Development of human emotion circuits investigated using a Big-Data analytic approach: Stability, reliability, and robustness

https://doi.org/10.1523/JNEUROSCI.0220-19.2019

Cite as: J. Neurosci 2019; 10.1523/JNEUROSCI.0220-19.2019
Received: 25 January 2019
Revised: 27 June 2019
Accepted: 1 July 2019

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.
Development of human emotion circuits investigated using
a Big-Data analytic approach: Stability, reliability, and robustness

Short title: Development of human emotion circuits

Yuan Zhang*, Aarthi Padmanabhan1, James J. Gross2, Vinod Menon1,3,4*

1 Department of Psychiatry and Behavioral Sciences
2 Department of Psychology
3 Department of Neurology and Neurological Sciences
4 Stanford Neuroscience Institute

Stanford University
Stanford, CA 94305, USA

* Correspondence:
Yuan Zhang (yuanzh@stanford.edu) and Vinod Menon (menon@stanford.edu; Lead
Contact)

Number of pages: 51
Number of figures and tables: 6; 7
Number of words for abstract, introduction, and discussion: 230; 627; 1459

Declaration of interests

The authors declare no competing interests.

Acknowledgements

This work was supported by the NIH (grants EB022907 and NS086085), the Stanford Child
Health Research Institute Transdisciplinary Initiatives Program, and a Stanford Institute for
Computational & Mathematical Engineering GPU computing seed grant. We thank Drs.
Kaustubh Supekar, Shaozheng Qin, Stacie Warren, and Christina Young for helpful
comments on our manuscript, and Katherine Duberg for proofreading the manuscript. We
thank the PNC consortium for sharing their data, and Drs. Ted Satterthwaite, Daniel Wolf,
and Monica Calkins for graciously replying to our queries during the early stages of the study.
Abstract

Emotion perception is fundamental to affective and cognitive development and is thought to involve distributed brain circuits. Efforts to chart neurodevelopmental changes in emotion have been severely hampered by narrowly focused approaches centered on activation of individual brain regions and small sample sizes. Here we investigate the maturation of human functional brain circuits associated with identification of fear, anger, sadness, happiness, and neutral faces using a large sample of 759 children, adolescents, and adults (ages 8-23; female/male = 419/340). Network analysis of emotion-related brain circuits revealed three functional modules, encompassing lateral frontoparietal, medial prefrontal-posterior cingulate, and subcortical-posterior insular cortices, with hubs in medial prefrontal, but not posterior cingulate, cortex. This overall network architecture was stable by age eight, and it anchored maturation of circuits important for salience detection and cognitive control, as well as dissociable circuit patterns across distinct emotion categories. Our findings point to similarities and differences in functional circuits associated with identification of fearful, angry, sad, happy, and neutral faces, and reveal aspects of brain circuit organization underlying emotion perception that are stable over development as well as features that change with age. Reliability analyses demonstrated the robustness of our findings and highlighted the importance of large samples for probing functional brain circuit development. Our study emphasizes a need to focus beyond amygdala circuits and provides a robust neurodevelopmental template for investigating emotion perception and identification in psychopathology.
Significance Statement

Emotion perception is fundamental to cognitive and affective development. However, efforts to chart neurodevelopmental changes in emotion perception have been hampered by narrowly focused approaches centered on the amygdala and prefrontal cortex and small sample sizes. Using a large sample of 759 children, adolescents, and adults and a multipronged analytical strategy, we investigated the development of brain network organization underlying identification and categorization of fearful, happy, angry, sad, and neutral facial expressions. Results revealed a developmentally-stable modular architecture that anchored robust age-related and emotion category-related changes in brain connectivity across multiple brain systems that extend far beyond amygdala circuits and provide a new template for investigation of emotion processing in the developing brain.
Introduction

Identification of distinct facial expressions, which is fundamental to recognizing the emotional state of others, begins in infancy and continues to develop throughout childhood and adolescence (Batty and Taylor, 2006; Thomas et al., 2007; Rodger et al., 2015; Leitzke and Pollak, 2016; Theurel et al., 2016). How emotions are identified and categorized in the developing brain remains a fundamental open question. Analysis of task-evoked brain circuits using a ‘Big Data’ approach has the potential to accelerate foundational knowledge of emotion perception circuit development, and provide robust templates for understanding miswiring of functional circuits in developmental psychopathologies. Here, we leverage the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2014) to address critical gaps in our knowledge regarding the functional circuits of emotion perception and their development.

Meta-analyses of neuroimaging studies in adults have identified widely-distributed, and overlapping functional systems engaged by distinct emotion categories (Kober et al., 2008; Fusar-Poli et al., 2009). In particular, brain regions associated with three large-scale functional brain networks are relevant for emotion processing based on their extensive role in integrating salient cues across cortical and limbic systems. The ‘salience network’ facilitates rapid detection of personally relevant and salient emotional cues in the environment (Seeley et al., 2007; Menon and Uddin, 2010; Menon, 2011), the ‘default mode network’ is important for mentalizing and inferring the emotional states of others (Greicius et al., 2003; Blakemore,
2008), and the frontoparietal ‘central executive network’ is involved in attention, reappraisal, and regulation of reactivity to emotional stimuli (Dosenbach et al., 2007; Gyurak et al., 2011; Tupak et al., 2014; Braunstein et al., 2017; Sperduti et al., 2017). However, previous neurodevelopmental studies of emotion perception have focused almost exclusively on the amygdala and its connectivity with the medial prefrontal cortex (mPFC) (Gee et al., 2013; Vink et al., 2014; Kujawa et al., 2016; Wolf and Herringa, 2016; Wu et al., 2016), with inconsistent findings (see Table 1 for a summary). While some studies found no age-related changes (Kujawa et al., 2016; Wolf and Herringa, 2016; Wu et al., 2016), others have reported increases (Vink et al., 2014; Wolf and Herringa, 2016) or decreases with age (Gee et al., 2013; Kujawa et al., 2016; Wu et al., 2016) in amygdala activation or connectivity with the mPFC. In contrast, intrinsic functional connectivity studies have highlighted protracted developmental changes in distributed brain networks (Power et al., 2010; Menon, 2013), including developmental changes in amygdala connectivity with subcortical, paralimbic, and limbic structures, polymodal association, and ventromedial prefrontal cortex (Qin et al., 2012). These findings highlight a critical need for characterizing the development of task-evoked functional brain circuits and network organization underlying emotion perception (Barrett and Satpute, 2013).

Here, we characterize the neurodevelopment of emotion perception circuits leveraging high-quality data (N=759) from the emotion-identification fMRI task in the PNC, an unprecedented “Big Data” initiative for charting brain development in a large community
sample (Satterthwaite et al., 2014) (Figure 1). We first used meta-analysis and community
detection methods to identify and characterize the modular architecture of brain circuits
associated with perception and identification of fearful, angry, sad, happy, and neutral faces.
We then investigated aspects of brain circuit organization that are stable over development in
addition to aspects that changed with age. Lastly, given the challenges of small sample sizes
(Button et al., 2013; Szucs and Ioannidis, 2017; Turner et al., 2018), high false positive rates
(Eklund et al., 2016; Szucs and Ioannidis, 2017), and lack of replicability in neuroimaging
research (Ioannidis, 2005; Eklund et al., 2016; Szucs and Ioannidis, 2017), we evaluated the
reliability of our age-related findings on randomly subsampled independent datasets across a
wide range of sample sizes (Figure 1d) and assessed the replicability of previously reported
findings. Our findings advance a new understanding of emotion circuit development in the
human brain.

Materials and Methods

Participants

Participants were selected from the Philadelphia Neurodevelopmental Cohort (PNC) aged 8 –
23 years (N = 1495). Data were excluded based on the following criteria: (1) lack of complete
emotion task fMRI data; (2) high in-scanner motion (mean scan-to-scan displacement > 0.2
mm or maximum scan-to-scan displacement > 1 mm or maximum displacement > 3 mm); (3)
missing brain coverage; (4) medical rating > 2 (i.e., having significant or major medical
problems); and (5) overall task performance (accuracy) < 70%. 759 participants (fMRI
sample) survived these exclusion criteria and were included in further behavioral and
euroimaging analyses (Figure 1a and Table 2). Additional behavioral analysis was
performed to investigate age-related effects on emotion categorization accuracy and reaction
time in a larger sample of 1338 participants (behavioral sample) who had valid behavioral
data. 157 participants with medical issues or incomplete data or incapable of performing the
task were excluded from any analysis. All study procedures were approved by the
institutional review boards of both the University of Pennsylvania and the Children’s
Hospital of Philadelphia (Satterthwaite et al., 2014).

Behavioral and demographic differences between the neuroimaging and excluded
samples
We examined whether participants in the fMRI (N = 759) and excluded (N = 579) samples are
similar cross-sections of the overall population in terms of age, sex, and behavioral
performance (accuracy and RT). The excluded sample included participants who did not meet
our criteria for inclusion in fMRI data analysis, but were included in the behavioral analysis.

PNC Emotion identification fMRI task
Participants were presented with faces and were required to categorize them as one of five
emotion categories: neutral, fear, anger, sadness, or happiness. Each face was presented for
5.5 seconds followed by a jittered inter-stimulus interval ranging between 0.5 and 18.5
seconds, during which a complex crosshair that matched perceptual qualities of the face was
displayed. There were 12 faces per emotion category resulting in 60 total trials (Satterthwaite et al., 2014).

Neuroimaging data acquisition and preprocessing

Images were acquired using a 3T Siemens TIM Trio scanner (32 channel head coil), with a single-shot, interleaved multi-slice, gradient-echo, echo planar imaging sequence (TR/TE = 3000/32 ms, flip angle = 90°, field of view = 192 × 192 mm², matrix = 64 × 64, slice thickness = 3 mm). FMRI data were preprocessed and analyzed using Statistical Parametric Mapping software SPM8 (Wellcome Trust Department of Cognitive Neurology, London, UK). Images were slice-time corrected, motion corrected, re-sampled to 3 mm × 3 mm × 3 mm isotropic voxels, normalized to MNI space, spatially smoothed with a 6 mm FWHM Gaussian filter, and temporally filtered using a high-pass filter with a cutoff frequency of 1/128 Hz.

Head Motion

Participants with mean scan-to-scan displacement less than or equal to 0.2 mm and maximum scan-to-scan displace less than or equal to 1 mm and maximum displacement less than or equal to 3 mm were included in further analysis. In addition to correction for motion correction at the whole brain level during fMRI preprocessing (as described above), to rule out any potential residual effects of motion, mean scan-to-scan displacement was included as a nuisance covariate in all our analyses.
Network analysis

Network nodes. To overcome limitations of extant neurodevelopmental studies of emotion perception and categorization, which have predominantly focused on the amygdala (Gee et al., 2013; Vink et al., 2014; Kujawa et al., 2016; Wolf and Herringa, 2016; Wu et al., 2016), we conducted a meta-analysis to identify cortical and subcortical brain regions that are consistently involved in emotion-related processes. Specifically, using the meta-analysis toolbox Neurosynth (Yarkoni et al., 2011) and the search term “emotion”, we identified local peaks (at least 8 mm apart) within each cluster in the forward and reverse inference maps. We selected peaks in frontal, parietal, medial temporal, limbic and subcortical areas. Regions of Interest (ROIs) were defined as a 6 mm sphere with an identified peak as center. If a unilateral peak was detected at a specific region, we included its contralateral counterpart by flipping the peak coordinate along the x axis.

Taken together, nodes encompassed the ventromedial, dorsomedial, ventrolateral and dorsolateral prefrontal cortex (vmPFC, vmPFC, vIPFC, vIPFC), lateral orbitofrontal cortex (IOFC), inferior and superior parietal lobule (IPL, SPL), subgenual, pregenual, dorsal anterior and posterior cingulate cortex (sgACC, pgACC, dACC, PCC), pre-supplementary motor area (preSMA), basolateral amygdala (BLA), dorsal anterior and posterior insula (dAI, PI), hippocampus and fusiform gyrus (FFG). An additional pair of dorsolateral prefrontal ROIs superior to the pair identified by Neurosynth were included based on two recent emotion-regulation meta-analyses (Buhle et al., 2014; Kohn et al., 2014). We also included
nodes in the ventral anterior insula (vAI), centromedial amygdala (CMA), bed nucleus stria terminalis (BNST) and nucleus accumbens (NAc) based on their well-known roles in perception of negative and positive emotions (LeDoux, 2007; Deen et al., 2011; Chang et al., 2013; Floresco, 2015; Lebow and Chen, 2016; Namkung et al., 2017). The vAI node was identified using Neurosynth and search term “anxiety” (Paulus and Stein, 2006; Etkin, 2010; Menon, 2011; Blackford and Pine, 2012; Qin et al., 2014), the CMA was identified based on its cytoarchitectonically-defined probabilistic map using the Anatomy toolbox (Eickhoff et al., 2005), the BNST node was based on the meta-analysis by Avery et al. (2014), and NAc was identified based on the Harvard-Oxford atlas distributed with FSL (http://www.fmrib.ox.ac.uk/fsl/). These procedures identified 50 ROIs consistently associated with emotion processing (Figure 1b and Table 3). Finally, we further verified that the activation maps of emotion-related topics encompassed the maps reported in a large meta-analytic database of fMRI studies (Rubin et al., 2017).

Emotion-related node-level activity. Task-related brain activation was identified using a general linear model (GLM) implemented in SPM8. At individual level, brain activation representing correct trials for each emotion task condition (neutral, fear, anger, sad, or happy) was modeled using boxcar functions convolved with a canonical hemodynamic response function and a temporal derivative to account for voxelwise latency differences in hemodynamic response. An error regressor was also included in the model to account for the influence of incorrect trials. Additionally, six head movement parameters generated from the
realignment procedure were included as regressors of no interest. Serial correlations were accounted for by modeling the fMRI time series as a first-degree autoregressive process. For each emotion task condition, ROI activity was obtained by averaging the estimated beta values across voxels within that ROI.

*Emotion-related inter-node connectivity.* Functional connectivity was assessed using a beta series analysis approach (Rissman et al., 2004). At individual level, brain activation related to each trial was modeled using boxcar functions convolved with a canonical hemodynamic response function. Six head movement parameters were included as regressors of no interest. Serial correlations were accounted for by modeling the fMRI time series as a first-degree autoregressive process. Trial-wise ROI activity was obtained by averaging the estimated beta values across voxels within that ROI. A beta-series was derived for each ROI each emotion condition by concatenating trial-wise ROI activity belonging to the same emotion condition. Only ROI activity corresponding to correct trials were used to construct beta-series. Condition-wise beta series were demeaned and used to calculate Pearson correlation between all pairs of ROIs. Pearson correlation coefficients were Fisher-Z transformed and used to index functional connectivity strength and form emotion-related networks in each participant (*Figure 1c*).

*Emotion-related network structure.* To determine common patterns of community structure across all participants, we used a consensus-based approach for identifying functional
modules (Lancichinetti and Fortunato, 2012; Dwyer et al., 2014) associated with each emotion. In each participant, the Louvain algorithm (Blondel et al., 2008), as implemented in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010), was used to first determine the modular structure of the signed, weighted connectivity matrix (Newman, 2004, 2006). To handle potential degeneracy of community assignments, we repeated this procedure 1000 times in each participant (Lancichinetti and Fortunato, 2012; Dwyer et al., 2014). In each iteration, we generated a co-classification matrix in which each element contained 1 if two nodes were part of the same module and 0 otherwise. The resulting 1000 co-classification matrices were averaged to generate a co-occurrence matrix indicating the probability of two nodes being in the same module across 1000 iterations. Co-occurrence matrices were then averaged across participants to produce a positive, weighted group consistency matrix, with higher weight in elements of this matrix indicating that two nodes were more frequently classified in the same module across participants. This group consistency matrix was used to determine a consensus partition (Lancichinetti and Fortunato, 2012). To evaluate the influence of resolution parameter Gamma, which tunes the size and number of communities obtained from modularity maximization (Fortunato and Barthelemy, 2007; Sporns and Betzel, 2016), the above consensus-based approach was repeated at Gamma values ranging from 0.1 to 2.0 in increments of 0.1.

To assess the significance of the modular structure we compared the actual Q to the Q derived from null networks. Specifically, the group consistency matrix was used to create 1000 null
networks with preserved weight and degree distributions. For each null network, weak edges
with weight below 0.1 were first set to 0 and Q was calculated 1000 times and used to
establish an empirical null distribution of Q. Q was considered significant if it was larger than
the 95th percentile of the empirical null distribution.

Emotion-related hubs. To identify hubs in the affective circuits and their development with
age, we first computed nodal degree for each individual and each emotion category. Degree
of a node $i$ is defined as the number of edges connected to that node:

$$k_i = \sum_{j \in n, j \neq i} e_{ij}$$

where $e_{ij}$ is the connection status (0 or 1) between node $i$ and node $j$, and $n$ is the total number
of nodes (Rubinov and Sporns, 2010). Because the functional matrix is fully connected, a
sparsity condition, based on the ratio of the actual edge number to the maximum possible
edge number in a network, was applied to the connectivity matrix before computing nodal
degree. To minimize the impact of setting an arbitrary level of sparsity, we calculated degree
over a wide range of sparsity (i.e., $10\% \leq$ sparsity $\leq 40\%$ with an interval of 5%) and then
computed area under the curve (AUC), to obtain an integrated metric of node degree as in
previous studies examining functional circuits (Yan et al., 2013; Lv et al., 2016). A node with
high degree indicates high interconnectivity with other brain nodes. Hub nodes were defined
as nodes with integrated degree metric at least 1 standard deviation above the mean across all
nodes.
Generalization of ROIs to the Brainnetome atlas. We used the Brainnetome atlas to determine robustness and generalizability of findings from the 50 emotion-related metanalysis-derived ROIs. An advantage of the Brainnetome atlas, compared to other many other functional brain parcellations, is that it includes both cortical and subcortical clusters with appropriate anatomical labels (Fan et al., 2016). Specifically, we identified Brainnetome clusters in which each of the 50 ROIs was located; because 2 ROI pairs were located in the same Brainnetome cluster, this procedure yielded 46 matching clusters. We then repeated all our analyses using these spatially extended clusters.

Statistical analysis

Linear vs. Nonlinear age models. The remainder of our analyses involved assessing age-related changes in task performance, regional activity, regional connectivity, and intra- and inter-module connectivity. For task performance, we analyzed both the fMRI (N= 759) and behavioral (N=1338) samples. We examined three models – linear, quadratic, and cubic spline – to test for linear and nonlinear age-related effects, and used Bayesian Information Criteria (BIC) to determine the best model. The vast majority of the analyses indicated that non-linear models did not fit the data better than linear models. Where appropriate we report nonlinear age effects that survived FDR correction for multiple comparisons.

Linear mixed-effect model. To examine main effects of age, emotion, and age by emotion interactions on task performance and brain measures, we used R package lme4 (Bates et al.,
298 2015) to perform a linear mixed-effect analysis. As fixed effects, we entered age, emotion,
299 and age by emotion interactions (regressors of interest) as well as sex, race, and head
300 movement (regressors of no interest) into the model. As random effects, we had intercepts for
301 participants. Significance of regressor coefficients were obtained by F-tests.
302
303 Multiple linear regression of age-related changes. For significant age by emotion
304 interactions, we performed additional analyses to examine age-related changes in each
305 emotion category. Specifically, we employed multiple linear regression models, with
306 behavioral or brain measure as dependent variable, age as a covariate of interest, and sex,
307 race, and head movement as covariates of no interest.
308
309 Development of emotion circuits
310 Development of emotion-related network structure. To determine how network structure
311 changes with age, we first determined community structure for each participant and each
312 group as described above. We then computed the Jaccard index between individual and group
313 consensus community structures. Specifically, overlap in community structure was computed
314 using a co-occurrence matrix, with 1 indicating two nodes belonging to the same module and
315 0 indicating two nodes belonging to different modules. The Jaccard index is defined as the
316 ratio of the intersection between two matrices divided by their union, ranging from 0 (no
317 overlap) to 1 (full overlap). A linear mixed-effect model was then applied to the Jaccard index
318 to determine main effects of age, emotion, and age by emotion interactions. We also
examined community structure in three distinct age groups, children (8-12 years), adolescents (13-17 years) and adults (18-23 years), to provide additional insights into the development of emotion circuits.

**Development of emotion-related intra- and inter-module connectivity.** To examine main effect of age, emotion, and age by emotion interactions on functional interactions between the module, we assessed the intra-module connectivity, by computing the mean strength of all connectional weights within a module, and inter-module interaction, by computing the mean strength of connectional weights between modules, for each individual and each emotion category.

**Development of emotion-related hubs.** To investigate the development of hub nodes, we first computed the probability of an ROI being a hub in a 1-year age window, separately for each emotion category. Specifically, for each participant and each emotion category, we identified hubs (see above) and counted, for each node, how many times it functioned as hub across participants under a specific emotion category. This count was divided by the total number of participants within that specific age window to derive a node’s hub probability. To assess the significance of hub probability, we constructed an empirical null distribution for each region and each age window under each emotion category. Specifically, we first permuted hub identity across all nodes within each individual, by randomly designating which nodes were hubs while keeping the total number of hubs constant over 1000 iterations, and then
calculated the probability of a brain region being a hub node within each age window for all
permutations. Only nodes with a probability above the 95th percentile of its empirical null
distribution were considered as a significant hub.

Classification analysis
We used a non-linear SVM classification algorithm with 10-fold cross-validation to assess the
discriminability of regional connectivity measures between all pairs of emotion categories.
The python scikit-learn package (https://scikit-learn.org/stable/) was used to perform this
analysis. Permutation test was used to assess the performance of the classifier. Specifically,
for each pair of emotion categories, data were divided into 10 folds, with those belonging to
the same participant included in the same fold. A non-linear SVM classifier was trained using
nine folds, leaving out one fold. The samples in the left-out fold were then predicted using
this classifier and an accuracy was derived by calculating the percentage of correct
predictions. This procedure was repeated ten times, and a final classification accuracy was
derived by calculating the mean of the 10 left-out accuracy, which was used as a measure of
model performance. Next, the statistical significance of the model was assessed using
permutation test. In each permutation, emotion category labels (e.g., neutral vs. fear) were
randomly exchanged with a probability of 0.5 for each participant. A classifier was trained,
and a classification accuracy was derived using the same scheme described above.
Classification accuracy from 100 permutations were used to construct the empirical null
distribution from which a p-value was computed.
Reliability analysis

Behavior. We examined the reliability of our developmental findings and assessed the effect of sample size on reliability for each emotion category. First, we randomly drew N (sample size) participants (50 \( \leq N \leq 700 \), incremented by 50) from either the fMRI (N=759), or behavioral (N=1338), sample without replacement and with the proportion of children (ages 8-12 years), adolescents (ages 13-17 years) and adults (ages 18-23 years) in each subsample being matched to the fMRI, or behavioral, sample. Second, we assessed developmental changes in the subsample using multiple linear regression models. Third, we repeated the first two steps 1000 times at each sample size. Finally, we calculated the probability of observing a significant age effect with the same direction as the one revealed in the full sample across 1000 runs for each sample size, which was used as the measure of reliability. We defined findings with reliability above 0.7 as reliable because it is unlikely (\( p < 2.2e-16 \)) to observe a significant effect 700 times or more out of 1000 runs, assuming equal probability to observe a significant or null effect in a run.

Brain networks. We randomly drew N participants from the fMRI sample (N = 759) and assessed age-related changes in the subsample using linear mixed-effect model. This was repeated 1000 times at each sample size across a wide range of sample sizes (50 \( \leq N \leq 400 \)). Next, correlation between age effects in regional activity or connectivity across all brain regions from a subsample and that from the full sample was calculated and then averaged.
across subsamples at each sample size. The resulting average correlation was used as the
measure of network-level reliability of overall developmental pattern across all brain regions.
We defined reliable findings as those with correlation above 0.7 (i.e., $R$-square $> 0.5$), the
point at which the model explains more variance than is unexplained. Finally, we also
constructed empirical null distributions for reliability measured as average correlation, using
the same approach described above, with an additional step that age was permuted in each
subsample before assessing developmental changes. Average correlation above the 95th
percentile of its empirical null distribution was considered as significantly different from null.

Reproduction analysis of previous developmental findings related to amygdala circuitry
We examined the development of amygdala activity and connectivity in the current PNC
cohort using coordinates reported in three published task-fMRI studies using similar
emotion face perception tasks, contrasts, and age range as the PNC (Gee et al., 2013; Kujawa
et al., 2016; Wu et al., 2016). Tables 1 and 4 provide details of the studies and coordinates
used. Moreover, we performed additional gPPI (generalized psychophysiological interaction)
analyses to establish amygdala-mPFC connectivity and examine age-related changes as a
similar method (gPPI or PPI) was used in the three published studies. Specifically, gPPI
model included one regressor of seed time series, five regressors of task conditions (neutral,
fear, anger, sadness, or happiness) and their interaction with seed time series, one regressor
of error trials and its interaction with seed time series, six regressors of head movement, and
one regressor of constant. Replication analysis was then applied to amygdala-mPFC gPPI 
connectivity.

Data and code availability
Raw data associated with this work are available through dbGaP

All data analysis scripts used in the study will be made publicly available upon publication.

Results
Behavioral maturation of emotion identification
We first examined the main effects of age, emotion, and age by emotion interactions on 
accuracy (i.e., proportion of correct responses) and median reaction time of emotion 
identification in the fMRI sample (N= 759), which included participants who had valid 
behavioral and neuroimaging data.

Accuracy vs age. Our analysis revealed no main effect of age (p > 0.1) but a significant effect 
of emotion (F (4, 3028) = 187.156, p < 2.2e-16) and age by emotion interactions (F (4, 3028) 
= 5.370, p = 0.0003). Further analysis revealed that accuracy increased with age for Fear (r = 
0.151, p = 4.71e-05) and Anger (r = 0.152, p = 3.83e-05) conditions (Figure 2a-i), but not for 
Sad (r = -0.028, p = 0.457), Happy (r = 0.064, p = 0.084) or Neutral (r = 0.055, p = 0.144) 
conditions.
Reaction time vs age. Results revealed a significant age effect ($F(1, 1785) = 5.370, p = 1.15e-07$), showing that median reaction time decreased with age such that older participants were faster at identifying emotions. Moreover, there is a significant effect of emotion ($F(4, 3028) = 397.167, p < 2.2e-16$) and age by emotion interactions ($F(4, 3028) = 4.023, p = 0.003$). Further analysis revealed that median reaction time decreased with age for all five emotions (Figure 2a-ii): Fear ($r = -0.230, p = 4.62e-10$), Anger ($r = -0.173, p = 3.11e-6$), Sad ($r = -0.111, p = 0.003$), Happy ($r = -0.235, p = 1.75e-10$) and Neutral ($r = -0.207, p = 2.01e-08$). The significant age by emotion interaction was characterized by differences in slope between the Sad and Fear conditions and between the Sad and Happy conditions.

Reliability of age-related changes. We examined the reliability of age-related changes in relation to sample size. For each emotion category, we used multiple linear regression model to assess age-accuracy and age-reaction time associations across a wide range of sample sizes (50 $\leq N \leq 700$, increased by 50) with the proportion of children (ages 8-12 years), adolescents (ages 13-17 years) and adults (ages 18-23 years) in each subsample matched to the full sample (1000 times per sample size). Reliability was measured as probability of observing a significant correlation between age and accuracy or reaction time with the same direction as the one revealed in the full sample. For Fear and Anger, the two emotion categories that showed a significant age-related increase in accuracy in the full sample, reliability increased with sample size (Figure 2c-i). Reliability of negative age-reaction time associations increased with sample size across all five emotion categories (Figure 2c-ii). In
general, a sample size of 300-350 was needed to identify reliable age-accuracy associations
and a sample size of 200-300 was needed to identify reliable age-reaction time association.
Results also demonstrate that age-related improvements in emotion identification were
strongest for fear and anger, the two high arousal negative emotions.

**Behavioral maturation of emotion identification assessed using the full behavioral sample**

Most developmental neuroimaging studies typically restrict analysis of fMRI-task behavioral
data to participants with high levels of accuracy and low head movement during scanning, as
noted above. The large PNC dataset allowed us to investigate potential biases in age-related
effects in a robust manner using a larger sample of 1338 participants.

**Accuracy vs age.** Results revealed significant main effect of age \((F (1, 4515) = 53.814, p < 2.6e-13)\), showing that accuracy increased with age such that older participants were better at
identifying emotions. Moreover, there was a significant main effect of emotion \((F (4, 5344) = 344.866, p < 1.86e-264)\), and age by emotion interactions \((F (4, 5344) = 10.434, p < 2.04e-08)\). Further analysis revealed that accuracy increased with age for Neutral \((r = 0.078, p = 0.006)\), Fear \((r = 0.150, p = 9.32-08)\), Anger \((r = 0.129, p = 6.43e-06)\), and Happy \((r = 0.058, p = 0.044)\) conditions, but not for the Sad condition \((r = 0.036, p = 0.203)\) (Figure 2b-i). Additional analysis revealed that the age by emotion interaction was driven by slope
differences between Fear and each of the other four conditions, between Anger and Happy
conditions, and between Neutral and Happy conditions. While all significant findings revealed in the fMRI sample were replicated in the larger behavioral sample, we also noted a few additional significant effects in the behavioral sample, including a main effect of age as well as associations between accuracy and age under Neutral and Happy conditions. However, these additional age-related effects during Neutral and Happy conditions were weaker than those during Fear and Anger.

Reaction time vs age. Results revealed a significant main age effect ($F(1, 3091) = 139.818, p = 1.38e-31$), showing that median reaction time decreased with age such that older participants were faster at identifying emotions. Moreover, there was a significant main effect of emotion ($F(4, 5344) = 594.997, p = 0.000$) and age by emotion interactions ($F(4, 5344) = 9.162, p = 2.25e-07$). Further analysis revealed that median reaction time decreased with age for all five emotions (Figure 2b-ii): Neutral ($r = -0.300, p = 9.57e-28$), Fear ($r = -0.146, p = 1.85e-21$), Anger ($r = -0.221, p = 4.76e-15$), Sad ($r = -0.185, p = 4.84e-11$), and Happy ($r = -0.260, p = 5.66e-21$). Additional analysis revealed that age by emotion interaction was driven by slope differences between Fear and Anger/Sad/Happy conditions and between Neutral and Anger/Sad/Happy conditions. Notably, all significant findings revealed in 759 participants were replicated in this larger behavioral sample.

Reliability of age-related changes. Reliability of age-related changes as a function of sample size were conducted using the same procedures as ones used above for the fMRI sample. For
Fear and Anger, the two emotion categories that showed a significant age-related increase in accuracy in both the fMRI and the full behavioral samples, reliability increased with sample size, and in general, a sample size of 300-350 was needed to identify reliable age-accuracy associations (Figure 2c-i). This finding is not surprising given that the effect size during Fear and Anger conditions were comparable across the two samples. Moreover, for Sad and Happy, the two emotion categories that showed a significant age effect only in the full behavioral sample, reliability increased with sample size but remained low (< 0.5) even at a sample size of 700, reflecting the small effect size under these two conditions. In addition, reliability of negative age-reaction time associations increased with sample size across all five emotion categories, and in general, a sample size of 150-200 was needed to identify reliable age-reaction time associations (Figure 2c-ii). Results also demonstrate that age-related improvements in emotion identification were strongest for Fear and Anger, the two high arousal negative emotions.

Behavioral and demographic differences between the neuroimaging and excluded samples

Finally, we examined whether participants in the fMRI (N = 759) and excluded (N = 579) samples are similar cross-sections of the overall population in terms of age, sex, and behavioral performance (accuracy and RT). The excluded sample included participants who did not meet our criteria for inclusion in fMRI data analysis, but were included in the behavioral analysis. We found a significant difference in age between the fMRI and excluded samples (t (1186) = 11.68, p < 2.2e-16), with participants in the fMRI sample (age: 16.20 ± 3.27) older than participants in the excluded sample (age: 13.99 ± 3.56). Moreover, sex and
exclusion/inclusion was not independent ($\chi^2(1) = 6.52, p = 0.011$), with higher female/male ratio in the fMRI sample (F/M = 419/340) than that in the excluded sample (F/M = 278/301).

Lastly, we found significant differences in accuracy ($t (781) = 8.83, p < 2.2e-16$) and RT ($t (998) = -8.87, p < 2.2e-16$) between the two samples, with participants in the fMRI sample showing higher accuracy and faster response (accuracy: $0.92 \pm 0.05$; RT: $2.29 \pm 0.36$) than participants in the excluded sample (accuracy: $0.88 \pm 0.10$; RT: $2.51 \pm 0.51$). These findings are not surprising as children, and boys in particular, usually tend to move more in the scanner (Yuan et al., 2009; Dosenbach et al., 2017; Engelhardt et al., 2017) so the excluded sample includes more boys and younger participants, leading to age and performance differences in the full sample of 1338 participants.

**Meta-analysis of brain areas involved in emotion**

To take a broad and inclusive view of cortical and subcortical brain regions involved in emotion perception, we proceeded in a two-step manner. First, we used the meta-analysis toolbox Neurosynth (Yarkoni et al., 2011) and the search term “emotion” to identify the most consistently reported cortical and subcortical regions implicated in emotion processing, including ventromedial, dorsomedial, ventrolateral and dorsolateral prefrontal cortex (vmPFC, dmPFC, vLPFC, dLPFC), lateral orbitofrontal cortex (lOFC), inferior and superior parietal lobule (IPL, SPL), subgenual, pregenual, dorsal anterior and posterior cingulate cortex (sgACC, pgACC, dACC, PCC), pre-supplementary motor area (preSMA), basolateral amygdala (BLA), ventral and dorsal anterior insula (vAI, dAI), posterior insula (PI),
hippocampus (Hipp), and fusiform gyrus (FFG) (Figure 1b and Table 3). Second, given recent theoretical focus on the salience network (Menon, 2011), specifically the subdivisions of the insula and the amygdala, we identified additional distinct functional subdivisions of the anterior insula (vAI: ventral anterior insula) and amygdala (CMA: centromedial amygdala) and additional subcortical regions, including the bed nucleus of the stria terminalis (BNST) and the nucleus accumbens (NAc), thought to play distinct functional roles in emotion processing (LeDoux, 2007; Deen et al., 2011; Chang et al., 2013; Floresco, 2015; Lebow and Chen, 2016; Namkung et al., 2017). The brain areas identified by this analysis showed strong overlap with salience, central executive, and default mode networks (Seeley et al., 2007; Menon, 2011).

Modular structure of emotion-related circuits is established in childhood and remains developmentally stable

To investigate how the functional organization of emotion perception circuits change with age, we first identified the modular structure of inter-regional connectivity, computed using beta series correlation (Rissman et al., 2004), for each emotion task condition. Next, we determined common patterns of modular structure across all participants for each emotion condition, which was used as a template for further analysis. Across all five emotion categories, we observed three consistent network modules corresponding to a frontoparietal (FP) module, a mPFC/PCC module, and a subcortical/posterior insula (SPI) module (Figure 3a). This 3-module structure was stable across emotion categories for all values of Gamma >
0.5; results are reported here for Gamma = 1, the most commonly used value in most fMRI studies. The modularity value $Q$ for Gamma = 1 was highly significant compared to empirical null networks with preserved weight and degree distributions ($p < 0.001$).

Next, we assessed main effects of age, emotion, and age by emotion interactions on the Jaccard overlap between individual and template modular structure. Results revealed a significant emotion effect ($F(4, 3028) = 3.656, p = 0.006$) but no significant effect of age or age by emotion interactions ($p > 0.1$), indicating that the overall modular structure for all emotion conditions was stable with age. Additional analysis using the Jaccard overlap between individual and adult-group modular structure revealed similar results. This was further confirmed by findings of modular structure within distinct age groups: children (8-12 years), adolescents (13-17 years) and adults (18-23 years) (Figure 3b). Notably, this modular architecture maps closely onto the triple-network model (Menon, 2011). Finally, a similar 3-module structure was observed when using the 46 spatially extended clusters, demonstrating the robustness and generalizability of our finding with respect to Brainnetome functional brain parcellations (Fan et al., 2016).

Development of hubs underlying emotion perception

We next identified which nodes served as hubs and examined whether these change with age. Briefly, we identified hub regions for each individual, computed the probability of a brain region being a hub within a one-year age window for each emotion category and assessed the
significance of hub probability by permutation. We found that hubs resided in the medial
prefrontal nodes (pgACC, vmPFC, and dmPFC) of the mPFC/PCC module (Figure 4a) and
were stable over age across all five emotion categories (Figure 4b and 4c).

Developmental changes in intra- and inter-module connectivity

We examined main effects of age, emotion, and age by emotion interactions on intra- and
inter-module connectivity. Results revealed a significant main effect of age for intra-FP
connectivity ($F(1, 2692) = 7.015, p = 0.008$) and for connectivity between FP and SPI ($F(1,
2766) = 5.253, p = 0.022$; Figure 3c), indicating that intra-FP and FP-SPI connectivity
increased with age. Results revealed a significant emotion effect in intra-FP ($F(4, 3028) =
7.487, p = 5.35e-06$), intra-SPI ($F(4, 3028) = 4.184, p = 0.002$), between FP and mPFC/PCC
($F(4, 3028) = 10.546, p = 1.75e-08$), between FP and SPI ($F(4, 3028) = 3.980, p = 0.003$),
and between mPFC/PCC and SPI ($F(4, 3028) = 5.531, p = 1.9e-04$) connectivity (Figure 3c
and 3d). Age by emotion interactions is significant for intra-mPFC/PCC connectivity ($F(4,
3028) = 2.724, p = 0.028$). Further analysis revealed that intra-mPFC/PCC connectivity
increased with age under Fear ($F(1, 754) = 3.910, p = 0.048$) and Anger ($F(1, 754) = 4.579,$
$p = 0.033$) but not the other three emotion categories ($ps > 0.1$).

Developmental changes in inter-regional connectivity

29
Using the modular architecture identified in the previous sections we examined main effects of age, emotion, and age by emotion interactions on inter-regional connectivity. All results are reported at $q < 0.05$, FDR corrected for multiple comparisons.

*Main effect of age.* We found age-related increases within FP and SPI modules, between FP and SPI modules and between mPFC/PCC and SPI modules, including left dlPFC connectivity with right dAI, right vAI connectivity with right sgACC and bilateral PI, right BNST connectivity with right IPL and SPL, and right BLA connectivity with left PI (Figure 5a). We found age-related decreases within mPFC/PCC and SPI modules, including right sgACC connectivity with left vmPFC and left CMA connectivity with right CMA and left PI.

In addition, we found two links – left NAc connectivity with bilateral dACC – which showed significant quadratic age effects, consistent with findings in the literature regarding enhanced sensitivity of dopaminergic reward circuits in adolescence (Padmanabhan et al., 2011; Padmanabhan and Luna, 2014).

*Main effect of emotion.* We found widespread emotion effect, especially within FP module, between FP and mPFC/PCC modules, between FP and SPI modules, and between mPFC/PCC and SPI modules (Figure 5a). Classification analysis revealed that it was sufficient to discriminate all pairs of emotion categories using all inter-regional links (Figure 5e), with classification accuracy ranging from 56.2% to 67.2% and significantly higher than chance level ($p < 0.01$) as assessed by permutation test.
Age by emotion interactions. No age by emotion interactions were found.

Reliability of age-related changes. To examine the robustness of our age-related changes and the effect of sample size on reliability, we performed reliability analysis (Figure 1d). Briefly, we randomly drew N (sample size) participants from the full sample and assessed age-related changes in the subsample using linear mixed-effect model. This was repeated 1000 times at each sample size across a wide range of sample sizes (50 $\leq$ N $\leq$ 400, increased by 50). Correlation between age effects from a subsample and that from the full sample was calculated and then averaged across subsamples at each sample size. The resulting average correlation was used to measure the reliability of overall developmental patterns across all inter-regional links. We found that reliability increased with sample size, that a minimum sample size of 50 was sufficient for reliability to surpass the 95th percentile of its corresponding empirical null distribution, and that a minimum sample size of 200 – 250 was needed to identify reliable (reliability > 0.7) age-related changes (Figure 5b). These results highlight the robustness of our main age-related effects and indicate that large samples are needed to produce reliable findings.

Amygdala responses are invariant with age while FP node responses decrease with age
We examined main effects of age, emotion, and age by emotion interactions on regional brain activity in each of the ROIs. All results reported here at $q < 0.05$, FDR corrected for multiple comparisons.

Main effect of age. We found significant age-related decreases in cognitive control regions, including bilateral preSMA, dACC and DLPFC, left SPL, right IPL, and right LOFC (Figure 6a-i).

Main effect of emotion. We found a significant emotion effect in all ROIs except bilateral BNST (Figure 6b).

Age by emotion interactions. No age by emotion interactions were found.

Reliability of age-related changes. We performed reliability analysis using the procedure same to that described in regional connectivity section (Figure 1d). We found that reliability increased with sample size, that a minimum sample size of 50 – 100 was sufficient for reliability to surpass the 95th percentile of its corresponding empirical null distribution, and that a minimum sample size of 150 – 200 was needed to identify reliable (reliability > 0.7) age-related changes (Figure 6a-ii). These results highlight the robustness of our main age-related effects and indicate that large samples are needed to produce reliable findings.
Replicability of previous developmental findings related to amygdala circuitry

Given the inconsistencies with regards to developmental changes in amygdala recruitment and connectivity in the extant literature, we further examined age-related changes in the current large sample using precise anatomical coordinates reported in three previously published studies with similar emotion face recognition tasks, experimental contrasts, and age ranges (Table 4) (Gee et al., 2013; Kujawa et al., 2016; Wu et al., 2016).

Amygdala activity. Gee et al. (2013) reported age-related decreases in right amygdala activity during fearful face processing. We first included all five emotion categories in our linear mixed-effect model and did not identify age effect or age by emotion interactions in their reported amygdala region in the current PNC sample ($F$s < 2, $ps > 0.1$). We further examined age-related changes for each of the five emotion categories and still did not identify any age-related changes (Table 5). Kujawa et al. (2016) and Wu et al. (2016) reported no age-related changes in amygdala activity during identification of happy, fearful, and angry faces. When we included all five emotion categories in our linear mixed-effect model, we found no age-related changes or age by emotion interactions ($F$s < 1, $ps > 0.1$). When we included only Fear, Anger and Happy, the three emotion categories used in Kujawa et al. (2016) and Wu et al. (2016) studies, we found significant age-related increases ($F\,(1,1886) = 4.357, p = 0.037, effect size = 0.076$) and age by emotion interactions ($F\,(2, 1514) = 5.804, p = 0.003$) in left amygdala using coordinates reported in Kujawa et al. (2016), and significant age-related increases ($F\,(1,1681) = 4.962, p = 0.026, effect size = 0.081$) in right amygdala.
using coordinates reported in Wu et al. (2016). We further examined age-related changes for each of the five emotion categories. Age-related increases were found in the left amygdala under Fear, Sad and Happy using coordinates reported in Kujawa et al. (2016) and in bilateral amygdala under Fear, Sad, and/or Happy using coordinates reported in Wu et al. (2016) (Table 5). In summary, our replication results from the large sample are partially consistent with Kujawa et al. (2016) and Wu et al. (2016) findings and inconsistent with Gee et al. (2013) finding (Table 7).

Amygdala connectivity. Gee et al. (2013) reported age-related decreases in the right amygdala-vmPFC connectivity during fearful face processing. We included five emotion categories in our linear mixed-effect model and did not identify age effect or age by emotion interactions in their reported amygdala-VMPFC regions in the current PNC sample ($F$s < 1, $p$s > 0.1). We further examined age-related changes for each of the five emotion categories and did not identify any age-related changes (Table 6). Kujawa et al. (2016) reported age-related decreases in bilateral amygdala-dACC connectivity during happy, angry, and fearful face identification in typically developing individuals. When we included five emotion categories in linear mixed-effect model, we found no age-related changes or age by emotion interactions using the reported coordinates ($F$s < 1, $p$s > 0.1). When we included only Fear, Anger and Happy, the three emotion categories used in Kujawa et al. (2016) study, we found a significant age-related increases in right amygdala-dACC connectivity ($F$ (1, 2011) = 8.250, $p$ = 0.004, effect size = 0.104). We further examined age-related changes for each of
the five emotion categories. Age-related increases were found in left amygdala-dACC under Fear and Happy and in right amygdala-dACC under Fear, Anger and Sad (Table 6). Finally, Wu et al. (2016) reported age-related decreases in bilateral amygdala-mPFC/ACC during identification of fearful, angry, and happy faces. When we included all five emotion categories in linear mixed-effect model, we found no age effect or age by emotion interactions ($F < 2, ps > 0.1$). When we included only Fear, Anger and Happy, the three emotion categories used in Wu et al. (2016) study, we found a significant age-related increases in right amygdala connectivity with left mPFC ($F (1, 2041) = 8.046, p = 0.005, effect size = 0.103$) and right ACC ($F (1, 2037) = 6.749, p = 0.009, effect size = 0.094$). We further examined age-related changes for each of the five emotion categories and found age-related increases in bilateral amygdala-mPFC/ACC connectivity under Fear and/or Anger (Table 6).

Amygdala connectivity (gPPI analysis). Because previous developmental studies of amygdala connectivity have used PPI (Gee et al., 2013) or gPPI techniques (Kujawa et al., 2016; Wu et al., 2016), we conducted additional analyses using a similar approach. When we included the five emotion categories in linear mixed-effect model, we did not find a main effect of age, or age by emotion interactions in analyses using coordinates from the three studies ($F < 1.5, ps > 0.1$). When we included only Fear, Anger and Happy in linear mixed-effect model, we did not identify age effect or age by emotion interactions in any of the amygdala connectivity using coordinates from Kujawa et al. (2016) and Wu et al. (2016)
studies ($Fs < 2.2, ps > 0.1$). Finally, we examined age-related changes for each emotion
category. Results revealed no age effect using coordinates reported in Gee et al. (2013) study
($Fs < 2, ps > 0.1$) but significant age-related increases in bilateral amygdala-dACC
connectivity under Fear (left: $F(1, 754) = 5.602$, $p = 0.018$, effect size = 0.088; right: $F(1,
754) = 4.605$, $p = 0.032$, effect size = 0.080) using coordinates reported in Kujawa et al.
(2016) study as well as in left amygdala-mPFC connectivity under Fear ($F(1, 754) = 7.041$, $p$
= 0.008, effect size = 0.099) using coordinates reported in Wu et al. (2016) study. These
results are consistent with those derived using beta time series analysis, as described in the
previous section.

Taken together, these results point to poor replicability of previous developmental findings
on amygdala connectivity associated with emotion processing (Table 7).

Discussion

A large sample of children, adolescents, and young adults with high quality emotion
perception task fMRI data from the PNC (Satterthwaite et al., 2014) afforded us an
unprecedented opportunity to address critical gaps in our knowledge regarding the
development of human emotion identification circuits. We employed a multipronged
analytical strategy to investigate the development of functional circuits associated with
perception and identification of five distinct emotion categories: fear, anger, sadness,
happiness, and neutral. Our analysis disentangled aspects of emotion-related brain circuitry
that were stable over development and those that changed with age. We found a
developmentally-stable modular architecture anchored in the salience, default mode and
central executive networks, with hubs in the mPFC and emotion-related reconfiguration of
the salience network. Developmental changes were most prominent in fronto-parietal circuits
important for salience detection and cognitive control. Crucially, reliability analyses provided
quantitative evidence for the robustness of our findings. Our findings provide a new template
for investigation of emotion processing in the developing brain.

Maturation of emotion identification and specificity with respect to high arousal

negative emotions

Behaviorally, facial emotion identification emerges early in infancy with continued
improvements through childhood and adolescence (Batty and Taylor, 2006; Thomas et al.,
2007; Rodger et al., 2015; Leitzke and Pollak, 2016; Theurel et al., 2016). We found that
although emotion identification accuracy was high for all five emotion categories even in the
youngest individuals (age 8), identification of Fear and Anger emotion categories improved
significantly with age, consistent with previous reports (Gee et al., 2013; Kujawa et al., 2016;
Wu et al., 2016). These findings suggest that high arousal negative (“threat-related”)
emotions show more prominent differentiation, allowing for further sharpening of emotion
concepts with age (Widen, 2013; Nook et al., 2017).

Architecture and developmental stability of emotion circuitry
Network analysis of inter-regional connectivity provided novel evidence for a developmentally stable architecture of emotion-related circuitry. We found three distinct communities (FP, mPFC/PCC, and SPI) that shared similarities with canonical intrinsic salience, central executive, and default mode networks (Menon, 2011), with some key differences.

We found significant reconfiguration of the salience network during emotion identification, including segregation of individual insula subdivisions, amygdala, and dACC, into separate modules. The dorsal AI and the dACC grouped with the FP community of cognitive control systems while the ventral AI, PI, and amygdala grouped with the SPI module of subcortical/limbic systems. This is noteworthy because the dorsal, ventral, anterior, and posterior subdivisions of the insula play distinct and integrative roles in cognitive-affective aspects of emotion perception (Craig, 2009; Menon and Uddin, 2010; Uddin et al., 2014).

Findings provide novel evidence that the overall functional architecture of emotion identification is largely stable across development and different categories of emotion.

**Multiple mPFC regions form a developmentally stable core of emotion circuitry**

Network analysis also revealed developmentally stable hubs in mPFC regions pgACC, vmPFC, and dmPFC across all five emotion categories. These regions play central roles in socio–emotional processing (Phillips et al., 2003; Ochsner et al., 2004; Peelen et al., 2010)

The vmPFC and dmPFC are thought to be important for self-referential processing and
mentalizing (Ochsner et al., 2012; Dixon et al., 2017), while the pgACC is important for awareness of one’s own emotions (Lane et al., 2015; Dixon et al., 2017). Lastly, our results are in line with the theoretical framework that the mPFC plays a central role in emotion categorization by tuning other brain systems to distinct emotion categories (Heberlein et al., 2008; Roy et al., 2012; Barrett and Satpute, 2013; Satpute et al., 2013; Satpute et al., 2016).

Surprisingly, while the PCC was part of the same module as the mPFC, it was not a hub region, despite being consistently identified as hub in intrinsic functional connectivity studies (Buckner et al., 2008; Fransson and Marrelec, 2008; Greicius et al., 2009; Andrews-Hanna et al., 2010). These key differences between intrinsic and task-based network organization argues for the critical importance of characterizing task-dependent brain circuits.

Modular and regional connectivity changes are largely driven by emotion category and are stable with age.

Anchored on a stable modular architecture, we found significant differences in intra- and inter- module and regional connectivity associated with emotion category, and to a lesser degree age. Prominent differences between emotion categories provides evidence that distinct emotions, although anchored in a common modular architecture, may be represented differently in the brain at the network level (Kragel and LaBar, 2015; Saarimaki et al., 2016).

Our SVM-based classification provides evidence of unique brain connectivity patterns.
elicited by each emotion category and indicate differential patterns of distributed network
connectivity patterns involving the same regions (Barrett, 2017).

While module connectivity was largely developmentally stable, two findings of age-related
changes are noteworthy. First, significant age-related increases in FP intra-module
connectivity point to continued maturation of FP control systems underlying emotion
categorization over development (Ahmed et al., 2015; Guyer et al., 2016). Second,
age-related increases in mPFC/PCC intra-module connectivity during identification of fearful
and angry faces suggest that maturation of midline DMN systems underlie continued
developmental improvements in perception of threat and high emotion-arousal emotions.
Lastly, significant developmental changes in regional connectivity centered predominantly
on the insula, a core node of the salience network, highlighting a central and previously
underrecognized role of the insula in the development of emotion perception and
identification.

Developmental changes in amygdala reactivity and amygdala-PFC circuits

Surprisingly, we did not detect developmental changes in the amygdala and its functional
circuitry, given that amygdala and its connectivity with the mPFC have been the focus of
developmental studies of emotion processing, but with contradictory findings (see Table 1
for a summary). To address these discrepancies, we examined amygdala activity and
amygdala-mPFC connectivity in our PNC cohort using specific amygdala seed and mPFC
target regions from three published studies with similar emotion face recognition tasks, experimental contrasts, and age ranges (Gee et al., 2013; Kujawa et al., 2016; Wu et al., 2016). We were unable to replicate the findings of these previous studies, suggesting that small samples and not differences in amygdala and mPFC coordinates are a major contributor to the discrepancies between the current and previous studies. These findings suggest that a narrow focus on amygdala-mPFC connectivity has precluded characterization of other important brain regions, most notably the insula and other PFC regions, underlying the development of emotion perception and categorization.

Further studies are needed to determine the extent to which differences in experimental design contributed to inconsistent patterns of developmental changes in amygdala activity and connectivity reported in previous studies. Critically, addressing this question will require much larger developmental samples than used in previous studies, along with stability and reliability analyses as in the present study.

Neutral faces engage emotion circuitry in the context of emotion identification

Another surprising finding of our study was that neutral faces showed distributed and strong levels of activity and connectivity across the three modules. Disambiguation of neural faces from other emotions would require perception, evaluation, and decision-making and, as our data indicates, engage emotion perception and categorization circuits. Moreover, the classification rates for neutral faces were no different from those between the other emotions.
These findings suggest that previous findings with neutral faces as a control condition in emotion identification tasks should be reassessed and interpreted with caution. This is also an important consideration for future studies of emotion identification.

**Reliability analyses highlight robustness of findings**

Lack of replication arising from small sample sizes is a significant problem in neuroscience research (Button et al., 2013; Szucs and Ioannidis, 2017). We examined the robustness of our findings by assessing reliability of findings at different sample sizes (Schaer et al., 2015; Moser et al., 2018; Turner et al., 2018). We found that sample sizes of about 200 – 300 allowed us to achieve a reliability of 0.70 and higher, providing strong evidence for the role of sample size in generating potentially reproducible findings and that the modest sample sizes of 25-65 used in previous studies have low reliability (< 0.5).

It is also important to note that logistical challenges often preclude individual research groups from collecting large samples of data, which can significantly impede scientific progress (Nee, 2018). The recent curation and sharing of large-scale neurodevelopmental datasets, such as the PNC, provide a promising solution to this challenge and afford the opportunity to begin reconciliation of mixed findings in the extant literature.
Emotion identification and categorization are fundamental to affective and cognitive development and involve integration of multiple widely distributed brain systems. We employed quantitatively rigorous procedures that span multiple levels of regional, circuit, and network analysis and provide robust evidence for shared, unique, stable, and changing neural fingerprints of emotion perception development from childhood to adulthood. We stress the importance of large sample sizes and a systems-wide brain networks approach for understanding the development of human emotion perception.
References


Figure Legends

Figure 1. Schematic view of participant selection procedure and main analysis steps. a Problems, poor brain coverage, or not being able to perform the task. 567 participants were excluded due to incomplete data, medical problems, poor brain coverage, or not being able to perform the task. 567 participants were excluded due to head motion and 12 participants were excluded due to low task accuracy. 759 participants ages 8 – 23 years survived these exclusion criteria. b Brain regions of interest (ROI) involved in emotion perception were identified using meta-analysis implemented in Neurosynth (Yarkoni et al., 2011). They included ventromedial, dorsomedial, ventrolateral and dorsolateral prefrontal cortex (vmPFC, dPFC, vPFC, dPFC), lateral orbitofrontal cortex (IOFC), inferior and superior parietal lobule (IPL, SPL), subgenual, pregenual, dorsal anterior and posterior cingulate cortex (sgACC, pgACC, dACC, PCC), pre-supplementary motor area (preSMA), basolateral amygdala (BLA), ventral and dorsal anterior insula (vAI, dAI), posterior insula (PI), hippocampus (Hipp), and fusiform gyrus (FFG). Subcortical ROIs, including centromedial (CMA) nuclei of the amygdala, bed nucleus stria terminalis (BNST) and nucleus accumbens (NAc) were identified using anatomical atlases (see also Table 3). ROIs were overlaid on a reference brain surface using BrainNet Viewer (Xia et al., 2013). c Network analysis steps used to investigate development of the affective circuits. i. ROI beta series were derived from each task condition (fear, anger, sad, happy, or neutral) for each individual. ii. Functional connectivity matrices were computed for each individual. iii. Functional connectivity matrices were fed into a community detection algorithm to determine community structure of the ROIs. iv. ROI nodal degree was computed over a wide range of sparsity (10% ≤ sparsity ≤ 40% with an interval of 5%) and then an integrated metric of nodal degree was obtained by computing the area under the curve (AUC). v. Intra- and inter-module interactions were computed. vi. Deviation of individual community structure from template community structure were measured as the Jaccard overlap between the two. vii. ROIs were further categorized into hubs or non-hubs and then probability of a ROI to be a hub within each age year was calculated. Permutation test was used to assess hub significance. viii. These resulting brain metrics were further fed into linear mixed-effect models to assess age, emotion, and age×emotion effects. d Reliability analyses assessed the robustness of age-related findings and the effect of sample size. Briefly, we randomly drew N (sample size) participants from the full sample and assessed age-related changes in the subsample. This was repeated 1000 times across a wide range of sample sizes (50 <= N <= 400, increased by 50). Correlation between age effects from a subsample and that from the full sample were calculated and then averaged across subsamples for each sample size. The resulting average correlation was used to measure reliability.

Figure 2. Developmental changes in behavior and reliability. Accuracy The ability to correctly identify high arousal negative (threat-related) emotions (fear and anger) improved with age in both the fMRI (a-i) and full behavioral samples (b-i). Note that only reliability of significant effects were assessed. Reaction Time. Time to identify an emotion decreased with age across all five emotion categories in both the fMRI (a-ii) and full behavioral samples.
Reliability was measured as the probability of observing a significant correlation between accuracy and age or that between reaction time and age. i. *Accuracy*. Reliability of age-accuracy associations increased as sample size increased for conditions that showed significant age-related changes in accuracy in the fMRI (solid line) and the full behavioral (dashed line) samples. ii. *Reaction Time*. Reliability of age-reaction time associations increased as sample size increased for all five emotion categories in the fMRI (solid line) and the full behavioral (dashed line) samples. In general, a sample size of 300 – 350 was needed to identify reliable (reliability > 0.7) age-accuracy association and that of 200 was needed to identify reliable age-reaction time association.

**Figure 3. Main effects of age and emotion on modular organization and connectivity.** a i. Three functional modules were identified across all participants: the frontoparietal (FP) module (red), medial prefrontal cortex/posterior cingulate cortex (mPFC/PCC) module (green), and subcortical/posterior insula (SPl) module (yellow). ii. As shown in the group consistency matrix, the FP module (red border) includes bilateral dorsolateral and ventrolateral prefrontal cortex (dLPFC, vLPFC), dorsal anterior insula (dACC), dorsal anterior cingulate cortex (dACC), pre-supplementary motor area (preSMA), inferior and superior parietal lobule (IPL, SPL), and fusiform gyrus (FFG). The mPFC/PCC module (blue border) includes bilateral nucleus accumbens (NAc), subgenual and pregenual anterior cingulate (sgACC, pgACC), posterior cingulate cortex (PCC), and ventromedial and dorsomedial prefrontal cortex (vmPFC, dmPFC). The SPl module (yellow border) includes bilateral basolateral and centromedial amygdala (BLA, CMA), ventral anterior and posterior insula (vAI, pI), bed nucleus stria terminalis (BNST), and hippocampus (Hipp). The color bar indicates the probability of two nodes being classified in the same module across participants. b As shown in the group consistency matrices, overall modular structure is stable across all five emotion categories and age groups (children: 8-12 y/o, adolescents: 13-17 y/o, and adults: 18-23 y/o). c Intra-FP connectivity and FP-SPl connectivity increased with age; Intra-FP, intra-SPl, and all inter-module connectivity measures varied across emotions; age-related changes in intra-mPFC/PCC connectivity was modulated by emotion category. d Intra- and inter-module connectivity varied across emotions. Data are represented as mean ± SEM.

**Figure 4. Developmentally-stable hubs in medial prefrontal cortex.** a ROI nodal degree for each emotion category. Node size is in proportion to integrated nodal degree measure. b Probability measures averaged across all regions within each module at each age year. Overall, the medial prefrontal cortex/posterior cingulate cortex (mPFC/PCC) module has a higher probability to be a functional hub across age than the frontoparietal (FP) and subcortical/posterior insula (SPl) modules. Data are represented as mean ± SEM. c Multiple regions of the mPFC/PCC module, including the pregenual anterior cingulate cortex (pgACC), ventral medial prefrontal cortex (vmPFC), and dorsal medial prefrontal cortex (dmPFC) of the mPFC/PCC module were identified as functional hub regions for all five emotion categories and were stable over development. Here, probability matrices are masked.
by the 95th percentile of empirical null distribution at each region and age year, with
probability values less than or equal to the 95th percentile set as 0. An age year includes all
participants whose age was larger than or equal to the current age year and less than the next
age year (e.g., Age 8 includes all participants with age >= 8 and < 9).

Figure 5. Main effects of age and emotion on regional connectivity. a Connectivity within
FP and SPI modules, between FP and SPI modules and between mPFC/PCC and SPI modules
increased with age, including left dPFC connectivity with right dAI, right vAI connectivity
with right sgACC and bilateral PI, right BNST connectivity with right IPL and SPL, and right
BLA connectivity with left PI. Connectivity within mPFC/PCC and SPI modules decreased
with age, including right sgACC connectivity with left vmPFC and left CMA connectivity
with right CMA and left PI. Emotion effects are widespread, especially within FP module,
between FP and mPFC/PCC modules, between FP and SPI modules, and between
mPFC/PCC and SPI modules. b Reliability of age-related changes in regional connectivity
across all brain regions increased with sample size. A sample size of 50 is sufficient for
reliability to surpass the 95th percentile of its corresponding empirical null distribution
(dashed line), and a minimum sample size of 200 – 250 was needed to identify reliable
age-related changes. c Classification analysis showed that it was sufficient to discriminate all
pairs of emotion categories using all inter-regional links, with classification accuracy ranging
from 56.2% to 67.2% and significantly higher than chance level as assessed by permutation
test.

Figure 6. Main effects of age and emotion on regional activity. a i. Brain regions showing
age-related changes. Activity in cognitive control regions, including bilateral preSMA,
dACC and DLPFC, left SPM, right IPL, and right LOFC, decreased with age. Node size was
scaled to present effect size of age-related decrease. ii. Reliability of age-related changes
across all brain regions. A minimum sample size of 50 – 100 was needed for reliability to
surpass the 95th percentile of its corresponding empirical null distribution (dashed line), and a
minimum sample size of 50 – 100 was needed to identify reliable (correlation > 0.7)
age-related changes. b i. Brain regions showing emotion effects. Node size was scaled to
present F value of emotion effect. ii. Activity in all brain regions except bilateral BNST
varied across emotion category.
a Participant selection procedure

| Complete behavioral and fMRI data AND | Mean scan-to-scan Disp < 0.2 mm AND | Age distribution (N=759) |
| No significant or major medical problems AND | Max scan-to-scan Disp < 1 mm AND | Count |
| Good brain coverage AND | Max Disp < 3 mm? | 80 |
| At least made one correct response? | Overall accuracy > 70%? | 40 |

- Excluded
  - NO
  - YES
  - Excluded

N = 1,338

N = 771

N = 759

b Emotion-related brain regions identified using meta-analysis

![Brain images with marked regions]

c Network analysis

i ROI beta series

ii Functional connectivity matrix

iii Community detection

iv Nodal degree

 Anger

 Sad

 Happy

Regional connectivity

v Intra- and inter-module interactions

vi Jaccard overlap between individual and template community structure

vii Probability of region being a hub within each age year (permutation used to assess significance)

viii Linear mixed-effect model used to assess Age, Emotion, and Age x Emotion effects

d Reliability analysis

Sample size = N

(50 <= N <= 400, Increased by 50)

Sample 1

... Sample 500

... Sample 1000

Age effects derived from the full sample

Correlation

Average correlation (Reliability measure)

Effect of sample size on reliability

![Graph showing reliability increase with sample size]
a Age-related changes in emotion identification: accuracy and reaction time (N = 759)

b Age-related changes in emotion identification: accuracy and reaction time (N = 1338)

c Effect of sample size on reliability of emotion identification: accuracy and reaction time
a Modular structure of emotion identification circuits

- Frontoparietal (FP) Module
- mPFC/PCC Module
- Subcortical/Posterior Insula (SPIN) Module

b Modular structure for each emotion category over development

- Neutral
- Fear
- Anger
- Sad
- Happy

All participants (8 - 23 y)  Children (8 - 12 y)  Adolescents (13 - 17 y)  Adults (18 - 23 y)

c Main effects of age, emotion, and age by emotion interactions on intra- and inter-modular connectivity

- Age
- Emotion
- Age x Emotion

d Intra- and inter-module connectivity across emotions

- Connectivity Strength
- mPFC/PCC
- FP
- SPI
- FP-SPI
- FP-mPFC/PCC
- SPI-mPFC/PCC-SPI
- Neutral
- Fear
- Anger
- Sad
- Happy
a Nodal degree of circuit connectivity for each emotion category

Neutral  Fear  Anger  Sad  Happy

b Hub probability for each emotion category over development

Neutral  Fear  Anger  Sad  Happy

Age (year)

Probability

FP  mPFC/PCC  SPI

c Developmentally stable hubs in mPFC/PCC

Neutral  Fear  Anger  Sad  Happy

Age (year)

Probability
a Main effect of age and emotion on inter-regional connectivity (FDR corrected)

b Effect of sample size on reliability of age-related changes

c Classification accuracy for differentiating connectivity patterns of emotions
a Age-related changes in regional activity and reliability

i

Left hemisphere

preSMA  dACC  SPL

Right hemisphere

preSMA  dACC  IPL  DLPFC  LOFC

ii

Reliability

Sample size

Average observed
95th percentile of null distribution

b Effect of emotion on regional activity

i

L

R

Frontoparietal (FP) Module
mPFC/PCC Module
Subcortical/Posterior Insula (SPI) Module

ii

Activity (left hemisphere)

Activity (right hemisphere)

Neutral  Fear  Anger  Sad  Happy
### Tables

**Table 1. Studies that examined age-related changes in amygdala activity and connectivity underlying emotion perception.**

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Age (years)</th>
<th>Contrast and #trials</th>
<th>Task</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gee et al., 2013</td>
<td>45 (19F)</td>
<td>4 – 22</td>
<td>Fearful face vs. Fixation (#trials = 24)</td>
<td>Participants viewed fearful faces with interspersed with neutral faces in one run, and viewed happy faces interspersed with neutral faces in the other run</td>
<td>Amygdala activity decreased with age; amygdala-vmPFC connectivity decreased with age</td>
</tr>
<tr>
<td>Kujawa et al., 2016</td>
<td>61 (35F) TD</td>
<td>7 – 25</td>
<td>Emotional faces (happy, angry or fearful) vs. Shapes (#trials per emotion category = 12)</td>
<td>Participants were instructed to match emotion (happy, angry or fearful) of the target face in face blocks and match shapes in shape blocks</td>
<td>No age-related effect in amygdala activity; amygdala-dACC connectivity decreased with age in TD but increased with age in AD for all emotions</td>
</tr>
<tr>
<td>Wu et al., 2016</td>
<td>61 (35F)</td>
<td>7 – 25</td>
<td>Emotional faces (angry, fearful or happy) vs. Shapes (#trials per emotion category = 12)</td>
<td>Participants were instructed to match emotion (angry, fearful or happy) of the target face in face blocks and match shapes in shape blocks</td>
<td>No age-related effect in amygdala activity; amygdala-ACC/mPFC connectivity decreased with age for all emotions</td>
</tr>
<tr>
<td>Vink et al., 2014</td>
<td>60 (28F)</td>
<td>10 – 24</td>
<td>Positive vs. Neutral (#trials = 32), Negative vs. Neutral (#trials = 32)</td>
<td>Participants viewed each picture selected from the IAPS and label the valence (negative, neutral, or positive)</td>
<td>Amygdala activity decreased with age; amygdala-mOFC connectivity increased with age</td>
</tr>
</tbody>
</table>
Wolf & Herringa, 2016

<table>
<thead>
<tr>
<th>TD</th>
<th>24 (13F)</th>
<th>8 – 18</th>
<th>Threat vs. Neutral</th>
<th>Participants viewed threat and neutral images selected from the IAPS and label the valence</th>
<th>No age-related effect in amygdala activity; amygdala-mOFC connectivity increased with age in TD but decreased with age in PTSD</th>
</tr>
</thead>
</table>

Note, if not specified, healthy individuals were recruited in the listed studies. Functional connectivity was measured by PPI in Gee et al. (2013), Vink et al. (2014) and Wolf & Herringa (2016) and by gPPI in Kujawa et al. (2016) and Wu et al. (2016). AD = Anxiety Disorders, gPPI = Generalized Psychophysiological Interaction, IAPS = International Affective Picture System, PPI = Psychophysiological Interaction, PTSD = Posttraumatic Stress Disorder, TD = Typically Developing. See also Table 4 for MNI coordinates of amygdala and prefrontal cortex regions used in these studies, and Tables 5, 6, and 7 for results of replication analysis.
Table 2. Participant characteristics and task performance.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean ± SD (n = 759)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td>16.20 ± 3.27</td>
</tr>
<tr>
<td>Sex (female/male)</td>
<td>419/340</td>
</tr>
<tr>
<td>Race (Caucasian/Other)</td>
<td>408/351</td>
</tr>
<tr>
<td>Motion (mm)</td>
<td>0.073 ± 0.03</td>
</tr>
<tr>
<td>Fluid Intelligence</td>
<td>12.62 ± 3.97</td>
</tr>
<tr>
<td>Fear Accuracy</td>
<td>0.89 ± 0.11</td>
</tr>
<tr>
<td>Anger Accuracy</td>
<td>0.92 ± 0.09</td>
</tr>
<tr>
<td>Sad Accuracy</td>
<td>0.88 ± 0.11</td>
</tr>
<tr>
<td>Neutral Accuracy</td>
<td>0.91 ± 0.09</td>
</tr>
<tr>
<td>Happy Accuracy</td>
<td>0.99 ± 0.03</td>
</tr>
<tr>
<td>Fear RT (sec)</td>
<td>2.65 ± 0.46</td>
</tr>
<tr>
<td>Anger RT (sec)</td>
<td>2.48 ± 0.44</td>
</tr>
<tr>
<td>Sad RT (sec)</td>
<td>2.68 ± 0.46</td>
</tr>
<tr>
<td>Happy RT (sec)</td>
<td>2.09 ± 0.40</td>
</tr>
<tr>
<td>Neutral RT (sec)</td>
<td>2.34 ± 0.44</td>
</tr>
</tbody>
</table>

Fluid Intelligence is measured by Penn Matrix Reasoning Test scores (i.e., total correct responses for all test trials). RT = median Response Time. See also Figure 1a for age distribution.
<table>
<thead>
<tr>
<th>Region/Abbreviation</th>
<th>MNI Coordinates</th>
<th>Brodmann Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left basolateral amygdala (BLA.L)</td>
<td>-20  -6  -20</td>
<td></td>
</tr>
<tr>
<td>Right basolateral amygdala (BLA.R)</td>
<td>26   -4  -20</td>
<td></td>
</tr>
<tr>
<td>Left centromedial amygdala (CMA.L)</td>
<td>-24  -9   9</td>
<td></td>
</tr>
<tr>
<td>Right centromedial amygdala (CMA.R)</td>
<td>27   -10  -9</td>
<td></td>
</tr>
<tr>
<td>Left subgenual anterior cingulate cortex (sgACC.L)</td>
<td>-2   14   -16</td>
<td>BA 25</td>
</tr>
<tr>
<td>Right subgenual anterior cingulate cortex (sgACC.R)</td>
<td>6    20   -12</td>
<td>BA 25</td>
</tr>
<tr>
<td>Left pregenual anterior cingulate cortex (pgACC.L)</td>
<td>-2   42   6</td>
<td>BA 32</td>
</tr>
<tr>
<td>Right pregenual anterior cingulate cortex (pgACC.R)</td>
<td>2    42   8</td>
<td>BA 32</td>
</tr>
<tr>
<td>Left dorsal anterior cingulate cortex (dACC.L)</td>
<td>-4   22   30</td>
<td>BA 32</td>
</tr>
<tr>
<td>Right dorsal anterior cingulate cortex (dACC.R)</td>
<td>4    22   30</td>
<td>BA 32</td>
</tr>
<tr>
<td>Left posterior cingulate cortex (PCC.L)</td>
<td>-2   -54  28</td>
<td>BA 31</td>
</tr>
<tr>
<td>Right posterior cingulate cortex (PCC.R)</td>
<td>4    -54  28</td>
<td>BA 31</td>
</tr>
<tr>
<td>Left ventromedial prefrontal cortex (vmPFC.L)</td>
<td>-8   50   -4</td>
<td>BA 10</td>
</tr>
<tr>
<td>Right ventromedial prefrontal cortex (vmPFC.R)</td>
<td>8    50   -4</td>
<td>BA 10</td>
</tr>
<tr>
<td>Left ventromedial prefrontal cortex (vmPFC.L)</td>
<td>-2   50   -12</td>
<td>BA 32</td>
</tr>
<tr>
<td>Right ventromedial prefrontal cortex (vmPFC.R)</td>
<td>4    50   -12</td>
<td>BA 32</td>
</tr>
<tr>
<td>Left ventromedial prefrontal cortex (vmPFC.L)</td>
<td>-4   38   -10</td>
<td>BA 32</td>
</tr>
<tr>
<td>Right ventromedial prefrontal cortex (vmPFC.R)</td>
<td>4    38   -10</td>
<td>BA 32</td>
</tr>
<tr>
<td>Left dorsomedial prefrontal cortex (dmPFC.L)</td>
<td>-2   58   20</td>
<td>BA 9</td>
</tr>
<tr>
<td>Right dorsomedial prefrontal cortex (dmPFC.R)</td>
<td>2    58   20</td>
<td>BA 9</td>
</tr>
<tr>
<td>Left pre-supplementary motor area (preSMA.L)</td>
<td>-2   18   46</td>
<td>BA 6</td>
</tr>
<tr>
<td>Right pre-supplementary motor area (preSMA.R)</td>
<td>6    24   38</td>
<td>BA 32</td>
</tr>
<tr>
<td>Left lateral orbitofrontal cortex (IOFC.L)</td>
<td>-46  26   -6</td>
<td>BA 47</td>
</tr>
<tr>
<td>Right lateral orbitofrontal cortex (IOFC.R)</td>
<td>46   26   -4</td>
<td>BA 47</td>
</tr>
<tr>
<td>Left lateral orbitofrontal cortex (IOFC.L)</td>
<td>-28  18   -16</td>
<td>BA 47</td>
</tr>
<tr>
<td>Right lateral orbitofrontal cortex (IOFC.R)</td>
<td>26   22   -18</td>
<td>BA 47</td>
</tr>
<tr>
<td>Left ventrolateral prefrontal cortex (vLPFC.L)</td>
<td>-48  28   10</td>
<td>BA 45</td>
</tr>
<tr>
<td>Right ventrolateral prefrontal cortex (vLPFC.R)</td>
<td>48   28   10</td>
<td>BA 45</td>
</tr>
<tr>
<td>Left dorsolateral prefrontal cortex (dLPFC.L)</td>
<td>-46  8    30</td>
<td>BA 9</td>
</tr>
<tr>
<td>Right dorsolateral prefrontal cortex (dLPFC.R)</td>
<td>46   8    30</td>
<td>BA 9</td>
</tr>
<tr>
<td>Left dorsolateral prefrontal cortex (dLPFC.L)</td>
<td>-41  22   43</td>
<td>BA 8</td>
</tr>
<tr>
<td>Right dorsolateral prefrontal cortex (dLPFC.R)</td>
<td>41   22   43</td>
<td>BA 8</td>
</tr>
<tr>
<td>Left inferior parietal lobule (IPL.L)</td>
<td>-54  -50  44</td>
<td>BA 40</td>
</tr>
<tr>
<td>Right inferior parietal lobule (IPL.R)</td>
<td>46   -48  44</td>
<td>BA 40</td>
</tr>
<tr>
<td>Region</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Left superior parietal lobule (SPL.L)</td>
<td>-30</td>
<td>-56</td>
</tr>
<tr>
<td>Right superior parietal lobule (SPL.R)</td>
<td>38</td>
<td>-50</td>
</tr>
<tr>
<td>Left fusiform gyrus (FFG.L)</td>
<td>-42</td>
<td>-48</td>
</tr>
<tr>
<td>Right fusiform gyrus (FFG.R)</td>
<td>42</td>
<td>-48</td>
</tr>
<tr>
<td>Left dorsal anterior insula (dAI.L)</td>
<td>-34</td>
<td>22</td>
</tr>
<tr>
<td>Right dorsal anterior insula (dAI.R)</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>Left ventral anterior insula (vAI.L)</td>
<td>-38</td>
<td>-2</td>
</tr>
<tr>
<td>Right ventral anterior insula (vAI.R)</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>Left posterior insula (PII.L)</td>
<td>-38</td>
<td>-18</td>
</tr>
<tr>
<td>Right posterior insula (PII.R)</td>
<td>38</td>
<td>-14</td>
</tr>
<tr>
<td>Left bed nucleus stria terminalis (BNST.L)</td>
<td>-6</td>
<td>2</td>
</tr>
<tr>
<td>Right bed nucleus stria terminalis (BNST.R)</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Left nucleus accumbens (NAc.L)</td>
<td>-10</td>
<td>12</td>
</tr>
<tr>
<td>Right nucleus accumbens (NAc.R)</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Left hippocampus (Hipp.L)</td>
<td>-34</td>
<td>-18</td>
</tr>
<tr>
<td>Right hippocampus (Hipp.R)</td>
<td>32</td>
<td>-14</td>
</tr>
</tbody>
</table>
Table 4. MNI Coordinates of amygdala and prefrontal cortex regions used in previous developmental studies of face emotion perception.

<table>
<thead>
<tr>
<th>Study</th>
<th>Amygdala [x, y, z]</th>
<th>Connectivity Seed [x, y, z] and target [x, y, z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gee et al. 2013</td>
<td>[32, -1, -16]</td>
<td>Right amygdala [32, -1, -16] and right vmPFC [2, 32, 8]</td>
</tr>
<tr>
<td>Kujawa et al., 2016*</td>
<td>[-20, -2, -20],</td>
<td>Left amygdala (AAL) and left dACC [-4, 30, 16]</td>
</tr>
<tr>
<td></td>
<td>[24, -2, -22]</td>
<td>Right amygdala (AAL) and right dACC [2, 34, 14]</td>
</tr>
<tr>
<td>Wu et al., 2016*</td>
<td>Left and right amygdala from AAL atlas</td>
<td>Left amygdala (AAL) and left ACC/mPFC [-6, 34, 16]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right amygdala (AAL) and left ACC/mPFC [-4, 36, 14]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left amygdala (AAL) and left ACC [-8, 28, 18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right amygdala (AAL) and right ACC [6, 36, 12]</td>
</tr>
</tbody>
</table>

AAL = Automated Anatomical Labeling. *While no age-related effects were revealed in amygdala activity in Kujawa et al. (2016) and Wu et al. (2016), we created amygdala ROI based on coordinates showing significant main effect of emotion category in Kujawa et al. (2016) or used amygdala AAL atlas, which was a seed region in connectivity analysis in Wu et al. (2016), to examine age-related effects of amygdala activity in our PNC cohort.
Table 5. Age-related changes in amygdala activity in the PNC sample using coordinates from previous studies.

<table>
<thead>
<tr>
<th>Region</th>
<th>Emotion</th>
<th>F (1, 754)</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Amygdala in Gee et al. (2013) ↓</td>
<td>Neutral</td>
<td>1.121</td>
<td>0.290</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>0.228</td>
<td>0.633</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>0.174</td>
<td>0.677</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>1.965</td>
<td>0.161</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>0.109</td>
<td>0.741</td>
<td>-0.012</td>
</tr>
<tr>
<td>Left Amygdala in Kujawa et al. (2016) ←</td>
<td>Neutral</td>
<td>0.953</td>
<td>0.329</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>12.617</td>
<td>0.0004</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>1.844</td>
<td>0.175</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>15.140</td>
<td>0.0001</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>9.020</td>
<td>0.003</td>
<td>0.111</td>
</tr>
<tr>
<td>Right Amygdala in Kujawa et al. (2016) ←</td>
<td>Neutral</td>
<td>0.566</td>
<td>0.452</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>2.650</td>
<td>0.104</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>0.123</td>
<td>0.726</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>2.801</td>
<td>0.095</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>0.726</td>
<td>0.394</td>
<td>0.032</td>
</tr>
<tr>
<td>Left Amygdala (AAL) in Wu et al. (2016) ←</td>
<td>Neutral</td>
<td>1.583</td>
<td>0.209</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>15.796</td>
<td>7.73e-05</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>4.571</td>
<td>0.033</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>20.755</td>
<td>6.08e-06</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>10.009</td>
<td>0.002</td>
<td>0.117</td>
</tr>
<tr>
<td>Right Amygdala (AAL) in Wu et al. (2016) ←</td>
<td>Neutral</td>
<td>0.277</td>
<td>0.599</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>5.741</td>
<td>0.017</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>1.946</td>
<td>0.163</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>5.784</td>
<td>0.016</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>1.553</td>
<td>0.213</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Effect size is measured by the standardized partial coefficient of the age term. AAL = Automated Anatomical Labeling. AAL indicates that the amygdala ROI was derived based on AAL atlas. ↓ indicates that developmental decreases in amygdala with mPFC was reported in the cited study. ← indicates no significant developmental changes were found in the cited study.
Table 6. Age-related changes in amygdala connectivity in the PNC sample using coordinates from previous studies.

<table>
<thead>
<tr>
<th>Connectivity</th>
<th>Emotion</th>
<th>F (I, 754)</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMY.R-vmPFC.R in Gee et al. (2013)</td>
<td>Neutral</td>
<td>0.210</td>
<td>0.647</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>3.497</td>
<td>0.062</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>1.880</td>
<td>0.171</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>0.963</td>
<td>0.327</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>0.288</td>
<td>0.591</td>
<td>0.020</td>
</tr>
<tr>
<td>AMY.L-dACC.L in Kujawa et al. (2016)</td>
<td>Neutral</td>
<td>0.008</td>
<td>0.929</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>8.921</td>
<td>0.003</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>2.569</td>
<td>0.109</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>0.388</td>
<td>0.533</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>4.596</td>
<td>0.032</td>
<td>0.079</td>
</tr>
<tr>
<td>AMY.R-dACC.R in Kujawa et al. (2016)</td>
<td>Neutral</td>
<td>0.214</td>
<td>0.644</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>6.602</td>
<td>0.010</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>7.661</td>
<td>0.006</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>4.420</td>
<td>0.036</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>3.232</td>
<td>0.073</td>
<td>0.066</td>
</tr>
<tr>
<td>AMY.L-mPFC.L in Wu et al. (2016)</td>
<td>Neutral</td>
<td>0.012</td>
<td>0.914</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>4.139</td>
<td>0.042</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>1.198</td>
<td>0.274</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>0.003</td>
<td>0.958</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>1.199</td>
<td>0.274</td>
<td>0.040</td>
</tr>
<tr>
<td>AMY.R-mPFC.L in Wu et al. (2016)</td>
<td>Neutral</td>
<td>0.334</td>
<td>0.563</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>3.712</td>
<td>0.054</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>7.566</td>
<td>0.006</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>1.333</td>
<td>0.249</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>2.743</td>
<td>0.098</td>
<td>0.061</td>
</tr>
<tr>
<td>AMY.L-ACC.L in Wu et al. (2016)</td>
<td>Neutral</td>
<td>0.019</td>
<td>0.890</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>4.351</td>
<td>0.037</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>3.000</td>
<td>0.084</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>0.602</td>
<td>0.438</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>2.769</td>
<td>0.097</td>
<td>0.061</td>
</tr>
<tr>
<td>AMY.R-ACC.R in Wu et al. (2016)</td>
<td>Neutral</td>
<td>0.013</td>
<td>0.909</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>4.551</td>
<td>0.033</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>Effect Size</td>
<td>p-value</td>
<td>Significance</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>---------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>5.901</td>
<td>0.015</td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>4.496</td>
<td>0.034</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>1.261</td>
<td>0.262</td>
<td>0.041</td>
<td></td>
</tr>
</tbody>
</table>

Effect size is measured by the standardized partial coefficient of age term. ↓ or ↑ indicates that developmental decrease or increase in amygdala-mPFC connectivity was reported in the cited study respectively. — indicates no significant developmental changes in amygdala-mPFC connectivity was found in the cited study.
<table>
<thead>
<tr>
<th>Study</th>
<th>Findings</th>
<th>Replicability amygdala activity</th>
<th>Replicability amygdala connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gee et al., 2013</td>
<td>Amygdala activity decreased with age; amygdala-vmPFC connectivity decreased with age</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kujawa et al., 2016</td>
<td>No age-related effect in amygdala activity; amygdala-dACC connectivity decreased with age in TD but increased with age in AD for all emotions</td>
<td>Partially Yes</td>
<td>No</td>
</tr>
<tr>
<td>Wu et al., 2016</td>
<td>No age-related effect in amygdala activity; amygdala-ACC/mPFC connectivity decreased with age for all emotions</td>
<td>Partially Yes</td>
<td>No</td>
</tr>
</tbody>
</table>