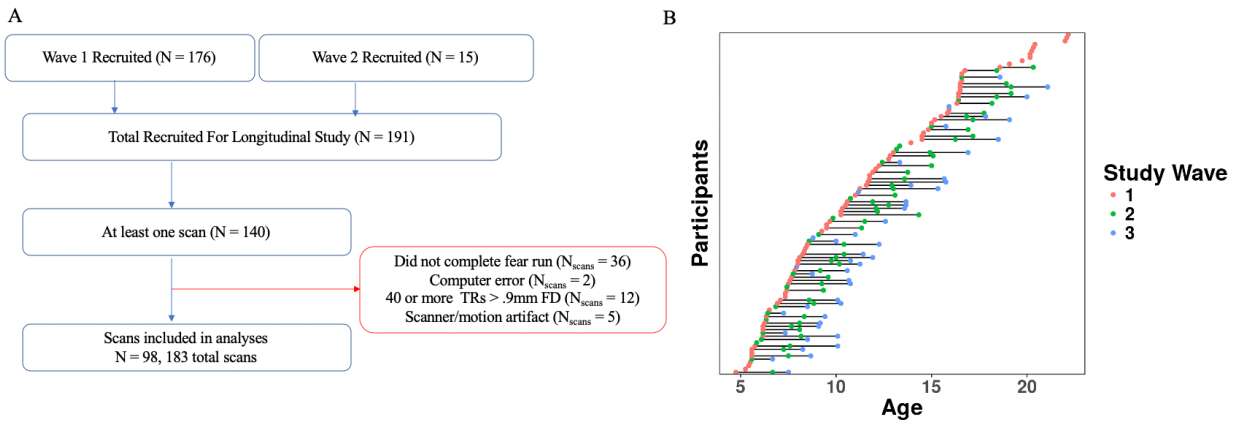

Figure 1. A. Schematic showing study inclusion criteria. **B.** Included scans at each study wave, with each dot representing one scan, and horizontal lines connecting participants across study waves.

Results

Age-related change in amygdala reactivity

We used multilevel regression models and specification curve analyses to examine age-related changes in amygdala reactivity to faces in an accelerated longitudinal sample ranging from ages 4-22 years (Figure 2). Across specifications, we found consistent evidence for negative age-related change in anatomically-defined (Harvard-Oxford atlas and Freesurfer-defined) amygdala reactivity to fear faces > baseline, such that the vast majority of analysis specifications (99.6%) estimated linear slopes at the group level that were negative in sign, and the majority (60.0%) of 95% posterior intervals about these slopes excluded 0 (Figure 2A; interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse). Thus, over half of models indicated that on average, increases in age were associated with decreases in amygdala reactivity to fear faces > baseline. Because the timepoint 1 data in the current study included the 42 scans used by Gee et al. (2013) to age-related changes in amygdala—mPFC circuitry for the fear > baseline contrast, results including these scans may have been more

likely to find similar change (particularly for fear > baseline, see sFigures 11-13, & 25). Estimated age-related change was on average weaker, though still largely negative (98.1% negative, 25.3% of posterior intervals excluding 0) when 42 previously analyzed scans were excluded to provide stricter independence from previously analyzed data (see sFigure 11, Gee et al., 2013).

Parallel multiverse analyses found 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, 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[44,45]. 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 18). We did not find consistent evidence for quadratic age-related changes in amygdala reactivity (see sFigures 14-17).

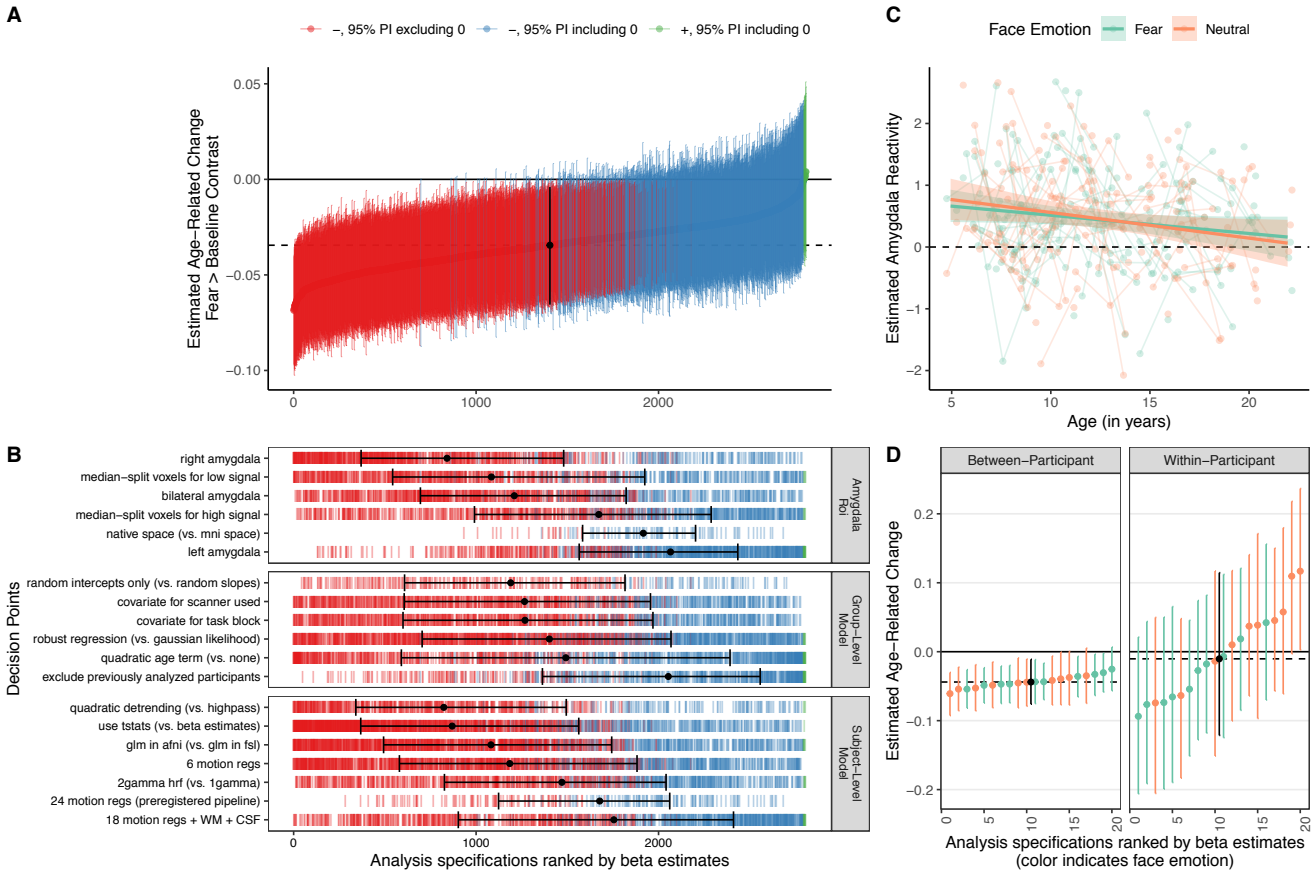

Figure 2. Multiverse analyses of age-related change in amygdala reactivity. **A.** Specification curve of age-related change in fear > baseline amygdala reactivity. Points represent estimated linear age-related change and lines are corresponding 95% posterior intervals (PIs). Models are ordered by age-related 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.

Age-related differences in within-scan amygdala reactivity change

To ask whether age-related changes in amygdala reactivity could be due to developmental changes in patterns of amygdala reactivity across face trials (within a run), we examined whether within-scan change in amygdala reactivity varied with age (see Table 1 Aim 1b). Analyses included 42 specifications (3 amygdala regions of interest [ROIs] x 2 global signal correction options x 7 group-level models). Across both fear and neutral trials, linear slopes of amygdala reactivity were negative on average, indicating higher amygdala reactivity at the beginning of the run (Figure 3A, sFigure 30). Across specifications, for both fear (100% of estimates had the same sign, 95.2% of posterior intervals excluding 0 in the same direction) and neutral trials (100% of estimates in the same direction, 38.1% of posterior intervals excluding 0), there was evidence that these within-scan slopes were steeper (i.e., more negative) at younger ages, though evidence did not meet our rule-of-thumb criteria for 'robustness' (>=50% of 95% posterior intervals excluding 0 in the same direction) for neutral trials (Figure 3D-E). Specifications with a global signal subtraction step also tended to find stronger age-related change.

Similarly, when splitting trials into the first half (trials 1-12) versus second half (trials 13-24), there was consistent evidence (100% of estimates had the same sign, 69.2% with posterior interval excluding 0) for an interaction between age and trial half, such that average reactivity to 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).

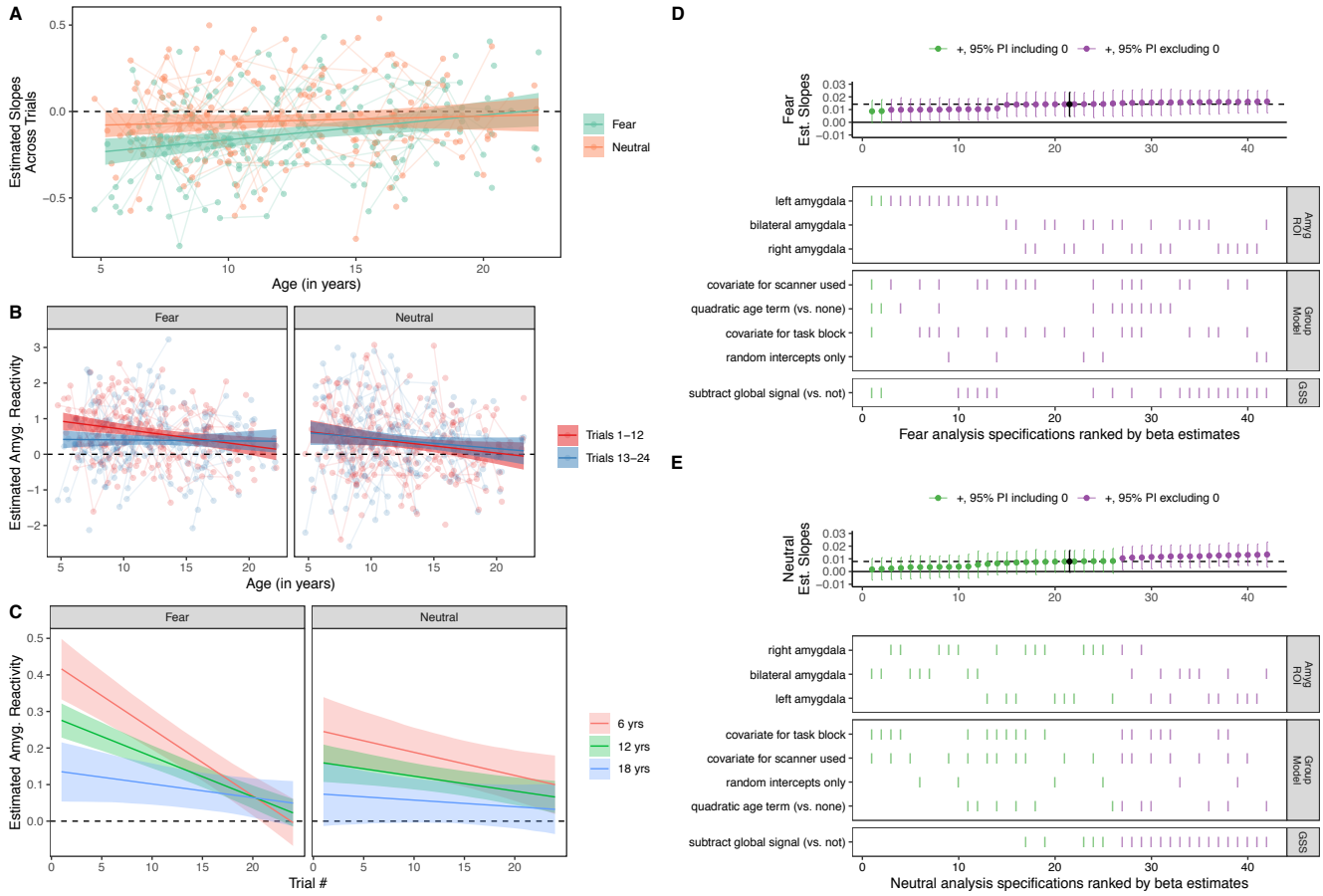

Figure 3. Age-related change in amygdala reactivity across trials. **A.** An example model of estimated age-related change in slopes of beta estimates across both fear (green) and neutral (orange) trials. Negative slopes represent higher amygdala activity in earlier trials relative to later trials. **B.** 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). 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.

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, as 59.0% of models found point estimates in the positive direction, and 23% of posterior intervals excluded 0 (Figure 4C-D, interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/). Through specification curve analyses, we 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 resulted in positive age-related change estimates (96.5% of point estimates positive, 47.9% of posterior intervals excluding 0). A visualization of the effects of the deconvolution step on amygdala FC with each of four mPFC ROIs is presented in Figure 4B. While mPFC ROI definition and other analysis decision points also influenced estimates of age-related change in

gPPI (Figure 4D), follow-up regression models indicated that the effect of including the deconvolution step was several times larger for the fear > baseline contrast (see sFigures 41-43).

Through equivalent multiverse analyses we also found no evidence of consistent linear age-related change in amygdala–mPFC gPPI for the neutral > baseline and fear > neutral contrasts (see sFigures 39-40), or quadratic change for any contrast (see sFigures 44-46). In addition, we did not see consistent evidence for group average amygdala–mPFC gPPI for any contrast, though such results often differed as a function of whether a deconvolution step was included (see sFigure 38). Though we included gPPI analysis specifications excluding the 42 scans at timepoint 1 studied by Gee et al.[10], exclusion of these scans had little impact on age-related change results (see Figure 4D).

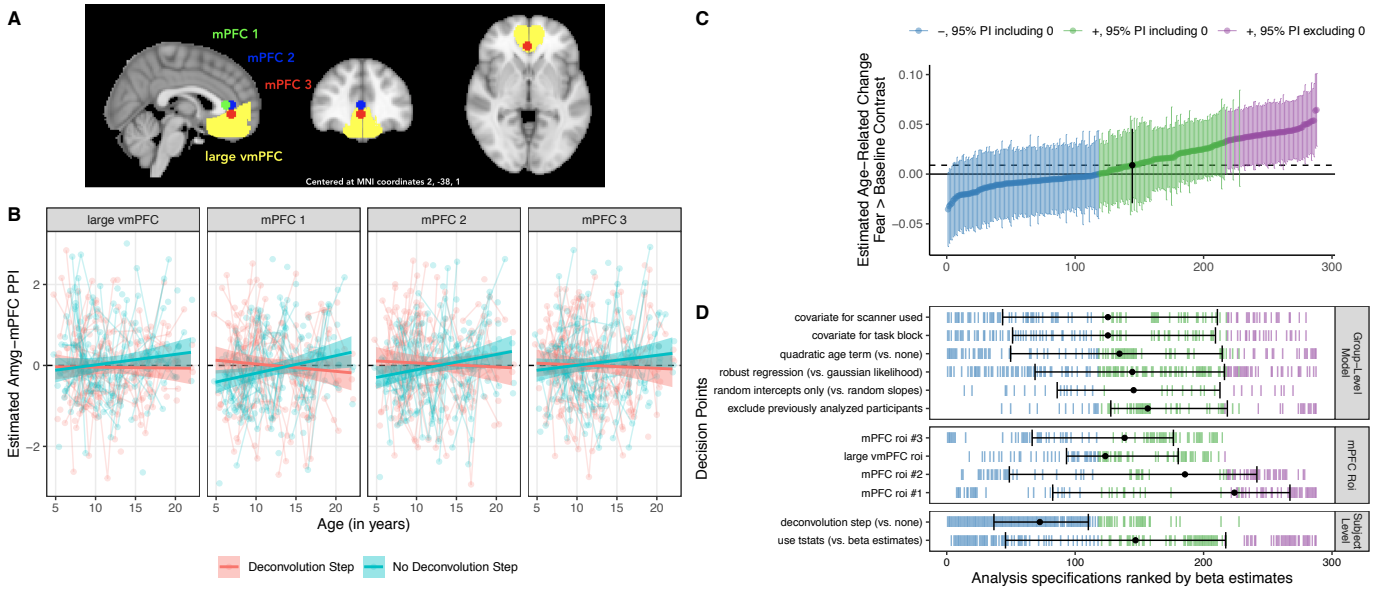

Figure 4. Multiverse analyses of age-related change in amygdala–mPFC connectivity using gPPI methods. **A.** MNI space mPFC ROIs used in connectivity analyses. **B.** Example participant-level data and model predictions for age-related related change in amygdala–mPFC gPPI for analysis pipelines with a deconvolution step (red), or without (blue) for each of the four regions shown in A. **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.

In addition to gPPI analyses, we used beta series correlation (BSC) analyses to examine age-related changes in task-evoked amygdala–mPFC connectivity (see Table 1 Aim 2b). As with gPPI, multiverse analyses of amygdala–mPFC BSC (168 total specifications; 3 amygdala ROI definitions x 4 mPFC ROI definitions x 2 global signal options x 7 group-level models) for fear trials (vs baseline) did not yield consistent evidence of age-related change across pipelines (84.5% of point estimates in the same direction as the median estimate, 24.4% of posterior interval excluding 0; Figure 5A, interactive version at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse). Unlike gPPI analyses, however, choice of mPFC ROI (as well as amygdala ROI, though this was not examined for gPPI) most impacted age-related change in BSC estimates, rather than preprocessing or modeling analytical choices (Figure 5B, sFigures 53-55). Accordingly, while global signal subtraction resulted in weaker amygdala–mPFC BSC on average, inclusion of this step did not consistently affect age-related change estimates (Figure 4C). We did not find consistent evidence for age-related change in amygdala–mPFC BSC for neutral trials (vs baseline), or for fear > neutral trials (sFigures 51-52). We did not find consistent evidence for quadratic age-related change for any contrast (sFigures 56-58).

Additionally, we constructed a correlation matrix using rank-order correlations of scan-level BSC and gPPI estimates for the fear (vs baseline) condition. Across scans, there was little

evidence of correspondence between BSC and gPPI metrics for amygdala–mPFC connectivity (Figure 5D, sFigures 60-63). Further, FC estimates tended to be positively correlated within a method type (BSC, gPPI) across mPFC ROIs, though less strongly for gPPI estimates with versus without a deconvolution step.

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

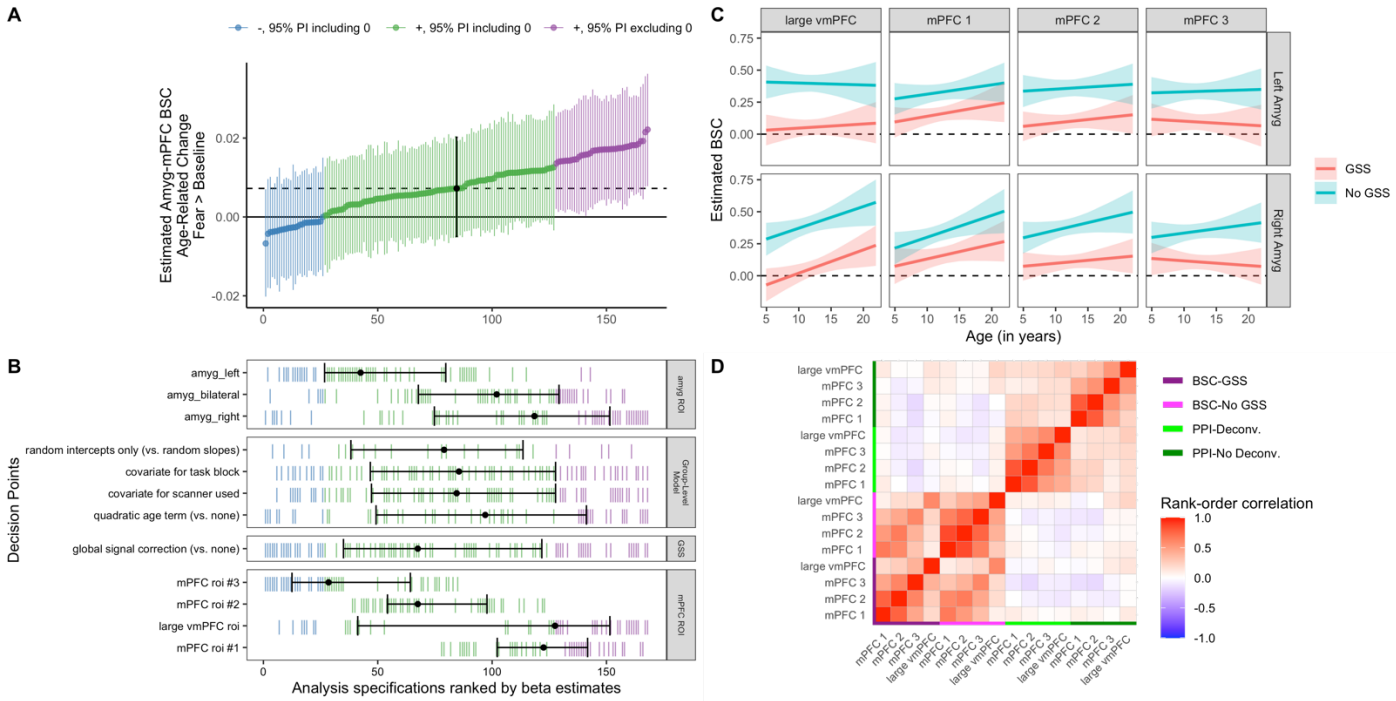

Figure 5. Multiverse analyses of age-related change in amygdala–mPFC connectivity using beta-series correlation (BSC) methods. **A.** Specification curve of age-related change in amygdala–mPFC BSC for fear trials. Points represent estimated linear age-related change and lines are corresponding 95% posterior 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. **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.

Amygdala–mPFC Measures & Separation Anxiety

We conducted multiverse analyses of associations between several amygdala–mPFC measures (amygdala reactivity, amygdala–mPFC FC, within-scan changes in amygdala reactivity) and separation anxiety behaviors (see Table 1 Aim 3). Separation anxiety behaviors on average decreased with age, as indicated by the RCADS-P and SCARED-P raw scores (Figure 6A-C). Neither specification curves for amygdala reactivity (18 total specifications, 56% of point estimates in the same direction as median estimate, 0% of posterior intervals excluding 0), amygdala–mPFC FC (90 total specifications, 51% of point estimates in the same direction as median estimate, 1% of posterior intervals excluding 0), nor slope of amygdala responses across trials (12 total specifications, 75% of point estimates in the same direction as median estimate, 17% of posterior intervals excluding 0), found consistent evidence for associations between brain measures and separation anxiety. All specifications controlled for age (see supplemental methods p.26).

To more specifically follow up on previous work reporting associations between separation anxiety behaviors and amygdala–mPFC gPPI for fear > baseline, we plotted model

predictions for such models from the above multiverse analysis for each of the four mPFC ROIs, across all three separation anxiety outcome measures, and both 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.

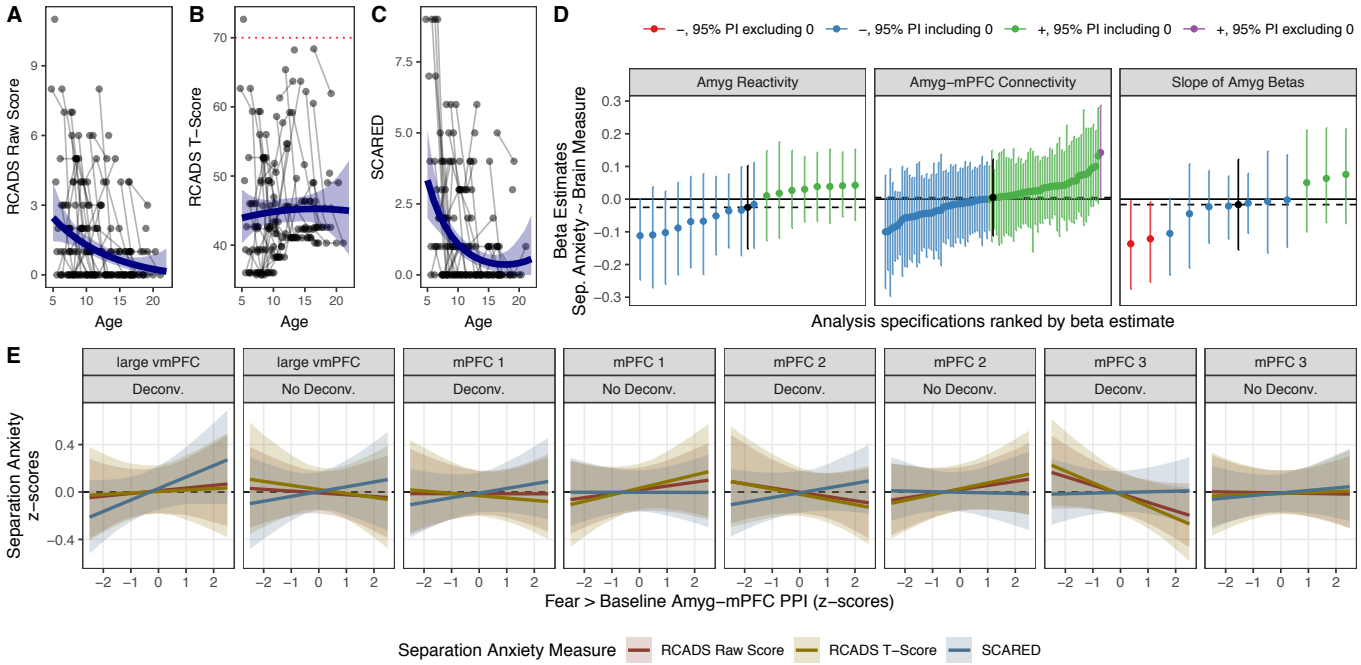

Figure 6. Multiverse analyses of associations between amygdala–mPFC circuitry and separation anxiety. **A-C.** Age-related change in SCARED and RCADS raw and t-scores for parent-reported separation anxiety subscales. The red dotted line in B represents the clinical threshold. **D.** Separate specification curves for associations of amygdala reactivity (left), amygdala–mPFC connectivity (both gPPI and BSC; center), and amygdala reactivity slopes across trials (right) with the three separation anxiety outcomes shown in A-C. Points represent estimated associations between brain measures and separation anxiety (controlling for mean FD 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. **E.** 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.

Discussion

Our accelerated longitudinal design and multiverse analysis approach afforded unique opportunities to examine age-related changes in task-evoked amygdala-prefrontal activity within a cohort recruited from the Los Angeles area. 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). 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). Specification curves also revealed that gPPI analyses were highly sensitive to analysis methods (deconvolution). We did not find associations between any task-related amygdala–mPFC measures (reactivity or functional connectivity) and separation anxiety behaviors (see Table 1 Aim 3).

Amygdala reactivity

Across specifications, we found 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[2,10,11,14], the other 40% of specifications would have been consistent with investigations that found little age-related change[23,26,28].

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[46], where age-

349 related changes may not necessarily indicate 'true' developmental growth. High uncertainty in
350 estimating average within-participant change could be driven by several factors, including true
351 heterogeneity in individual trajectories, low measurement reliability[47–49], scanner differences
352 across longitudinal timepoints, or unmodeled variables impacting amygdala reactivity.
353 Additionally, the within-participants terms represent a smaller age range (a maximum of 4
354 years for any given participant), relative to the broader age range assessed by the between-
355 participants terms (18 years), which may have placed additional limits on identifying reliable
356 within-participant change.

357 Age-related change in amygdala responses to fear faces over baseline seemed largely
358 the result of earlier trials in the task (see sFigures 32-34). While differences in task design and
359 contrast across studies have been highlighted as potential sources of discrepant findings on the
360 development of amygdala function[14,32,50], this result indicates that attention to trial structure and
361 task duration may also be necessary in comparing studies. Because the paradigm used in the
362 current study involved a task requiring participants to press for one face ('neutral') and not press
363 for 'fear' faces, findings specific to fear faces over baseline under the current paradigm may also
364 be driven by behavioral task demands.

365

366 *Amygdala–mPFC Functional Connectivity*

367 Multiverse analyses did not find consistent evidence for age-related change in
368 amygdala–mPFC connectivity using gPPI or BSC (Figures 4-5). Crucially, however,
369 specification curves did not find strong evidence against such age-related change, as we did
370 not observe precise and consistent 'null' estimates across specifications. This absence of
371 evidence for age-related changes in task-evoked amygdala–mPFC connectivity contrasts prior
372 work suggesting a developmental shift in such connectivity[10,23,26], but is consistent with recent
373 work[28].

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 (see Figure 4B). While deconvolution has been argued to be a necessary step for event-related PPI analyses[51], recent work has shifted guidelines on its use, and it may not be recommended for block designs [39,52]. 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 depending on the centering of the task regressors[53]. 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[54] (see supplemental results p.62-66, sFigures 35-37). 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[40].

Amygdala–mPFC circuitry and separation anxiety

Specification curve analyses did not yield consistent evidence of associations between any measures of amygdala–mPFC function (amygdala reactivity & change in amygdala

reactivity across trials, in addition to FC) and separation anxiety behaviors. This finding stands in contrast to associations between amygdala–mPFC connectivity and anxiety identified in previous developmental work[10,23,55,56]. However, given that analyses of brain-behavior associations may require imaging cohorts much larger than the current sample [31,57], the absence of relationships here may not be strong evidence against the existence of potential associations between amygdala–mPFC circuitry and developing anxiety-related behaviors.

Advantages and pitfalls of the multiverse approach

Our findings contribute to a body of work demonstrating that preprocessing and modeling choices can meaningfully influence results[34]. Indeed, most studies involving many analytical decision points could benefit from multiverse analyses. Such specification curves can help to examine the stability of findings in both exploratory and confirmatory research[58]. Particularly when methodological 'gold standards' have not been determined, specification curves may be informative for examining the impacts of potential analysis decisions[59]. Further, wider use of specification curves might help to resolve discrepancies between study findings stemming from different analysis pipelines.

While specification curve analyses may benefit much future research, we also note that multiverses are only as comprehensive as the included specifications[60], and such analyses alone do not solve problems related to unmodeled confounds, design flaws, inadequate statistical power, circular analyses, or non-representative sampling. Further, unless all specifications are decided *a priori*, analyses are vulnerable to problems of analytic flexibility[61], and inclusion of less justified specifications can bias results[62]. Because specification curves can include 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.

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 number of participants studied[31,63], the number of longitudinal study sessions per participant[44], and the duration of the task[64]. In particular, the low statistical power of our rapid event-related task design may be a major contributor to the variance in outcomes across analysis specifications, as 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[65,66]). 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[58]. The fact that some specifications used data included in previous similar analyses of the same cohort[10] also limits the confirmatory power of the present study[67]. 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.[10].

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[9]. 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. 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[68,69].

Finally, our findings on age-related change in amygdala and mPFC function may be biased or confounded by age-related differences in head motion[70], anatomical image quality and alignment[71,72], signal dropout, and physiological artifacts[73–75]. 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 concurs with prior investigations in finding evidence for age-related decreases in amygdala reactivity to fear faces over baseline. Though our work did not find evidence for previously reported developmental ‘shifts’ in positive to negative task-related amygdala–mPFC functional connectivity, our results highlight that both longitudinal study designs and specification curve analyses can be valuable tools for examining the robustness of such analyses, as well as for studies of human brain development more broadly.

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

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 years-old (M = 11.9 years old) who enrolled to participate in a study on emotional development. All participants were reported to be physically and psychiatrically healthy (no medical or psychiatric disorders), as indicated by a telephone screening before participation, and free of MRI contraindications. All except 4 participants fell below clinical cutoffs (see sFigure 2) on the Child Behavior Checklist (CBCL) Total Problems, Internalizing Problems, and Externalizing Problems scales[76]. The larger study also included youths with a history of institutional and/or foster care outside of the United States, who are not included here. Participants from the greater Los Angeles area were recruited through flyers, state birth records, community events, online advertising, lab website and newsletters, psychologists' offices, psychology courses at a local university (participants ages 18-22 years old only), and word-of-mouth. Each participant

completed up to 3 MRI scans spaced at an average interval of 18 months between visits. Parents provided written consent, children 7+ years old gave written assent, and children under 7 years old gave verbal assent. Study protocols were approved by the local university institutional review board. These data were collected between 2009 and 2015.

An accelerated longitudinal design was used such that participants' starting ages at scan 1 comprised the entire range of sample ages (4-22 years old), and coverage was approximately balanced across the entire age range (see Figure 1B). The design was structured into 3 study 'waves' corresponding with recruitment efforts and visit protocols. Participants were overenrolled at wave 1 to account for anticipated attrition (e.g., braces, relocation, etc) to achieve the desired sample size across the three waves. While most participants were recruited such that their first scan occurred at wave 1, a smaller set of participants were recruited at wave 2, such that some participants completed their first scan at wave 2 (see Figure 1). For these 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.[10].

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

522 sample representative of the Los Angeles or United States populations[78], and Hispanic or Latinx
523 participants were particularly underrepresented. Further sample demographics can be found in
524 the supplementary materials (see sTables 1-2, sFigures 1-2).

525 *Separation Anxiety*

526 For each participant (except for 10 adults 18-22 years), a parent completed both the
527 Revised Children's Anxiety and Depression Scale (RCADS-P) and the Screen for Child Anxiety
528 Related Emotional Disorders (SCARED-P) to assess the frequency of symptoms of anxiety and
529 low mood[79,80]. Following prior work suggesting associations between task-evoked amygdala–
530 mPFC functional connectivity and separation anxiety[10,43], we used the separation anxiety
531 subscales from both the SCARED-P and RCADS-P as measures of anxiety-related behaviors in
532 asking whether such functional connectivity may be linked to anxiety levels during childhood
533 and adolescence. For 11 participants who had missing items on the SCARED-P, indicating
534 parents had skipped or forgotten to answer a question, we imputed responses using 5-Nearest
535 Neighbor imputation using only the other items included in the SCARED-P separation anxiety
536 subscale[81]. As expected, raw separation anxiety scores on both measures decreased as a
537 function of age, while standardized scores were consistent across development with few
538 children at or near clinical threshold (see Figure 6, and supplemental results p.31 for intraclass
539 correlation coefficients).
540

541 *Emotion Discrimination Task*

542 Participants completed either two (at wave 3) or three (at waves 1 and 2) runs of a
543 modified 'go/no-go' task with emotional faces during fMRI scanning. Runs varied by emotional
544 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[82] 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[82]. Face images were adult White female faces from the Karolinska Directed Emotional Faces database [83]. Each run (130 TRs, duration of 4:20) consisted of 48 trials (24 neutral faces, 24 fearful faces), each presented for 350ms. All fMRI analyses of this task used event-related designs.

MRI Acquisition

Participants under 18-years-old completed a mock scanning session before the MRI scan to acclimate to the scanner environment and practice lying still for data collection. Waves 1 and 2 were collected on a Siemens 3T TIM Trio MRI scanner using a standard radiofrequency head coil. A 2D spin echo image (TR, 4000 ms; TE, 40 ms; matrix size, 256 x 256; 4 mm thick; 0mm gap) was acquired in the oblique plane to guide slice configuration in both structural and functional scans. A whole-brain high-resolution T1-weighted anatomical scan (MPRAGE; 256 x 256 in-plane resolution; 256mm FOV; 192 x 1 mm sagittal slices) was acquired for each participant for registration of functional data. The task was presented through MR-compatible

goggles during scanning. T2*-weighted echoplanar images (interleaved slice acquisition) were collected at an oblique angle of ~30 degrees (130 volumes/run; TR=2000ms; TE=30 ms; flip angle=90°; matrix size=64 x 64; FOV=192 mm; 34 slices; 4 mm slice thickness; skip=0 mm). Wave 3 was collected on a Siemens 3T TIM Trio MRI scanner at a different location using identical acquisition parameters.

Behavioral Analyses

We used multilevel logistic regression models to estimate age-related changes in several task performance metrics. We fit separate models for the d' performance metric, overall 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, and cubic regressions with identical random effects structures. Model equations and results for all behavioral analyses can be found in the supplement (see supplemental methods p.8-9, sFigures 3-4).

Preregistered fMRI Pipeline

MPRAGE scans were skull-stripped using 3dSkullStrip from the Analysis of Functional NeuroImages (AFNI) software package[84] or FSL's brain extraction tool (BET)[85], depending on visual inspection of image quality. For each scan, an experimenter visually compared skull-stripped images using both software packages and chose the one judged to be more accurate. 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[86].

Scans over a threshold of >40 volumes with > .9mm framewise displacement were excluded from analysis (12 out of an initial 195, or 6.2%). The mean age of participants with excluded scans was 7.16 years 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. We included event-related regressors for fear and neutral faces (convolved with a double-gamma HRF), their temporal derivatives[87], and 24 head motion nuisance regressors (the 6 head realignment parameters, their temporal derivatives, and their squares[88]. Volumes with FD > .9mm were downweighed to 0 in the GLM. Pre-whitening was used to estimate and remove temporal autocorrelation[89]. 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[90]) by VanTieghem et al.[91].

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')[60]. Because analyzing fMRI data using all reasonable specifications was infeasible (i.e., possibilities are virtually infinite), we took the approach of 'sampling' from the

619 many reasonable or commonly-used analysis choices for each multiverse. Despite not being
620 completely comprehensive, this approach still allowed for thorough investigation into the
621 robustness of results. For all multiverse analyses, we constructed specification curves by
622 ranking models by their beta estimates (ascending) for parameters of interest for interpretation
623 and visualization[92–96]. Because specification choices were not preregistered, we did not conduct
624 formal null hypothesis testing of specification curves. Instead, we used a rough rule-of-thumb for
625 whether a finding was 'robust', or consistent across specification choices, by asking whether
626 more than 50% of the specifications resulted in a 95% posterior interval excluding 0 in the same
627 direction. In addition, to analyze in more detail the impact of specific choices, we submitted point
628 estimates for parameters of interest across all specifications to multiple regression models.
629 From these models, we examined the conditional effects of each analysis decision point on the
630 parameter of interest (see supplemental methods p.26, sFigures 11-13, 41-43, & 53-55).
631

632 *Multiverse amygdala reactivity analyses*

633 For amygdala reactivity analyses, we examined the robustness of age-related change
634 estimates to a variety of analytical decisions. In addition to the preregistered FSL-based
635 pipeline, we preprocessed data using C-PAC software (v1.4.1[97]). We used C-PAC to take
636 advantage of features supporting running multiple pipeline 'forks' in parallel (for example
637 multiple nuisance regression forks using the same registration). No spatial smoothing was used
638 in C-PAC pipelines (see supplemental methods p.10). Following C-PAC and FSL
639 preprocessing, we examined the impact of different sets of commonly-used analysis methods
640 on age-related change in amygdala reactivity. We varied analysis choices of GLM software,
641 hemodynamic response function, nuisance regressors, first-level GLM estimates, amygdala
642 ROI, exclusion criteria (exclude vs. include scans analyzed by Gee et al.[10]), group-level model
643 outlier treatment, and group-level model covariates (see Table 2 & supplemental methods p.10-

13). 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 quadratic age-related changes (see sFigures 14-17) and ran a smaller set of analyses (24) to ask whether we could differentiate within-participant change over time from between-participant differences through alternative model parametrization (see supplemental methods p.14).

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 FD[88,98,99]). 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 FD (see sFigure 27).

Across preprocessing specifications, we also examined within-scan similarity of amygdala and whole-brain voxelwise reactivity patterns (see sFigures 19-20) and between-scan correlations of average amygdala reactivity estimates (sFigures 21-23). To understand the proportion of variance in amygdala reactivity explained by repeated measurements of the same participants across study timepoints, we estimated intraclass correlation coefficients (ICC) for amygdala reactivity for each preprocessing specification (see sFigure 23).

672 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. 25 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

673 [1]**Bolded** choices indicate those most closely matching preregistered pipelines.

674

675 *Change in amygdala reactivity across trials*

676 To probe whether amygdala reactivity exhibited within-scan change in an age-dependent
677 manner, we modeled reactivity to each face trial using a Least Squares Separate method
678 (LSS^{100}). After preprocessing, we used FEAT to fit 48 separate GLMs corresponding to each
679 trial in each scan. A given trial was modeled with its own regressor and the remaining 47 trials
680 were modeled with a single regressor. Each GLM also included 24 head motion nuisance
681 regressors and had TRs with framewise displacement > .9mm downweighted to 0. BOLD data
682 were high-pass filtered at .01Hz before the GLM. From each of the 48 GLMs, we extracted the
683 mean amygdala beta estimates corresponding to a contrast for each single trial > baseline.

684 We constructed separate multiverse analyses using three different methods for
685 measuring change in amygdala reactivity across trials. For method 1 (*slopes*), we measured
686 rank-order correlations between trial number and trial-wise amygdala betas. For method 2 (*trial*
687 *halves*), we split trials into the first half (trials 1-12) and second half (trials 13-24), and modeled
688 age-related change in each half. For method 3 (*single-trial models*), we constructed larger
689 multilevel models with individual trials as the unit of observation. We conducted several analysis
690 specifications for each method (see Table 2 & supplemental methods p.17-19), and generated
691 corresponding specification curves.

692

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

694 We applied multiverse analysis techniques towards examining age-related changes in
695 amygdala–mPFC FC using gPPI and beta-series correlation (BSC) methods. Briefly, gPPI
696 estimates functional connectivity by constructing an interaction term between the timecourse in
697 a seed region of interest and a stimulus (task) regressor. Voxels whose activity is well fit by this
698 interaction term (a psychological-physiological interaction, or PPI) are assumed to be

699 “functionally coupled” with the seed region in a way that depends on the behavioral task [54,101].
700 BSC offers a different way of estimating functioning connectivity, by constructing ‘timeseries’ of
701 beta values (i.e., a beta series) in a condition of interest for two regions of interest, and
702 calculating the product-moment correlation between those beta series.

703 We constructed separate specification curves for age-related change in gPPI and BSC
704 for each contrast. Across gPPI specifications, we varied whether to use a deconvolution step in
705 creating interaction regressors[51,52], as well as several other analysis decision points (see Table
706 2 & supplemental methods p.19). The deconvolution step applies to the preprocessed BOLD
707 data from the seed timecourse: these data are first deconvolved to estimate the “underlying
708 neural activity” that produced the BOLD signal[51], then these deconvolved signals are multiplied
709 with the task regressor (e.g., for fear faces). Finally, this new interaction term is convolved with
710 a hemodynamic response to produce the BOLD functional connectivity regressor of interest. We
711 preregistered constructing an mPFC ROI containing 120 voxels centered at the peak
712 coordinates reported by Gee at al.[10] for age-related change in fear > baseline gPPI (Talairach
713 2,32,8; or MNI 3,35,8). However, after preregistration we discovered that these peak
714 coordinates were quite close to the corpus collosum, and this ROI would have contained a high
715 proportion of white matter voxels relative to cortical voxels (though this was not true for the
716 mPFC ROI identified by Gee et al.[10]). To address this issue, we instead constructed three
717 spherical ROIs with 5mm radii; the first centered at the above peak coordinates, the second
718 shifted slightly anterior, and the third shifted slightly ventral relative to the second (see Figure 4).
719 Lastly, to examine amygdala functional connectivity with a more broadly-defined mPFC, we also
720 used a ‘large vmPFC’ mask encompassing many of the areas within the ventromedial prefrontal
721 cortex derived from Mackey & Petrides[102].

722 For BSC analyses, we used beta estimates from the LSS GLMs described above for
723 analyses of within-scan change in amygdala reactivity. Across BSC specifications we varied

analyses across several decision points (see Table 2 & supplemental methods p.21), including whether to include a correction for global signal (post-hoc distribution centering[103]). 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[39]. 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 (see supplemental methods p.22). We focused primarily on linear age-related change, but also examined quadratic associations (see sFigures 44-46 & 56-58). We separately examined group mean gPPI and BSC for each contrast (see sFigures 38 & 50), as well as associations between mean FD and both FC measures across specifications (see sFigures 47 & 59). 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 48-49).

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 and SCARED separation anxiety subscales as outcomes (see Table 2 & supplemental methods p.25).

749

750 *Model-fitting*

751 All statistical models fit at the group level were run in the R (v 3.6.1) computing
752 environment [104]. In order to most accurately model age-related changes in each of our
753 measures, we attempted to take into account both between-participants information and
754 repeated measurements within participants over time. Unless otherwise indicated, models were
755 estimated using Hamiltonian Monte Carlo sampling as implemented in the Stan programming
756 language through the brms package in R[105,106]. Unless otherwise indicated, all models used
757 package default weakly-informative priors and were run with 4 chains of 2000 sampling
758 iterations (1000 warmup) each (see supplemental methods p.14 and p.26 for syntax).

759

760 *Interactive visualizations*

761 Because static plots visualizing the model predictions for all models in each multiverse
762 would require far more page space than available, we created web-based interactive
763 visualization tools for exploring different model specifications and viewing the corresponding raw
764 (participant-level) data and fitted model predictions using R and Shiny[107]. These visualizations
765 can be found at https://pbloom.shinyapps.io/amygdala_mpfc_multiverse/

766

767 *Deviations from preregistration*

768 Although we largely completed the preregistered analyses, the current study includes
769 many analyses beyond those proposed in the initial preregistration. Because the additional
770 analyses (i.e., all multiverses) conducted here give us substantial analytical flexibility over that
771 initially indicated by preregistration, we consider all results here to be at least in part exploratory
772 (rather than completely confirmatory), despite the preregistered hypotheses. Additionally, we
773 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, emphasizing that without further methodological work, we consider all such choices in tandem as providing equal evidential value.

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

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.

Code Sharing & Materials 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/

References

1. Adolphs R. Fear, Faces, and the Human Amygdala. *Curr Opin Neurobiol*. 2008;18(2):166-172. doi:10.1016/j.conb.2008.06.006

2. Forbes EE, Phillips ML, Silk JS, Ryan ND, Dahl RE. Neural Systems of Threat Processing in Adolescents: Role of Pubertal Maturation and Relation to Measures of Negative Affect. *Developmental Neuropsychology*. 2011;36(4):429-452. doi:10.1080/87565641.2010.550178

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

4. Silvers JA, Shu J, Hubbard AD, Weber J, Ochsner KN. Concurrent and Lasting Effects of Emotion Regulation on Amygdala Response in Adolescence and Young Adulthood. *Developmental Science*. 2015;18(5):771-784. doi:10.1111/desc.12260

5. Sullivan RM, Perry RE. Mechanisms and Functional Implications of Social Buffering in Infants: Lessons from Animal Models. *Social Neuroscience*. 2015;10(5):500-511. doi:10.1080/17470919.2015.1087425

6. Pattwell SS, Liston C, Jing D, et al. Dynamic Changes in Neural Circuitry during Adolescence Are Associated with Persistent Attenuation of Fear Memories. *Nature Communications*. 2016;7:11475. doi:10.1038/ncomms11475

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

8. Meyer HC, Lee FS. Translating Developmental Neuroscience to Understand Risk for Psychiatric Disorders. *AJP*. 2019;176(3):179-185. doi:10.1176/appi.ajp.2019.19010091

9. Hariri AR, Tessitore A, Mattay VS, Fera F, Weinberger DR. The Amygdala Response to Emotional Stimuli: A Comparison of Faces and Scenes. *NeuroImage*. 2002;17(1):317-323. doi:10.1006/nimg.2002.1179

10. Gee DG, Humphreys KL, Flannery J, et al. A Developmental Shift from Positive to Negative Connectivity in Human Amygdala–Prefrontal Circuitry. *J Neurosci*. 2013;33(10):4584-4593. doi:10.1523/JNEUROSCI.3446-12.2013

11. Guyer AE, Monk CS, McClure-Tone EB, et al. A Developmental Examination of Amygdala Response to Facial Expressions. *J Cogn Neurosci*. 2008;20(9):1565-1582. doi:10.1162/jocn.2008.20114

12. Killgore WDS, Oki M, Yurgelun-Todd DA. Sex-specific developmental changes in amygdala responses to affective faces. *NeuroReport*. 2001;12(2):427.

13. Passarotti AM, Sweeney JA, Pavuluri MN. Neural Correlates of Incidental and Directed Facial Emotion Processing in Adolescents and Adults. *Soc Cogn Affect Neurosci*. 2009;4(4):387-398. doi:10.1093/scan/nsp029

14. Swartz JR, Carrasco M, Wiggins JL, Thomason ME, Monk CS. Age-Related Changes in the Structure and Function of Prefrontal Cortex–Amygdala Circuitry in Children and Adolescents: A Multi-Modal Imaging Approach. *NeuroImage*. 2014;86:212-220. doi:10.1016/j.neuroimage.2013.08.018

15. Telzer EH, Flannery J, Humphreys KL, et al. “The Cooties Effect”: Amygdala Reactivity to Opposite-versus Same-Sex Faces Declines from Childhood to Adolescence. *J Cogn Neurosci*. 2015;27(9):1685-1696. doi:10.1162/jocn_a_00813

16. Decety J, Michalska KJ, Kinzler KD. The Contribution of Emotion and Cognition to Moral Sensitivity: A Neurodevelopmental Study. *Cereb Cortex*. 2012;22(1):209-220. doi:10.1093/cercor/bhr111

17. Silvers JA, Insel C, Powers A, et al. 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*. 2017;25:128-137. doi:10.1016/j.dcn.2016.06.005

18. Vink M, Derks JM, Hoogendam JM, Hillegers M, Kahn RS. Functional Differences in Emotion Processing during Adolescence and Early Adulthood. *NeuroImage*. 2014;91:70-76. doi:10.1016/j.neuroimage.2014.01.035

19. Joseph JE, Zhu X, Gundran A, et al. Typical and Atypical Neurodevelopment for Face Specialization: An fMRI Study. *J Autism Dev Disord*. 2015;45(6):1725-1741. doi:10.1007/s10803-014-2330-4

20. Todd RM, Evans JW, Morris D, Lewis MD, Taylor MJ. The Changing Face of Emotion: Age-Related Patterns of Amygdala Activation to Salient Faces. *Soc Cogn Affect Neurosci*. 2011;6(1):12-23. doi:10.1093/scan/nsq007

21. Hare TA, Tottenham N, Galvan A, Voss HU, Glover GH, Casey BJ. Biological Substrates of Emotional Reactivity and Regulation in Adolescence During an Emotional Go-Nogo Task. *Biological Psychiatry*. 2008;63(10):927-934. doi:10.1016/j.biopsych.2008.03.015

22. Vijayakumar N, Pfeifer JH, Flournoy JC, Hernandez LM, Dapretto M. Affective reactivity during adolescence: Associations with age, puberty and testosterone. *Cortex*. 2019;117:336-350. doi:10.1016/j.cortex.2019.04.024

882 23. Kujawa A, Wu M, Klumpp H, et al. Altered Development of Amygdala-Anterior Cingulate
883 Cortex Connectivity in Anxious Youth and Young Adults. *Biological Psychiatry: Cognitive*
884 *Neuroscience and Neuroimaging*. 2016;1(4):345-352. doi:10.1016/j.bpsc.2016.01.006

885 24. Pfeifer JH, Masten CL, Moore WE, et al. Entering Adolescence: Resistance to Peer
886 Influence, Risky Behavior, and Neural Changes in Emotion Reactivity. *Neuron*. 2011;69(5).
887 doi:10.1016/j.neuron.2011.02.019

888 25. Pine DS, Grun J, Zarahn E, et al. Cortical Brain Regions Engaged by Masked Emotional Faces
889 in Adolescents and Adults: An fMRI Study. *Emotion*. 2001;1(2):137-147. doi:10.1037/1528-
890 3542.1.2.137

891 26. Wu M, Kujawa A, Lu LH, et al. Age-related changes in amygdala–frontal connectivity during
892 emotional face processing from childhood into young adulthood. *Human Brain Mapping*.
893 2016;37(5):1684-1695. doi:10.1002/hbm.23129

894 27. Yurgelun-Todd DA, Killgore WDS. Fear-related activity in the prefrontal cortex increases
895 with age during adolescence: A preliminary fMRI study. *Neuroscience Letters*.
896 2006;406(3):194-199. doi:10.1016/j.neulet.2006.07.046

897 28. Zhang Y, Padmanabhan A, Gross JJ, Menon V. Development of human emotion circuits
898 investigated using a Big-Data analytic approach: Stability, reliability, and robustness. *J*
899 *Neurosci*. Published online July 22, 2019:0220-19. doi:10.1523/JNEUROSCI.0220-19.2019

900 29. Silvers JA, Insel C, Powers A, et al. vlPFC–vmPFC–Amygdala Interactions Underlie Age-
901 Related Differences in Cognitive Regulation of Emotion. *Cereb Cortex*. 2017;27(7):3502-
902 3514. doi:10.1093/cercor/bhw073

903 30. Perlman SB, Pelphrey KA. Developing Connections for Affective Regulation: Age-Related
904 Changes in Emotional Brain Connectivity. *Journal of Experimental Child Psychology*.
905 2011;108(3):607-620. doi:10.1016/j.jecp.2010.08.006

906 31. Marek S, Tervo-Clemmens B, Calabro FJ, et al. Towards Reproducible Brain-Wide
907 Association Studies. *bioRxiv*. Published online August 22, 2020:2020.08.21.257758.
908 doi:10.1101/2020.08.21.257758

909 32. Lieberman MD, Eisenberger NI, Crockett MJ, Tom SM, Pfeifer JH, Way BM. Putting feelings
910 into words: Affect labeling disrupts amygdala activity in response to affective stimuli.
911 *Psychological Science*. 2007;18(5):421-428. doi:10.1111/j.1467-9280.2007.01916.x

912 33. Sergerie K, Chochol C, Armony JL. The role of the amygdala in emotional processing: A
913 quantitative meta-analysis of functional neuroimaging studies. *Neuroscience &*
914 *Biobehavioral Reviews*. 2008;32(4):811-830. doi:10.1016/j.neubiorev.2007.12.002

34. Botvinik-Nezer R, Holzmeister F, Camerer CF, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. 2020;582(7810):84-88. doi:10.1038/s41586-020-2314-9

35. Bowring A, Maumet C, Nichols TE. Exploring the impact of analysis software on task fMRI results. *Human Brain Mapping*. 2019;40(11):3362-3384. doi:10.1002/hbm.24603

36. Jo HJ, Lee J-M, Kim J-H, et al. Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques. *NeuroImage*. 2007;34(2):550-564. doi:10.1016/j.neuroimage.2006.09.047

37. Achterberg M, van der Meulen M. Genetic and environmental influences on MRI scan quantity and quality. *Developmental Cognitive Neuroscience*. 2019;38:100667. doi:10.1016/j.dcn.2019.100667

38. Bryce NV, Flournoy JC, Guassi Moreira JF, et al. Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *NeuroImage*. 2021;243:118487. doi:10.1016/j.neuroimage.2021.118487

39. Di X, Zhang Z, Biswal BB. Understanding psychophysiological interaction and its relations to beta series correlation. *bioRxiv*. Published online January 31, 2020:322073. doi:10.1101/322073

40. Cisler JM, Bush K, Steele JS. A Comparison of Statistical Methods for Detecting Context-Modulated Functional Connectivity in fMRI. *NeuroImage*. 2014;84:1042-1052. doi:10.1016/j.neuroimage.2013.09.018

41. Crone EA, Elzinga BM. Changing brains: how longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. *WIREs Cognitive Science*. 2015;6(1):53-63. doi:10.1002/wcs.1327

42. Madhyastha T, Peverill M, Koh N, et al. Current Methods and Limitations for Longitudinal fMRI Analysis across Development. *Developmental Cognitive Neuroscience*. Published online November 22, 2017. doi:10.1016/j.dcn.2017.11.006

43. Carpenter KLH, Angold A, Chen N-K, et al. Preschool Anxiety Disorders Predict Different Patterns of Amygdala-Prefrontal Connectivity at School-Age. *PLOS ONE*. 2015;10(1):e0116854. doi:10.1371/journal.pone.0116854

44. King KM, Littlefield AK, McCabe CJ, Mills KL, Flournoy J, Chassin L. Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. *Developmental Cognitive Neuroscience*. 2018;33:54-72. doi:10.1016/j.dcn.2017.11.009

45. Madhyastha T, Peverill M, Koh N, et al. Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*. 2018;33:118-128. doi:10.1016/j.dcn.2017.11.006

46. Glenn ND. Distinguishing Age, Period, and Cohort Effects. In: Mortimer JT, Shanahan MJ, eds. *Handbook of the Life Course*. Handbooks of Sociology and Social Research. Springer US; 2003:465-476. doi:10.1007/978-0-306-48247-2_21

47. Elliott ML, Knodt AR, Ireland D, et al. What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis: *Psychological Science*. Published online June 3, 2020. doi:10.1177/0956797620916786

48. Sauder CL, Hajcak G, Angstadt M, Phan KL. Test-retest reliability of amygdala response to emotional faces. *Psychophysiology*. 2013;50(11):1147-1156. doi:10.1111/psyp.12129

49. Zuo X-N, Xu T, Milham MP. Harnessing reliability for neuroscience research. *Nat Hum Behav*. 2019;3(8):768-771. doi:10.1038/s41562-019-0655-x

50. Killgore WDS, Yurgelun-Todd DA. Unconscious Processing of Facial Affect in Children and Adolescents. *Social Neuroscience*. 2007;2(1):28-47. doi:10.1080/17470910701214186

51. Gitelman DR, Penny WD, Ashburner J, Friston KJ. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage*. 2003;19(1):200-207. doi:10.1016/S1053-8119(03)00058-2

52. Di X, Biswal BB. Psychophysiological Interactions in a Visual Checkerboard Task: Reproducibility, Reliability, and the Effects of Deconvolution. *Front Neurosci*. 2017;11. doi:10.3389/fnins.2017.00573

53. Di X, Reynolds RC, Biswal BB. Imperfect (de)Convolution May Introduce Spurious Psychophysiological Interactions and How to Avoid It. *Human Brain Mapping*. 2017;38(4):1723-1740. doi:10.1002/hbm.23413

54. O'Reilly JX, Woolrich MW, Behrens TEJ, Smith SM, Johansen-Berg H. Tools of the Trade: Psychophysiological Interactions and Functional Connectivity. *Soc Cogn Affect Neurosci*. 2012;7(5):604-609. doi:10.1093/scan/nss055

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

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

57. Grady CL, Rieck JR, Nichol D, Rodrigue KM, Kennedy KM. Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. *Human Brain Mapping*. Published online 2020. doi:https://doi.org/10.1002/hbm.25217

58. Flournoy JC, Vijayakumar N, Cheng TW, Cosme D, Flannery JE, Pfeifer JH. Improving practices and inferences in developmental cognitive neuroscience. *Developmental Cognitive Neuroscience*. 2020;45:100807. doi:10.1016/j.dcn.2020.100807

59. Bridgeford EW, Wang S, Yang Z, et al. Big Data Reproducibility: Applications in Brain Imaging. *bioRxiv*. Published online January 18, 2020:802629. doi:10.1101/802629

60. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci*. 2016;11(5):702-712. doi:10.1177/1745691616658637

61. Gelman A, Loken E. 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*. 2014;102(6):460-466.

62. Del Giudice M, Gangestad SW. 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*. 2021;4(1):2515245920954925. doi:10.1177/2515245920954925

63. Bossier H, Roels SP, Seurinck R, et al. The empirical replicability of task-based fMRI as a function of sample size. *NeuroImage*. 2020;212:116601. doi:10.1016/j.neuroimage.2020.116601

64. Nee DE. fMRI replicability depends upon sufficient individual-level data. *Communications Biology*. 2019;2(1):1-4. doi:10.1038/s42003-019-0378-6

65. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*. 2013;14(5):365-376. doi:10.1038/nrn3475

66. Gelman A, Carlin J. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspect Psychol Sci*. 2014;9(6):641-651. doi:10.1177/1745691614551642

67. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. 2009;12(5):535-540. doi:10.1038/nn.2303

68. Richeson JA, Todd AR, Trawalter S, Baird AA. Eye-Gaze Direction Modulates Race-Related Amygdala Activity. *Group Processes & Intergroup Relations*. 2008;11(2):233-246. doi:10.1177/1368430207088040

1018 69. Telzer EH, Humphreys KL, Shapiro M, Tottenham N. Amygdala Sensitivity to Race Is Not
1019 Present in Childhood but Emerges over Adolescence. *Journal of Cognitive Neuroscience*.
1020 2012;25(2):234-244. doi:10.1162/jocn_a_00311

1021 70. Ciric R, Wolf DH, Power JD, et al. Benchmarking of participant-level confound regression
1022 strategies for the control of motion artifact in studies of functional connectivity.
1023 *Neuroimage*. 2017;154:174-187. doi:10.1016/j.neuroimage.2017.03.020

1024 71. Gilmore A, Buser N, Hanson JL. Variations in Structural MRI Quality Significantly Impact
1025 Commonly-Used Measures of Brain Anatomy. *bioRxiv*. Published online May 7,
1026 2020:581876. doi:10.1101/581876

1027 72. Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath H-O. Age-specific CT and MRI
1028 templates for spatial normalization. *NeuroImage*. 2012;61(4):957-965.
1029 doi:10.1016/j.neuroimage.2012.03.020

1030 73. Boubela RN, Kalcher K, Huf W, et al. fMRI measurements of amygdala activation are
1031 confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain
1032 regions. *Scientific Reports*. 2015;5(1):10499. doi:10.1038/srep10499

1033 74. Fair DA, Miranda-Dominguez O, Snyder AZ, et al. Correction of respiratory artifacts in MRI
1034 head motion estimates. *NeuroImage*. 2020;208:116400.
1035 doi:10.1016/j.neuroimage.2019.116400

1036 75. Gratton C, Dworetsky A, Coalson RS, et al. Removal of high frequency contamination from
1037 motion estimates in single-band fMRI saves data without biasing functional connectivity.
1038 *bioRxiv*. Published online April 21, 2020:837161. doi:10.1101/837161

1039 76. Achenbach TM. *Integrative Guide for the 1991 CBCL/4-18, Ysr, and Trf Profiles*. 1st US-1st
1040 Printing edition. Univ Vermont/Dept Psychiatry; 1991.

1041 77. US Census Bureau. U.S. Census Bureau QuickFacts: Los Angeles County, California.
1042 Published 2021. Accessed March 24, 2021.
1043 https://www.census.gov/quickfacts/losangelescountycalifornia

1044 78. Heeringa SG, Wagner J, Torres M, Duan N, Adams T, Berglund P. Sample designs and
1045 sampling methods for the Collaborative Psychiatric Epidemiology Studies (CPES).
1046 *International Journal of Methods in Psychiatric Research*. 2004;13(4):221-240.
1047 doi:https://doi.org/10.1002/mpr.179

1048 79. Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. Psychometric
1049 Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A
1050 Replication Study. *Journal of the American Academy of Child & Adolescent Psychiatry*.
1051 1999;38(10):1230-1236. doi:10.1097/00004583-199910000-00011

1052 80. Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of Symptoms of DSM-
1053 IV Anxiety and Depression in Children: A Revised Child Anxiety and Depression Scale.
1054 *Behaviour Research and Therapy*. 2000;38(8):835-855. doi:10.1016/S0005-7967(99)00130-
1055 8

1056 81. Beretta L, Santaniello A. Nearest Neighbor Imputation Algorithms: A Critical Evaluation.
1057 *BMC Med Inform Decis Mak*. 2016;16(Suppl 3). doi:10.1186/s12911-016-0318-z

1058 82. Wager TD, Nichols TE. Optimization of experimental design in fMRI: a general framework
1059 using a genetic algorithm. *Neuroimage*. 2003;18(2):293-309.

1060 83. Calvo MG, Lundqvist D. Facial expressions of emotion (KDEF): Identification under
1061 different display-duration conditions. *Behav Res*. 2008;40(1):109-115.
1062 doi:10.3758/BRM.40.1.109

1063 84. Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance
1064 Neuroimages. *Comput Biomed Res*. 1996;29(3):162-173.

1065 85. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*.
1066 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015

1067 86. Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and
1068 Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*.
1069 2002;17(2):825-841. doi:10.1006/nimg.2002.1132

1070 87. Pernet CR. Misconceptions in the use of the General Linear Model applied to functional
1071 MRI: a tutorial for junior neuro-imagers. *Front Neurosci*. 2014;8:1.
1072 doi:10.3389/fnins.2014.00001

1073 88. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic
1074 correlations in functional connectivity MRI networks arise from subject motion.
1075 *NeuroImage*. 2012;59(3):2142-2154. doi:10.1016/j.neuroimage.2011.10.018

1076 89. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate
1077 linear modeling of FMRI data. *Neuroimage*. 2001;14(6):1370-1386.
1078 doi:10.1006/nimg.2001.0931

1079 90. Fischl B. FreeSurfer. *NeuroImage*. 2012;62(2):774-781.
1080 doi:10.1016/j.neuroimage.2012.01.021

1081 91. VanTieghem M, Korom M, Flannery J, et al. Longitudinal changes in amygdala,
1082 hippocampus and cortisol development following early caregiving adversity.
1083 *Developmental Cognitive Neuroscience*. 2021;48:100916. doi:10.1016/j.dcn.2021.100916

1084 92. Cosme D, Lopez R. Neural indicators of food cue reactivity, regulation, and valuation and
1085 their associations with body composition and daily eating behavior. Published online
1086 March 7, 2020. doi:10.31234/osf.io/23mu5

1087 93. Klapwijk E, van den Bos W, Tamnes CK, Mills KL, Raschle N. *Opportunities for Increased*
1088 *Reproducibility and Replicability of Developmental Cognitive Neuroscience*. PsyArXiv; 2019.
1089 doi:10.31234/osf.io/fxjzt

1090 94. Orben A, Przybylski AK. The association between adolescent well-being and digital
1091 technology use. *Nat Hum Behav*. 2019;3(2):173-182. doi:10.1038/s41562-018-0506-1

1092 95. Simonsohn U, Simmons J, Nelson L. Specification Curve: Descriptive and Inferential
1093 Statistics on All Reasonable Specifications. *Marketing Papers*. Published online January 1,
1094 2015. doi:10.2139/ssrn.2694998

1095 96. Simonsohn U, Simmons JP, Nelson LD. Specification curve analysis. *Nat Hum Behav*.
1096 2020;4(11):1208-1214. doi:10.1038/s41562-020-0912-z

1097 97. Craddock C, Sikka S, Cheung B, et al. Towards Automated Analysis of Connectomes: The
1098 Configurable Pipeline for the Analysis of Connectomes (C-PAC). In: ; 2013.
1099 doi:10.3389/conf.fninf.2013.09.00042

1100 98. Satterthwaite TD, Wolf DH, Loughead J, et al. Impact of in-scanner head motion on
1101 multiple measures of functional connectivity: Relevance for studies of neurodevelopment
1102 in youth. *NeuroImage*. 2012;60(1):623-632. doi:10.1016/j.neuroimage.2011.12.063

1103 99. Satterthwaite TD, Wolf DH, Ruparel K, et al. Heterogeneous Impact of Motion on
1104 Fundamental Patterns of Developmental Changes in Functional Connectivity during Youth.
1105 *Neuroimage*. 2013;83:45-57. doi:10.1016/j.neuroimage.2013.06.045

1106 100. Abdulrahman H, Henson RN. Effect of trial-to-trial variability on optimal event-related
1107 fMRI design: Implications for Beta-series correlation and multi-voxel pattern analysis.
1108 *NeuroImage*. 2016;125:756-766. doi:10.1016/j.neuroimage.2015.11.009

1109 101. McLaren DG, Ries ML, Xu G, Johnson SC. A generalized form of context-dependent
1110 psychophysiological interactions (gPPI): A comparison to standard approaches.
1111 *NeuroImage*. 2012;61(4):1277-1286. doi:10.1016/j.neuroimage.2012.03.068

1112 102. Mackey S, Petrides M. Architecture and Morphology of the Human Ventromedial
1113 Prefrontal Cortex. *Eur J Neurosci*. 2014;40(5):2777-2796. doi:10.1111/ejn.12654

1114 103. Fox MD, Zhang D, Snyder AZ, Raichle ME. The Global Signal and Observed Anticorrelated
1115 Resting State Brain Networks. *Journal of Neurophysiology*. 2009;101(6):3270-3283.
1116 doi:10.1152/jn.90777.2008

104. R Core Team. R: The R Project for Statistical Computing. Published 2017. Accessed October 30, 2019. https://www.r-project.org/

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

106. Bürkner P-C. *Brms: Bayesian Regression Models Using "Stan."*; 2019. Accessed December 6, 2019. https://CRAN.R-project.org/package=brms

107. Beeley C. *Web Application Development with R Using Shiny*. Packt Publishing Ltd; 2013.