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

Paul Alexander Bloom¹, Michelle VanTieghem², Laurel Gabard-Durnam³, Dylan G. Gee⁴, Jessica Flannery⁵, Christina Caldera⁶, Bonnie Goff⁶, Eva H. Telzer⁷, Kathryn L. Humphreys⁸, Dominic S. Fareri⁹, Mor Shapiro¹⁰, Sameah Algharazi¹¹, Niall Bolger¹, Mariam Aly¹, & Nim Tottenham¹

Affiliations:
¹Columbia University
²Virta Health
³Northeastern University
⁴Yale University
⁵Limbix Health
⁶University of California Los Angeles
⁷University of North Carolina at Chapel Hill
⁸Vanderbilt University
⁹Adelphi University
¹⁰Kaiser Permanente
¹¹City 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. Choices including software package, spatial smoothing, treatment of head motion, parcellation, and functional connectivity approach can also impact results and qualitatively change findings. 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.

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 predicting that both amygdala reactivity and amygdala–mPFC connectivity 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, 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. 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 across earlier (left) and later (right) trials.
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 $\alpha = .05$), 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-
related changes may not necessarily indicate ‘true’ developmental growth. High uncertainty in estimating average within-participant change could be driven by several factors, including true heterogeneity in individual trajectories, low measurement reliability\textsuperscript{47–49}, scanner differences across longitudinal timepoints, or unmodeled variables impacting amygdala reactivity. Additionally, the within-participants terms represent a smaller age range (a maximum of 4 years for any given participant), relative to the broader age range assessed by the between-participants terms (18 years), which may have placed additional limits on identifying reliable within-participant change.

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

Amygdala–mPFC Functional Connectivity

Multiverse analyses did not find consistent evidence for age-related change in amygdala–mPFC connectivity using gPPI or BSC (Figures 4-5). Crucially, however, specification curves did not find strong evidence against such age-related change, as we did not observe precise and consistent ‘null’ estimates across specifications. This absence of evidence for age-related changes in task-evoked amygdala–mPFC connectivity contrasts prior work suggesting a developmental shift in such connectivity\textsuperscript{10,23,26}, but is consistent with recent work\textsuperscript{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 analyses⁵¹, recent work has shifted guidelines on its use, and it may not be recommended for block designs⁵⁹,⁶². 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 regressors⁵³. 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 connectivity⁵⁴ (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 generally⁴⁰.

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\textsuperscript{10,23,55,56}. However, given that analyses of brain-behavior associations may require imaging cohorts much larger than the current sample \textsuperscript{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.

\textit{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\textsuperscript{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\textsuperscript{58}. Particularly when methodological ‘gold standards’ have not been determined, specification curves may be informative for examining the impacts of potential analysis decisions\textsuperscript{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\textsuperscript{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 \textit{a priori}, analyses are vulnerable to problems of analytic flexibility\textsuperscript{61}, and inclusion of less justified specifications can bias results\textsuperscript{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\textsuperscript{31,63}, the number of longitudinal study sessions per participant\textsuperscript{44}, and the duration of the task\textsuperscript{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\textsuperscript{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\textsuperscript{58}. The fact that some specifications used data included in previous similar analyses of the same cohort\textsuperscript{10} also limits the confirmatory power of the present study\textsuperscript{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.\textsuperscript{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. 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.

Finally, our findings on age-related change in amygdala and mPFC function may be biased or confounded by age-related differences in head motion, anatomical image quality and alignment, signal dropout, and physiological artifacts. 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

<table>
<thead>
<tr>
<th>Aim</th>
<th>Preregistered Hypothesis</th>
<th>Analysis Methodology</th>
<th>Key Findings</th>
</tr>
</thead>
</table>
| 1a. Age-related change in amygdala reactivity to fearful faces | Amygdala reactivity to fearful faces will decrease with age, such that younger children will have more positive amygdala reactivity (higher BOLD response to fearful 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 fearful 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 fearful 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 fearful 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/](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](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. 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
sample representative of the Los Angeles or United States populations\textsuperscript{78}, and Hispanic or Latinx participants were particularly underrepresented. Further sample demographics can be found in the supplementary materials (see sTables 1-2, sFigures 1-2).

\textit{Separation Anxiety}

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

\textit{Emotion Discrimination Task}

Participants completed either two (at wave 3) or three (at waves 1 and 2) runs of a modified ‘go/no-go’ task with emotional faces during fMRI scanning. Runs varied by emotional expression (fear, happy, sad), and within each run participants viewed emotional faces
interspersed with neutral faces. To ensure that participants were paying attention, they were asked to press a button whenever they saw a neutral face (no response was required for any other face expression). The order of the runs was counterbalanced across participants; the stimuli within each run were pseudorandomized\(^2\) 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\(^2\). Face images were adult White female faces from the Karolinska Directed Emotional Faces database. 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=2000 ms; 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 or FSL’s brain extraction tool (BET), 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.
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\(^6^7\), and 24 head motion nuisance regressors (the 6 head realignment parameters, their temporal derivatives, and their squares\(^8^8\). Volumes with FD > .9mm were downweighed to 0 in the GLM. Pre-whitening was used to estimate and remove temporal autocorrelation\(^8^9\). 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\(^9^0\)) by VanTieghem et al.\(^9^1\).

**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’)\(^6^0\). Because analyzing fMRI data using all reasonable specifications was infeasible (i.e., possibilities are virtually infinite), we took the approach of ‘sampling’ from the...
many reasonable or commonly-used analysis choices for each multiverse. Despite not being completely comprehensive, this approach still allowed for thorough investigation into the robustness of results. For all multiverse analyses, we constructed specification curves by ranking models by their beta estimates (ascending) for parameters of interest for interpretation and visualization. Because specification choices were not preregistered, we did not conduct formal null hypothesis testing of specification curves. Instead, we used a rough rule-of-thumb for whether a finding was ‘robust’, or consistent across specification choices, by asking whether more than 50% of the specifications resulted in a 95% posterior interval excluding 0 in the same direction. In addition, to analyze in more detail the impact of specific choices, we submitted point estimates for parameters of interest across all specifications to multiple regression models. From these models, we examined the conditional effects of each analysis decision point on the parameter of interest (see supplemental methods p.26, sFigures 11-13, 41-43, & 53-55).

Multiverse amygdala reactivity analyses

For amygdala reactivity analyses, we examined the robustness of age-related change estimates to a variety of analytical decisions. In addition to the preregistered FSL-based pipeline, we preprocessed data using C-PAC software (v1.4.1). We used C-PAC to take advantage of features supporting running multiple pipeline ‘forks’ in parallel (for example multiple nuisance regression forks using the same registration). No spatial smoothing was used in C-PAC pipelines (see supplemental methods p.10). Following C-PAC and FSL preprocessing, we examined the impact of different sets of commonly-used analysis methods on age-related change in amygdala reactivity. We varied analysis choices of GLM software, hemodynamic response function, nuisance regressors, first-level GLM estimates, amygdala ROI, exclusion criteria (exclude vs. include scans analyzed by Gee et al.), group-level model outlier treatment, and group-level model covariates (see Table 2 & supplemental methods p.10-
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\textsuperscript{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).
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 | Method of quantifying within-scan change | Slopes across trials, trials split into halves, single-trial models |
FSL preproc & GLM, high-pass filter, 24 motion regressors, 2G HRF, beta estimates, included previously analyzed scans, and robust group-level regression | 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 |
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 |

2a. Age-related change in amygdala–mPFC functional connectivity (FC) to fear faces > baseline, as measured by (gPPI) | Deconvolution step | Yes, No |
FSL preproc & GLM, high-pass filter, 24 motion regressors, 2G HRF, bilateral amygdala ROI in MNI space | 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 |
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 |

2b. Age-related change in amygdala–mPFC functional connectivity to fear faces > baseline, as measured by (BSC) | 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 | 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 |

Table 2: Summary of forking pipelines used in analyses for each aim

1Bolded choices indicate those most closely matching preregistered pipelines.
Change in amygdala reactivity across trials

To probe whether amygdala reactivity exhibited within-scan change in an age-dependent manner, we modeled reactivity to each face trial using a Least Squares Separate method (LSS100). After preprocessing, we used FEAT to fit 48 separate GLMs corresponding to each trial in each scan. A given trial was modeled with its own regressor and the remaining 47 trials were modeled with a single regressor. Each GLM also included 24 head motion nuisance regressors and had TRs with framewise displacement > .9mm downweighted to 0. BOLD data were high-pass filtered at .01Hz before the GLM. From each of the 48 GLMs, we extracted the mean amygdala beta estimates corresponding to a contrast for each single trial > baseline.

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

Multiverse amygdala–mPFC functional connectivity (FC) analyses

We applied multiverse analysis techniques towards examining age-related changes in amygdala–mPFC FC using gPPI and beta-series correlation (BSC) methods. Briefly, gPPI estimates functional connectivity by constructing an interaction term between the timecourse in a seed region of interest and a stimulus (task) regressor. Voxels whose activity is well fit by this interaction term (a psychological-physiological interaction, or PPI) are assumed to be
“functionally coupled” with the seed region in a way that depends on the behavioral task \(^{54,101}\).

BSC offers a different way of estimating functioning connectivity, by constructing ‘timeseries’ of beta values (i.e., a beta series) in a condition of interest for two regions of interest, and calculating the product-moment correlation between those beta series.

We constructed separate specification curves for age-related change in gPPI and BSC for each contrast. Across gPPI specifications, we varied whether to use a deconvolution step in creating interaction regressors\(^{51,52}\), as well as several other analysis decision points (see Table 2 & supplemental methods p.19). The deconvolution step applies to the preprocessed BOLD data from the seed timecourse: these data are first deconvolved to estimate the “underlying neural activity” that produced the BOLD signal\(^{51}\), then these deconvolved signals are multiplied with the task regressor (e.g., for fear faces). Finally, this new interaction term is convolved with a hemodynamic response to produce the BOLD functional connectivity regressor of interest. We preregistered constructing an mPFC ROI containing 120 voxels centered at the peak coordinates reported by Gee at al.\(^{10}\) for age-related change in fear > baseline gPPI (Talairach 2,32,8; or MNI 3,35,8). However, after preregistration we discovered that these peak coordinates were quite close to the corpus collosum, and this ROI would have contained a high proportion of white matter voxels relative to cortical voxels (though this was not true for the mPFC ROI identified by Gee et al.\(^{10}\)). To address this issue, we instead constructed three spherical ROIs with 5mm radii; the first centered at the above peak coordinates, the second shifted slightly anterior, and the third shifted slightly ventral relative to the second (see Figure 4).

Lastly, to examine amygdala functional connectivity with a more broadly-defined mPFC, we also used a ‘large vmPFC’ mask encompassing many of the areas within the ventromedial prefrontal cortex derived from Mackey & Petrides\(^{102}\).

For BSC analyses, we used beta estimates from the LSS GLMs described above for analyses of within-scan change in amygdala reactivity. Across BSC specifications we varied
analyses across several decision points (see Table 2 & supplemental methods p.21), including whether to include a correction for global signal (post-hoc distribution centering103). 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 regions39. 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).
Model-fitting

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

Interactive visualizations

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

Deviations from preregistration

Although we largely completed the preregistered analyses, the current study includes many analyses beyond those proposed in the initial preregistration. Because the additional analyses (i.e., all multiverses) conducted here give us substantial analytical flexibility over that initially indicated by preregistration, we consider all results here to be at least in part exploratory (rather than completely confirmatory), despite the preregistered hypotheses. Additionally, we note that BSC analyses, analyses of change in amygdala reactivity across trials, and analyses
of associations between all brain measures and separation anxiety were exploratory, and conducted after we had seen the results of the preregistered reactivity and gPPI analyses. In addition, to avoid possible selection bias introduced by the analytical flexibility inherent in running many parallel analyses, we consider all analysis specifications simultaneously, 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
Kathryn L. Humphreys: Investigation, Writing – Review & Editing
Dominic Fareri: Investigation, Writing – Review & Editing
Eva Telzer: Investigation, Writing – Review & Editing
Sameah Algharazi: Validation, Writing – Review & Editing
Mariam Aly: Methodology, Formal Analysis, Supervision, Writing – Review & Editing
Niall Bolger: Methodology, Formal Analysis, Data Curation, Writing – Original Draft, Supervision, Funding
Kathryn L. Humphreys: Investigation, Methodology, Resources, Data Curation, Writing - Original Draft, Supervision, Funding

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


