JNeurosci Accepted Manuscript

JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

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.

Copyright © 2019 the authors

JNeurosci Accepted Manuscript

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[1], Aarthi Padmanabhan[1], James J. Gross[2], Vinod Menon[1,3,4*]*

[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

JNeurosci Accepted Manuscript

34 thank the PNC consortium for sharing their data, and Drs. Ted Satterthwaite, Daniel Wolf,

35 and Monica Calkins for graciously replying to our queries during the early stages of the study.

JNeurosci Accepted Manuscript

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.

JNeurosci Accepted Manuscript

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.

JNeurosci Accepted Manuscript

68 **Introduction**

69 Identification of distinct facial expressions, which is fundamental to recognizing the
70 emotional state of others, begins in infancy and continues to develop throughout childhood
71 and adolescence (Batty and Taylor, 2006; Thomas et al., 2007; Rodger et al., 2015; Leitzke
72 and Pollak, 2016; Theurel et al., 2016). How emotions are identified and categorized in the
73 developing brain remains a fundamental open question. Analysis of task-evoked brain
74 circuits using a 'Big Data' approach has the potential to accelerate foundational knowledge of
75 emotion perception circuit development, and provide robust templates for understanding
76 miswiring of functional circuits in developmental psychopathologies. Here, we leverage the
77 Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2014) to address
78 critical gaps in our knowledge regarding the functional circuits of emotion perception and
79 their development.
80
81 Meta-analyses of neuroimaging studies in adults have identified widely-distributed, and
82 overlapping functional systems engaged by distinct emotion categories (Kober et al., 2008;
83 Fusar-Poli et al., 2009). In particular, brain regions associated with three large-scale
84 functional brain networks are relevant for emotion processing based on their extensive role in
85 integrating salient cues across cortical and limbic systems. The 'salience network' facilitates
86 rapid detection of personally relevant and salient emotional cues in the environment (Seeley
87 et al., 2007; Menon and Uddin, 2010; Menon, 2011), the 'default mode network' is important
88 for mentalizing and inferring the emotional states of others (Greicius et al., 2003; Blakemore,

JNeurosci Accepted Manuscript

89 2008), and the frontoparietal 'central executive network' is involved in attention, reappraisal,
90 and regulation of reactivity to emotional stimuli (Dosenbach et al., 2007; Gyurak et al., 2011;
91 Tupak et al., 2014; Braunstein et al., 2017; Sperduti et al., 2017). However, previous
92 neurodevelopmental studies of emotion perception have focused almost exclusively on the
93 amygdala and its connectivity with the medial prefrontal cortex (mPFC) (Gee et al., 2013;
94 Vink et al., 2014; Kujawa et al., 2016; Wolf and Herringa, 2016; Wu et al., 2016), with
95 inconsistent findings (see **Table 1** for a summary). While some studies found no age-related
96 changes (Kujawa et al., 2016; Wolf and Herringa, 2016; Wu et al., 2016), others have
97 reported increases (Vink et al., 2014; Wolf and Herringa, 2016) or decreases with age (Gee et
98 al., 2013; Kujawa et al., 2016; Wu et al., 2016) in amygdala activation or connectivity with
99 the mPFC. In contrast, intrinsic functional connectivity studies have highlighted protracted
100 developmental changes in distributed brain networks (Power et al., 2010; Menon, 2013),
101 including developmental changes in amygdala connectivity with subcortical, paralimbic, and
102 limbic structures, polymodal association, and ventromedial prefrontal cortex (Qin et al.,
103 2012). These findings highlight a critical need for characterizing the development of
104 task-evoked functional brain circuits and network organization underlying emotion
105 perception (Barrett and Satpute, 2013).
106
107 Here, we characterize the neurodevelopment of emotion perception circuits leveraging
108 high-quality data (N=759) from the emotion-identification fMRI task in the PNC, an
109 unprecedented "Big Data" initiative for charting brain development in a large community

JNeurosci Accepted Manuscript

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 displace > 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

JNeurosci Accepted Manuscript

sample) survived these exclusion criteria and were included in further behavioral and neuroimaging 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

JNeurosci Accepted Manuscript

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 mm^2, 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.

JNeurosci Accepted Manuscript

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, vlPFC, vlPFC), 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), 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

JNeurosci Accepted Manuscript

194 nodes in the ventral anterior insula (vAI), centromedial amygdala (CMA), bed nucleus stria
195 terminalis (BNST) and nucleus accumbens (NAc) based on their well-known roles in
196 perception of negative and positive emotions (LeDoux, 2007; Deen et al., 2011; Chang et al.,
197 2013; Floresco, 2015; Lebow and Chen, 2016; Namkung et al., 2017). The vAI node was
198 identified using Neurosynth and search term "anxiety" (Paulus and Stein, 2006; Etkin, 2010;
199 Menon, 2011; Blackford and Pine, 2012; Qin et al., 2014), the CMA was identified based on
200 its cytoarchitectonically-defined probabilistic map using the Anatomy toolbox (Eickhoff et
201 al., 2005), the BNST node was based on the meta-analysis by Avery et al. (2014), and NAc
202 was identified based on the Harvard-Oxford atlas distributed with FSL
203 (http://www.fmrib.ox.ac.uk/fsl/). These procedures identified 50 ROIs consistently
204 associated with emotion processing (**Figure 1b and Table 3**). Finally, we further verified
205 that the activation maps of emotion-related topics encompassed the maps reported in a large
206 meta-analytic database of fMRI studies (Rubin et al., 2017).
207
208 *Emotion-related node-level activity.* Task-related brain activation was identified using a
209 general linear model (GLM) implemented in SPM8. At individual level, brain activation
210 representing correct trials for each emotion task condition (neutral, fear, anger, sad, or happy)
211 was modeled using boxcar functions convolved with a canonical hemodynamic response
212 function and a temporal derivative to account for voxelwise latency differences in
213 hemodynamic response. An error regressor was also included in the model to account for the
214 influence of incorrect trials. Additionally, six head movement parameters generated from the

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

236 modules (Lancichinetti and Fortunato, 2012; Dwyer et al., 2014) associated with each
237 emotion. In each participant, the Louvain algorithm (Blondel et al., 2008), as implemented in
238 the Brain Connectivity Toolbox (Rubinov and Sporns, 2010), was used to first determine the
239 modular structure of the signed, weighted connectivity matrix (Newman, 2004, 2006). To
240 handle potential degeneracy of community assignments, we repeated this procedure 1000
241 times in each participant (Lancichinetti and Fortunato, 2012; Dwyer et al., 2014). In each
242 iteration, we generated a co-classification matrix in which each element contained 1 if two
243 nodes were part of the same module and 0 otherwise. The resulting 1000 co-classification
244 matrices were averaged to generate a co-occurrence matrix indicating the probability of two
245 nodes being in the same module across 1000 iterations. Co-occurrence matrices were then
246 averaged across participants to produce a positive, weighted group consistency matrix, with
247 higher weight in elements of this matrix indicating that two nodes were more frequently
248 classified in the same module across participants. This group consistency matrix was used to
249 determine a consensus partition (Lancichinetti and Fortunato, 2012). To evaluate the
250 influence of resolution parameter Gamma, which tunes the size and number of communities
251 obtained from modularity maximization (Fortunato and Barthelemy, 2007; Sporns and
252 Betzel, 2016), the above consensus-based approach was repeated at Gamma values ranging
253 from 0.1 to 2.0 in increments of 0.1.
254
255 To assess the significance of the modular structure we compared the actual Q to the Q derived
256 from null networks. Specifically, the group consistency matrix was used to create 1000 null

JNeurosci Accepted Manuscript

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% ≤ sparsity ≤ 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.

JNeurosci Accepted Manuscript

277 *Generalization of ROIs to the Brainnetome atlas.* We used the Brainnetome atlas to
278 determine robustness and generalizability of findings from the 50 emotion-related
279 metanalysis-derived ROIs. An advantage of the Brainnetome atlas, compared to other many
280 other functional brain parcellations, is that it includes both cortical and subcortical clusters
281 with appropriate anatomical labels (Fan et al., 2016). Specifically, we identified Brainnetome
282 clusters in which each of the 50 ROIs was located; because 2 ROI pairs were located in the
283 same Brainnetome cluster, this procedure yielded 46 matching clusters. We then repeated all
284 our analyses using these spatially extended clusters.
285
286 **Statistical analysis**
287 *Linear vs. Nonlinear age models*. The remainder of our analyses involved assessing
288 age-related changes in task performance, regional activity, regional connectivity, and intra-
289 and inter-module connectivity. For task performance, we analyzed both the fMRI (N= 759)
290 and behavioral (N=1338) samples. We examined three models – linear, quadratic, and cubic
291 spline – to test for linear and nonlinear age-related effects, and used Bayesian Information
292 Criteria (BIC) to determine the best model. The vast majority of the analyses indicated that
293 non-linear models did not fit the data better than linear models. Where appropriate we report
294 nonlinear age effects that survived FDR correction for multiple comparisons.
295
296 *Linear mixed-effect model.* To examine main effects of age, emotion, and age by emotion
297 interactions on task performance and brain measures, we used R package lme4 (Bates et al.,

JNeurosci Accepted Manuscript

2015) to perform a linear mixed-effect analysis. As fixed effects, we entered age, emotion, and age by emotion interactions (regressors of interest) as well as sex, race, and head movement (regressors of no interest) into the model. As random effects, we had intercepts for participants. Significance of regressor coefficients were obtained by F-tests.

Multiple linear regression of age-related changes. For significant age by emotion interactions, we performed additional analyses to examine age-related changes in each emotion category. Specifically, we employed multiple linear regression models, with behavioral or brain measure as dependent variable, age as a covariate of interest, and sex, race, and head movement as covariates of no interest.

Development of emotion circuits

Development of emotion-related network structure. To determine how network structure changes with age, we first determined community structure for each participant and each group as described above. We then computed the Jaccard index between individual and group consensus community structures. Specifically, overlap in community structure was computed using a co-occurrence matrix, with 1 indicating two nodes belonging to the same module and 0 indicating two nodes belonging to different modules. The Jaccard index is defined as the ratio of the intersection between two matrices divided by their union, ranging from 0 (no overlap) to 1 (full overlap). A linear mixed-effect model was then applied to the Jaccard index to determine main effects of age, emotion, and age by emotion interactions. We also

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

340 calculated the probability of a brain region being a hub node within each age window for all
341 permutations. Only nodes with a probability above the 95th percentile of its empirical null
342 distribution were considered as a significant hub.
343
344 **Classification analysis**
345 We used a non-linear SVM classification algorithm with 10-fold cross-validation to assess the
346 discriminability of regional connectivity measures between all pairs of emotion categories.
347 The python scikit-learn package (https://scikit-learn.org/stable/) was used to perform this
348 analysis. Permutation test was used to assess the performance of the classifier. Specifically,
349 for each pair of emotion categories, data were divided into 10 folds, with those belonging to
350 the same participant included in the same fold. A non-linear SVM classifier was trained using
351 nine folds, leaving out one fold. The samples in the left-out fold were then predicted using
352 this classifier and an accuracy was derived by calculating the percentage of correct
353 predictions. This procedure was repeated ten times, and a final classification accuracy was
354 derived by calculating the mean of the 10 left-out accuracy, which was used as a measure of
355 model performance. Next, the statistical significance of the model was assessed using
356 permutation test. In each permutation, emotion category labels (e.g., neutral vs. fear) were
357 randomly exchanged with a probability of 0.5 for each participant. A classifier was trained,
358 and a classification accuracy was derived using the same scheme described above.
359 Classification accuracy from 100 permutations were used to construct the empirical null
360 distribution from which a p-value was computed.

JNeurosci Accepted Manuscript

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 <= N <= 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 <= N <= 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

JNeurosci Accepted Manuscript

382 across subsamples at each sample size. The resulting average correlation was used as the
383 measure of network-level reliability of overall developmental pattern across all brain regions.
384 We defined reliable findings as those with correlation above 0.7 (i.e., R-square > 0.5), the
385 point at which the model explains more variance than is unexplained. Finally, we also
386 constructed empirical null distributions for reliability measured as average correlation, using
387 the same approach described above, with an additional step that age was permuted in each
388 subsample before assessing developmental changes. Average correlation above the 95th
389 percentile of its empirical null distribution was considered as significantly different from null.
390
391 **Replication analysis of previous developmental findings related to amygdala circuitry**
392 We examined the development of amygdala activity and connectivity in the current PNC
393 cohort using coordinates reported in three published task-fMRI studies using similar
394 emotion face perception tasks, contrasts, and age range as the PNC (Gee et al., 2013; Kujawa
395 et al., 2016; Wu et al., 2016). **Tables 1 and 4** provide details of the studies and coordinates
396 used. Moreover, we performed additional gPPI (generalized psychophysiological interaction)
397 analysis to establish amygdala-mPFC connectivity and examine age-related changes as a
398 similar method (gPPI or PPI) was used in the three published studies. Specifically, gPPI
399 model included one regressor of seed time series, five regressors of task conditions (neutral,
400 fear, anger, sadness, or happiness) and their interaction with seed time series, one regressor
401 of error trials and its interaction with seed time series, six regressors of head movement, and

JNeurosci Accepted Manuscript

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 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v2.p2. 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.

JNeurosci Accepted Manuscript

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 <= N <= 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

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

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, $p < 2.2e\text{-}16$) and RT (t (998) = -8.87, $p < 2.2e\text{-}16$) between the two samples, with participants in the fMRI sample showing higher accuracy and faster response (accuracy: 0.92 ± 0.05; RT: 2.29 ± 0.36) than participants in the excluded sample (accuracy: 0.88 ± 0.10; RT: 2.51 ± 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),

JNeurosci Accepted Manuscript

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 >

JNeurosci Accepted Manuscript

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 ($ps > 0.1$), 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

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

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 5c**), with classification accuracy ranging from 56.2% to 67.2% and significantly higher than chance level ($ps < 0.01$) as assessed by permutation test.

JNeurosci Accepted Manuscript

612

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

614

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

629

630 **Amygdala responses are invariant with age while FP node responses decrease with age**

JNeurosci Accepted Manuscript

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.

JNeurosci Accepted Manuscript

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 ($Fs < 2$, $ps > 0.1$). 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 ($Fs < 1$, $ps > 0.1$). 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

JNeurosci Accepted Manuscript

673 using coordinates reported in Wu et al. (2016). We further examined age-related changes for
674 each of the five emotion categories. Age-related increases were found in the left amygdala
675 under Fear, Sad and Happy using coordinates reported in Kujawa et al. (2016) and in bilateral
676 amygdala under Fear, Sad, and/or Happy using coordinates reported in Wu et al. (2016)
677 (**Table 5**). In summary, our replication results from the large sample are partially consistent
678 with Kujawa et al. (2016) and Wu et al. (2016) findings and inconsistent with Gee et al.
679 (2013) finding (**Table 7**).
680
681 *Amygdala connectivity*. Gee et al. (2013) reported age-related decreases in the right
682 amygdala-vmPFC connectivity during fearful face processing. We included five emotion
683 categories in our linear mixed-effect model and did not identify age effect or age by emotion
684 interactions in their reported amygdala-VMPFC regions in the current PNC sample ($Fs < 1$,
685 $ps > 0.1$). We further examined age-related changes for each of the five emotion categories
686 and did not identify any age-related changes (**Table 6**). Kujawa et al. (2016) reported
687 age-related decreases in bilateral amygdala-dACC connectivity during happy, angry, and
688 fearful face identification in typically developing individuals. When we included five
689 emotion categories in linear mixed-effect model, we found no age-related changes or age by
690 emotion interactions using the reported coordinates ($Fs < 1$, $ps > 0.1$). When we included only
691 Fear, Anger and Happy, the three emotion categories used in Kujawa et al. (2016) study, we
692 found a significant age-related increases in right amygdala-dACC connectivity ($F(1, 2011) =$
693 8.250, $p = 0.004$, *effect size* $= 0.104$). We further examined age-related changes for each of

JNeurosci Accepted Manuscript

694 the five emotion categories. Age-related increases were found in left amygdala-dACC under
695 Fear and Happy and in right amygdala-dACC under Fear, Anger and Sad (**Table 6**). Finally,
696 Wu et al. (2016) reported age-related decreases in bilateral amygdala-mPFC/ACC during
697 identification of fearful, angry, and happy faces. When we included all five emotion
698 categories in linear mixed-effect model, we found no age effect or age by emotion
699 interactions (*Fs* < 2, *ps* > 0.1). When we included only Fear, Anger and Happy, the three
700 emotion categories used in Wu et al. (2016) study, we found a significant age-related
701 increases in right amygdala connectivity with left mPFC (F (1, 2041) = 8.046, p = 0.005,
702 *effect size* = 0.103) and right ACC (F (1, 2037) = 6.749, p = 0.009, *effect size* = 0.094). We
703 further examined age-related changes for each of the five emotion categories and found
704 age-related increases in bilateral amygdala-mPFC/ACC connectivity under Fear and/or
705 Anger (**Table 6**).
706
707 *Amygdala connectivity (gPPI analysis)*. Because previous developmental studies of
708 amygdala connectivity have used PPI (Gee et al., 2013) or gPPI techniques (Kujawa et al.,
709 2016; Wu et al., 2016), we conducted additional analyses using a similar approach. When we
710 included the five emotion categories in linear mixed-effect model, we did not find a main
711 effect of age, or age by emotion interactions in analyses using coordinates from the three
712 studies (*Fs* < 1.5, *ps* > 0.1). When we included only Fear, Anger and Happy in linear
713 mixed-effect model, we did not identify age effect or age by emotion interactions in any of the
714 amygdala connectivity using coordinates from Kujawa et al. (2016) and Wu et al. (2016)

JNeurosci Accepted Manuscript

715 studies ($Fs < 2.2$, $ps > 0.1$). Finally, we examined age-related changes for each emotion
716 category. Results revealed no age effect using coordinates reported in Gee et al. (2013) study
717 ($Fs < 2$, $ps > 0.1$) but significant age-related increases in bilateral amygdala-dACC
718 connectivity under Fear (left: F (1, 754) = 5.602, p = 0.018, *effect size* = 0.088; right: F (1,
719 754) = 4.605, p = 0.032, *effect size* = 0.080) using coordinates reported in Kujawa et al.
720 (2016) study as well as in left amygdala-mPFC connectivity under Fear (F (1, 754) = 7.041, p
721 = 0.008, *effect size* = 0.099) using coordinates reported in Wu et al. (2016) study. These
722 results are consistent with those derived using beta time series analysis, as described in the
723 previous section.
724
725 Taken together, these results point to poor replicability of previous developmental findings
726 on amygdala connectivity associated with emotion processing (**Table 7**).
727
728 **Discussion**
729 A large sample of children, adolescents, and young adults with high quality emotion
730 perception task fMRI data from the PNC (Satterthwaite et al., 2014) afforded us an
731 unprecedented opportunity to address critical gaps in our knowledge regarding the
732 development of human emotion identification circuits. We employed a multipronged
733 analytical strategy to investigate the development of functional circuits associated with
734 perception and identification of five distinct emotion categories: fear, anger, sadness,
735 happiness, and neutral. Our analysis disentangled aspects of emotion-related brain circuitry

JNeurosci Accepted Manuscript

736 that were stable over development and those that changed with age. We found a
737 developmentally-stable modular architecture anchored in the salience, default mode and
738 central executive networks, with hubs in the mPFC and emotion-related reconfiguration of
739 the salience network. Developmental changes were most prominent in fronto-parietal circuits
740 important for salience detection and cognitive control. Crucially, reliability analyses provided
741 quantitative evidence for the robustness of our findings. Our findings provide a new template
742 for investigation of emotion processing in the developing brain.
743
744 **Maturation of emotion identification and specificity with respect to high arousal**
745 **negative emotions**
746 Behaviorally, facial emotion identification emerges early in infancy with continued
747 improvements through childhood and adolescence (Batty and Taylor, 2006; Thomas et al.,
748 2007; Rodger et al., 2015; Leitzke and Pollak, 2016; Theurel et al., 2016). We found that
749 although emotion identification accuracy was high for all five emotion categories even in the
750 youngest individuals (age 8), identification of Fear and Anger emotion categories improved
751 significantly with age, consistent with previous reports (Gee et al., 2013; Kujawa et al., 2016;
752 Wu et al., 2016). These findings suggest that high arousal negative ("threat-related")
753 emotions show more prominent differentiation, allowing for further sharpening of emotion
754 concepts with age (Widen, 2013; Nook et al., 2017).
755
756 **Architecture and developmental stability of emotion circuitry**

JNeurosci Accepted Manuscript

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

JNeurosci Accepted Manuscript

778 mentalizing (Ochsner et al., 2012; Dixon et al., 2017), while the pgACC is important for
779 awareness of one's own emotions (Lane et al., 2015; Dixon et al., 2017). Lastly, our results
780 are in line with the theoretical framework that the mPFC plays a central role in emotion
781 categorization by tuning other brain systems to distinct emotion categories (Heberlein et al.,
782 2008; Roy et al., 2012; Barrett and Satpute, 2013; Satpute et al., 2013; Satpute et al., 2016).
783
784 Surprisingly, while the PCC was part of the same module as the mPFC, it was not a hub
785 region, despite being consistently identified as hub in intrinsic functional connectivity studies
786 (Buckner et al., 2008; Fransson and Marrelec, 2008; Greicius et al., 2009; Andrews-Hanna et
787 al., 2010). These key differences between intrinsic and task-based network organization
788 argues for the critical importance of characterizing task-dependent brain circuits.
789
790 **Modular and regional connectivity changes are largely driven by emotion category and**
791 **are stable with age**
792 Anchored on a stable modular architecture, we found significant differences in intra- and
793 inter- module and regional connectivity associated with emotion category, and to a lesser
794 degree age. Prominent differences between emotion categories provides evidence that
795 distinct emotions, although anchored in a common modular architecture, may be represented
796 differently in the brain at the network level (Kragel and LaBar, 2015; Saarimaki et al., 2016).
797 Our SVM-based classification provides evidence of unique brain connectivity patterns

JNeurosci Accepted Manuscript

798 elicited by each emotion category and indicate differential patterns of distributed network
799 connectivity patterns involving the same regions (Barrett, 2017).
800
801 While module connectivity was largely developmentally stable, two findings of age-related
802 changes are noteworthy. First, significant age-related increases in FP intra-module
803 connectivity point to continued maturation of FP control systems underlying emotion
804 categorization over development (Ahmed et al., 2015; Guyer et al., 2016). Second,
805 age-related increases in mPFC/PCC intra-module connectivity during identification of fearful
806 and angry faces suggest that maturation of midline DMN systems underlie continued
807 developmental improvements in perception of threat and high emotion-arousal emotions.
808 Lastly, significant developmental changes in regional connectivity centered predominantly
809 on the insula, a core node of the salience network, highlighting a central and previously
810 underrecognized role of the insula in the development of emotion perception and
811 identification.
812
813 **Developmental changes in amygdala reactivity and amygdala-PFC circuits**
814 Surprisingly, we did not detect developmental changes in the amygdala and its functional
815 circuity, given that amygdala and its connectivity with the mPFC have been the focus of
816 developmental studies of emotion processing, but with contradictory findings (see **Table 1**
817 for a summary). To address these discrepancies, we examined amygdala activity and
818 amygdala-mPFC connectivity in our PNC cohort using specific amygdala seed and mPFC

JNeurosci Accepted Manuscript

819 target regions from three published studies with similar emotion face recognition tasks,
820 experimental contrasts, and age ranges (Gee et al., 2013; Kujawa et al., 2016; Wu et al.,
821 2016). We were unable to replicate the findings of these previous studies, suggesting that
822 small samples and not differences in amygdala and mPFC coordinates are a major contributor
823 to the discrepancies between the current and previous studies. These findings suggest that a
824 narrow focus on amygdala-mPFC connectivity has precluded characterization of other
825 important brain regions, most notably the insula and other PFC regions, underlying the
826 development of emotion perception and categorization.
827
828 Further studies are needed to determine the extent to which differences in experimental
829 design contributed to inconsistent patterns of developmental changes in amygdala activity
830 and connectivity reported in previous studies. Critically, addressing this question will require
831 much larger developmental samples than used in previous studies, along with stability and
832 reliability analyses as in the present study.
833
834 **Neutral faces engage emotion circuitry in the context of emotion identification**
835 Another surprising finding of our study was that neutral faces showed distributed and strong
836 levels of activity and connectivity across the three modules. Disambiguation of neural faces
837 from other emotions would require perception, evaluation, and decision-making and, as our
838 data indicates, engage emotion perception and categorization circuits. Moreover, the
839 classification rates for neutral faces were no different from those between the other emotions.

JNeurosci Accepted Manuscript

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.

Conclusions

JNeurosci Accepted Manuscript

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.

JNeurosci Accepted Manuscript

867 **References**
868
869 Ahmed SP, Bittencourt-Hewitt A, Sebastian CL (2015) Neurocognitive bases of emotion
870 regulation development in adolescence. Dev Cogn Neurosci 15:11-25.

871 Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL (2010)
872 Functional-anatomic fractionation of the brain's default network. Neuron
873 65:550-562.

874 Avery SN, Clauss JA, Winder DG, Woodward N, Heckers S, Blackford JU (2014) BNST
875 neurocircuitry in humans. Neuroimage 91:311-323.

876 Barrett LF (2017) The theory of constructed emotion: an active inference account of
877 interoception and categorization. Soc Cogn Affect Neurosci 12:1-23.

878 Barrett LF, Satpute AB (2013) Large-scale brain networks in affective and social
879 neuroscience: towards an integrative functional architecture of the brain. Curr Opin
880 Neurobiol 23:361-372.

881 Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models
882 Using lme4. J Stat Softw 67:1-48.

883 Batty M, Taylor MJ (2006) The development of emotional face processing during
884 childhood. Dev Sci 9:207-220.

885 Blackford JU, Pine DS (2012) Neural substrates of childhood anxiety disorders: a review of
886 neuroimaging findings. Child Adolesc Psychiatr Clin N Am 21:501-525.

887 Blakemore SJ (2008) The social brain in adolescence. Nat Rev Neurosci 9:267-277.

888 Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of
889 communities in large networks. Journal of Statistical Mechanics: Theory and
890 Experiment 2008.

891 Braunstein LM, Gross JJ, Ochsner KN (2017) Explicit and implicit emotion regulation: a
892 multi-level framework. Soc Cogn Affect Neurosci 12:1545-1557.

893 Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: anatomy,
894 function, and relevance to disease. Ann N Y Acad Sci 1124:1-38.

895 Buhle JT, Silvers JA, Wager TD, Lopez R, Onyemekwu C, Kober H, Weber J, Ochsner KN
896 (2014) Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging
897 studies. Cereb Cortex 24:2981-2990.

JNeurosci Accepted Manuscript

898 Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafo MR (2013)
899 Power failure: why small sample size undermines the reliability of neuroscience. Nat
900 Rev Neurosci 14:365-376.

901 Chang LJ, Yarkoni T, Khaw MW, Sanfey AG (2013) Decoding the role of the insula in
902 human cognition: functional parcellation and large-scale reverse inference. Cereb
903 Cortex 23:739-749.

904 Craig AD (2009) How do you feel--now? The anterior insula and human awareness. Nat
905 Rev Neurosci 10:59-70.

906 Deen B, Pitskel NB, Pelphrey KA (2011) Three systems of insular functional connectivity
907 identified with cluster analysis. Cereb Cortex 21:1498-1506.

908 Dixon ML, Thiruchselvam R, Todd R, Christoff K (2017) Emotion and the prefrontal cortex:
909 An integrative review. Psychol Bull 143:1033-1081.

910 Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD,
911 Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct
912 brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U
913 S A 104:11073-11078.

914 Dosenbach NUF, Koller JM, Earl EA, Miranda-Dominguez O, Klein RL, Van AN, Snyder
915 AZ, Nagel BJ, Nigg JT, Nguyen AL, Wesevich V, Greene DJ, Fair DA (2017)
916 Real-time motion analytics during brain MRI improve data quality and reduce costs.
917 Neuroimage 161:80-93.

918 Dwyer DB, Harrison BJ, Yucel M, Whittle S, Zalesky A, Pantelis C, Allen NB, Fornito A
919 (2014) Large-scale brain network dynamics supporting adolescent cognitive control.
920 J Neurosci 34:14096-14107.

921 Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A
922 new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional
923 imaging data. Neuroimage 25:1325-1335.

924 Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial
925 extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113:7900-7905.

926 Engelhardt LE, Roe MA, Juranek J, DeMaster D, Harden KP, Tucker-Drob EM, Church JA
927 (2017) Children's head motion during fMRI tasks is heritable and stable over time.
928 Dev Cogn Neurosci 25:58-68.

929 Etkin A (2010) Functional neuroanatomy of anxiety: a neural circuit perspective. Curr Top
930 Behav Neurosci 2:251-277.

JNeurosci Accepted Manuscript

931 Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, Fox PT,
932 Eickhoff SB, Yu C, Jiang T (2016) The Human Brainnetome Atlas: A New Brain
933 Atlas Based on Connectional Architecture. Cereb Cortex 26:3508-3526.

934 Floresco SB (2015) The nucleus accumbens: an interface between cognition, emotion, and
935 action. Annu Rev Psychol 66:25-52.

936 Fortunato S, Barthelemy M (2007) Resolution limit in community detection. Proc Natl
937 Acad Sci U S A 104:36-41.

938 Fransson P, Marrelec G (2008) The precuneus/posterior cingulate cortex plays a pivotal role
939 in the default mode network: Evidence from a partial correlation network analysis.
940 Neuroimage 42:1178-1184.

941 Fusar-Poli P, Placentino A, Carletti F, Landi P, Allen P, Surguladze S, Benedetti F,
942 Abbamonte M, Gasparotti R, Barale F, Perez J, McGuire P, Politi P (2009)
943 Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105
944 functional magnetic resonance imaging studies. J Psychiatry Neurosci 34:418-432.

945 Gee DG, Humphreys KL, Flannery J, Goff B, Telzer EH, Shapiro M, Hare TA, Bookheimer
946 SY, Tottenham N (2013) A developmental shift from positive to negative
947 connectivity in human amygdala-prefrontal circuitry. J Neurosci 33:4584-4593.

948 Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting
949 brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A
950 100:253-258.

951 Greicius MD, Supekar K, Menon V, Dougherty RF (2009) Resting-state functional
952 connectivity reflects structural connectivity in the default mode network. Cereb
953 Cortex 19:72-78.

954 Guyer AE, Silk JS, Nelson EE (2016) The neurobiology of the emotional adolescent: From
955 the inside out. Neurosci Biobehav Rev 70:74-85.

956 Gyurak A, Gross JJ, Etkin A (2011) Explicit and implicit emotion regulation: a dual-process
957 framework. Cogn Emot 25:400-412.

958 Heberlein AS, Padon AA, Gillihan SJ, Farah MJ, Fellows LK (2008) Ventromedial frontal
959 lobe plays a critical role in facial emotion recognition. J Cogn Neurosci 20:721-733.

960 Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2:e124.

961 Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD (2008) Functional
962 grouping and cortical-subcortical interactions in emotion: a meta-analysis of
963 neuroimaging studies. Neuroimage 42:998-1031.

JNeurosci Accepted Manuscript

964 Kohn N, Eickhoff SB, Scheller M, Laird AR, Fox PT, Habel U (2014) Neural network of
965 cognitive emotion regulation--an ALE meta-analysis and MACM analysis.
966 Neuroimage 87:345-355.

967 Kragel PA, LaBar KS (2015) Multivariate neural biomarkers of emotional states are
968 categorically distinct. Soc Cogn Affect Neurosci 10:1437-1448.

969 Kujawa A, Wu M, Klumpp H, Pine DS, Swain JE, Fitzgerald KD, Monk CS, Phan KL
970 (2016) Altered Development of Amygdala-Anterior Cingulate Cortex Connectivity
971 in Anxious Youth and Young Adults. Biol Psychiatry Cogn Neurosci Neuroimaging
972 1:345-352.

973 Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep
974 2:336.

975 Lane RD, Weihs KL, Herring A, Hishaw A, Smith R (2015) Affective agnosia: Expansion
976 of the alexithymia construct and a new opportunity to integrate and extend Freud's
977 legacy. Neurosci Biobehav Rev 55:594-611.

978 Lebow MA, Chen A (2016) Overshadowed by the amygdala: the bed nucleus of the stria
979 terminalis emerges as key to psychiatric disorders. Mol Psychiatry 21:450-463.

980 LeDoux J (2007) The amygdala. Curr Biol 17:R868-874.

981 Leitzke BT, Pollak SD (2016) Developmental changes in the primacy of facial cues for
982 emotion recognition. Dev Psychol 52:572-581.

983 Lv Q, Yang L, Li G, Wang Z, Shen Z, Yu W, Jiang Q, Hou B, Pu J, Hu H, Wang Z (2016)
984 Large-Scale Persistent Network Reconfiguration Induced by Ketamine in
985 Anesthetized Monkeys: Relevance to Mood Disorders. Biol Psychiatry 79:765-775.

986 Menon V (2011) Large-scale brain networks and psychopathology: a unifying triple
987 network model. Trends Cogn Sci 15:483-506.

988 Menon V (2013) Developmental pathways to functional brain networks: emerging
989 principles. Trends Cogn Sci 17:627-640.

990 Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of
991 insula function. Brain Struct Funct 214:655-667.

992 Moser DA, Doucet GE, Lee WH, Rasgon A, Krinsky H, Leibu E, Ing A, Schumann G,
993 Rasgon N, Frangou S (2018) Multivariate Associations Among Behavioral, Clinical,
994 and Multimodal Imaging Phenotypes in Patients With Psychosis. JAMA Psychiatry
995 75:386-395.

JNeurosci Accepted Manuscript

996 Namkung H, Kim SH, Sawa A (2017) The Insula: An Underestimated Brain Area in
997 Clinical Neuroscience, Psychiatry, and Neurology. Trends Neurosci 40:200-207.

998 Nee DE (2018) Correspondence: fMRI replicability depends upon sufficient
999 individual-level data. bioRxiv:352633.

1000 Newman ME (2004) Fast algorithm for detecting community structure in networks. Phys
1001 Rev E Stat Nonlin Soft Matter Phys 69:066133.

1002 Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci
1003 U S A 103:8577-8582.

1004 Nook EC, Sasse SF, Lambert HK, McLaughlin KA, Somerville LH (2017) Increasing
1005 verbal knowledge mediates development of multidimensional emotion
1006 representations. Nat Hum Behav 1:881-889.

1007 Ochsner KN, Silvers JA, Buhle JT (2012) Functional imaging studies of emotion regulation:
1008 a synthetic review and evolving model of the cognitive control of emotion. Ann N Y
1009 Acad Sci 1251:E1-24.

1010 Ochsner KN, Knierim K, Ludlow DH, Hanelin J, Ramachandran T, Glover G, Mackey SC
1011 (2004) Reflecting upon feelings: an fMRI study of neural systems supporting the
1012 attribution of emotion to self and other. J Cogn Neurosci 16:1746-1772.

1013 Padmanabhan A, Luna B (2014) Developmental imaging genetics: linking dopamine
1014 function to adolescent behavior. Brain Cogn 89:27-38.

1015 Padmanabhan A, Geier CF, Ordaz SJ, Teslovich T, Luna B (2011) Developmental changes
1016 in brain function underlying the influence of reward processing on inhibitory control.
1017 Dev Cogn Neurosci 1:517-529.

1018 Paulus MP, Stein MB (2006) An insular view of anxiety. Biol Psychiatry 60:383-387.

1019 Peelen MV, Atkinson AP, Vuilleumier P (2010) Supramodal representations of perceived
1020 emotions in the human brain. J Neurosci 30:10127-10134.

1021 Phillips ML, Drevets WC, Rauch SL, Lane R (2003) Neurobiology of emotion perception II:
1022 Implications for major psychiatric disorders. Biol Psychiatry 54:515-528.

1023 Power JD, Fair DA, Schlaggar BL, Petersen SE (2010) The development of human
1024 functional brain networks. Neuron 67:735-748.

1025 Qin S, Young CB, Supekar K, Uddin LQ, Menon V (2012) Immature integration and
1026 segregation of emotion-related brain circuitry in young children. Proc Natl Acad Sci
1027 U S A 109:7941-7946.

JNeurosci Accepted Manuscript

1028 Qin S, Young CB, Duan X, Chen T, Supekar K, Menon V (2014) Amygdala subregional
1029 structure and intrinsic functional connectivity predicts individual differences in
1030 anxiety during early childhood. Biol Psychiatry 75:892-900.

1031 Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during
1032 distinct stages of a cognitive task. Neuroimage 23:752-763.

1033 Rodger H, Vizioli L, Ouyang X, Caldara R (2015) Mapping the development of facial
1034 expression recognition. Dev Sci 18:926-939.

1035 Roy M, Shohamy D, Wager TD (2012) Ventromedial prefrontal-subcortical systems and the
1036 generation of affective meaning. Trends Cogn Sci 16:147-156.

1037 Rubin TN, Koyejo O, Gorgolewski KJ, Jones MN, Poldrack RA, Yarkoni T (2017)
1038 Decoding brain activity using a large-scale probabilistic functional-anatomical atlas
1039 of human cognition. Plos Computational Biology 13.

1040 Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and
1041 interpretations. Neuroimage 52:1059-1069.

1042 Saarimaki H, Gotsopoulos A, Jaaskelainen IP, Lampinen J, Vuilleumier P, Hari R, Sams M,
1043 Nummenmaa L (2016) Discrete Neural Signatures of Basic Emotions. Cereb Cortex
1044 26:2563-2573.

1045 Satpute AB, Shu J, Weber J, Roy M, Ochsner KN (2013) The functional neural architecture
1046 of self-reports of affective experience. Biol Psychiatry 73:631-638.

1047 Satpute AB, Nook EC, Narayanan S, Shu J, Weber J, Ochsner KN (2016) Emotions in
1048 "Black and White" or Shades of Gray? How We Think About Emotion Shapes Our
1049 Perception and Neural Representation of Emotion. Psychol Sci 27:1428-1442.

1050 Satterthwaite TD, Elliott MA, Ruparel K, Loughead J, Prabhakaran K, Calkins ME, Hopson
1051 R, Jackson C, Keefe J, Riley M, Mentch FD, Sleiman P, Verma R, Davatzikos C,
1052 Hakonarson H, Gur RC, Gur RE (2014) Neuroimaging of the Philadelphia
1053 neurodevelopmental cohort. Neuroimage 86:544-553.

1054 Schaer M, Kochalka J, Padmanabhan A, Supekar K, Menon V (2015) Sex differences in
1055 cortical volume and gyrification in autism. Mol Autism 6:42.

1056 Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius
1057 MD (2007) Dissociable intrinsic connectivity networks for salience processing and
1058 executive control. J Neurosci 27:2349-2356.

JNeurosci Accepted Manuscript

1059 Sperduti M, Makowski D, Arcangeli M, Wantzen P, Zalla T, Lemaire S, Dokic J, Pelletier J,
1060 Piolino P (2017) The distinctive role of executive functions in implicit emotion
1061 regulation. Acta Psychol (Amst) 173:13-20.

1062 Sporns O, Betzel RF (2016) Modular Brain Networks. Annu Rev Psychol 67:613-640.

1063 Szucs D, Ioannidis JP (2017) Empirical assessment of published effect sizes and power in
1064 the recent cognitive neuroscience and psychology literature. PLoS Biol
1065 15:e2000797.

1066 Theurel A, Witt A, Malsert J, Lejeune F, Fiorentini C, Barisnikov K, Gentaz E (2016) The
1067 integration of visual context information in facial emotion recognition in 5- to
1068 15-year-olds. J Exp Child Psychol 150:252-271.

1069 Thomas LA, De Bellis MD, Graham R, LaBar KS (2007) Development of emotional facial
1070 recognition in late childhood and adolescence. Dev Sci 10:547-558.

1071 Tupak SV, Dresler T, Guhn A, Ehlis AC, Fallgatter AJ, Pauli P, Herrmann MJ (2014)
1072 Implicit emotion regulation in the presence of threat: neural and autonomic
1073 correlates. Neuroimage 85 Pt 1:372-379.

1074 Turner BO, Paul EJ, Miller MB, Barbey AK (2018) Small sample sizes reduce the
1075 replicability of task-based fMRI studies. Communications Biology 1.

1076 Uddin LQ, Kinnison J, Pessoa L, Anderson ML (2014) Beyond the tripartite
1077 cognition-emotion-interoception model of the human insular cortex. J Cogn
1078 Neurosci 26:16-27.

1079 Vink M, Derks JM, Hoogendam JM, Hillegers M, Kahn RS (2014) Functional differences
1080 in emotion processing during adolescence and early adulthood. Neuroimage
1081 91:70-76.

1082 Widen SC (2013) Children's Interpretation of Facial Expressions: The Long Path from
1083 Valence-Based to Specific Discrete Categories. Emotion Review 5:72-77.

1084 Wolf RC, Herringa RJ (2016) Prefrontal-Amygdala Dysregulation to Threat in Pediatric
1085 Posttraumatic Stress Disorder. Neuropsychopharmacology 41:822-831.

1086 Wu M, Kujawa A, Lu LH, Fitzgerald DA, Klumpp H, Fitzgerald KD, Monk CS, Phan KL
1087 (2016) Age-related changes in amygdala-frontal connectivity during emotional face
1088 processing from childhood into young adulthood. Hum Brain Mapp 37:1684-1695.

1089 Xia M, Wang J, He Y (2013) BrainNet Viewer: a network visualization tool for human
1090 brain connectomics. PLoS One 8:e68910.

JNeurosci Accepted Manuscript

1091 Yan CG, Craddock RC, He Y, Milham MP (2013) Addressing head motion dependencies
1092 for small-world topologies in functional connectomics. Front Hum Neurosci 7:910.

1093 Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale
1094 automated synthesis of human functional neuroimaging data. Nat Methods
1095 8:665-670.

1096 Yuan W, Altaye M, Ret J, Schmithorst V, Byars AW, Plante E, Holland SK (2009)
1097 Quantification of head motion in children during various fMRI language tasks. Hum
1098 Brain Mapp 30:1481-1489.
1099
1100

JNeurosci Accepted Manuscript

Figure Legends

Figure 1. Schematic view of participant selection procedure and main analysis steps. a From the original 1495 participants, 157 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 excess 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, 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). 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 ($10\% \leq \text{sparsity} \leq 40\%$ 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

JNeurosci Accepted Manuscript

(**b-ii**). **c** 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 (SPI) 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 SPI 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-SPI connectivity increased with age; Intra-FP, intra-SPI, 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 (SPI) 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

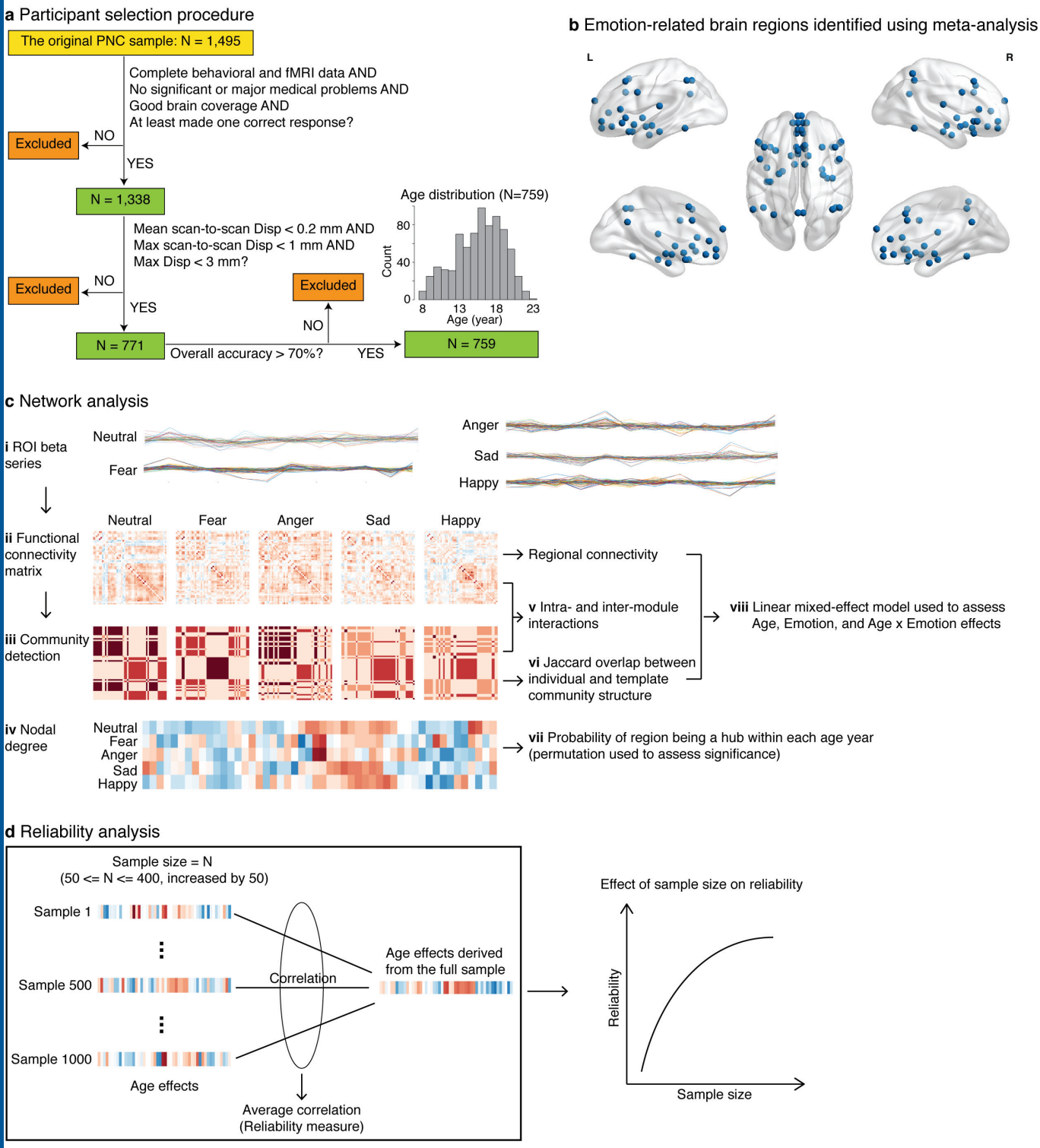
JNeurosci Accepted Manuscript

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 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. 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.

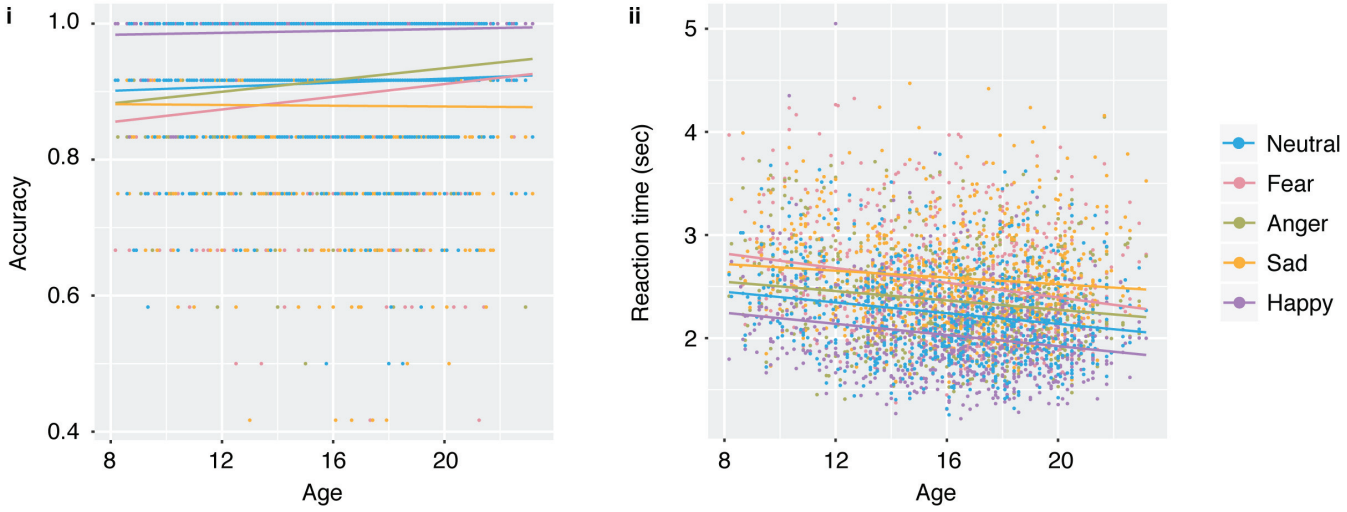
Figure 6. Main effects of age and emotion on regional activity. a i. Brain regions showing age-related changes. Activity in cognitive control regions, including bilateral preSMA, dACC and DLPFC, left SPM, right IPL, and right LOFC, decreased with age. Node size was scaled to present effect size of age-related 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.

JNeurosci Accepted Manuscript

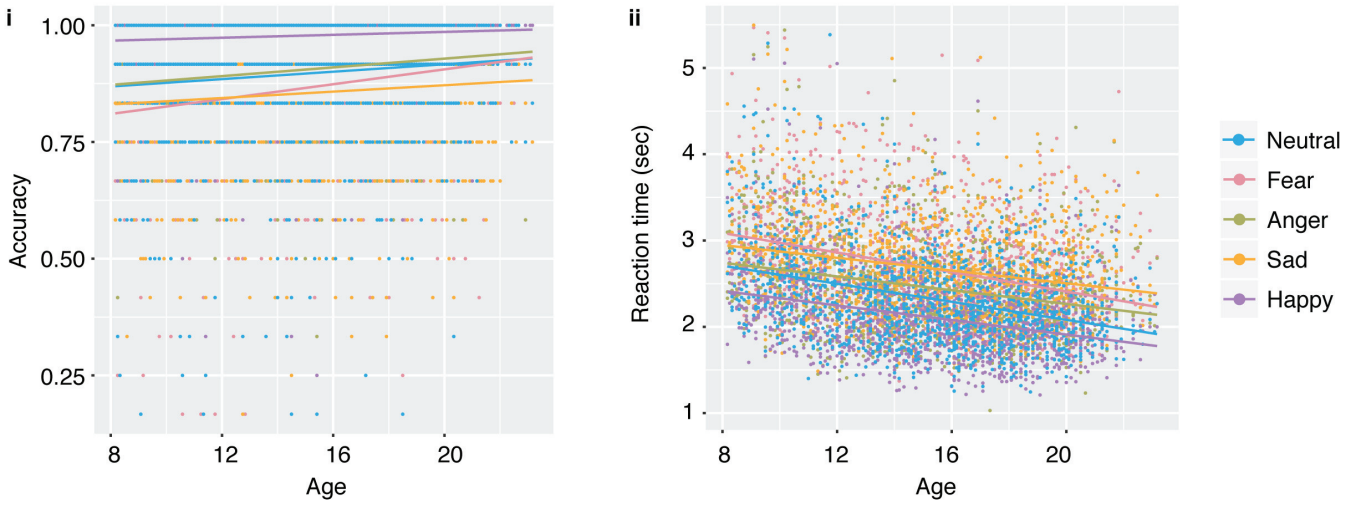


JNeurosci Accepted Manuscript

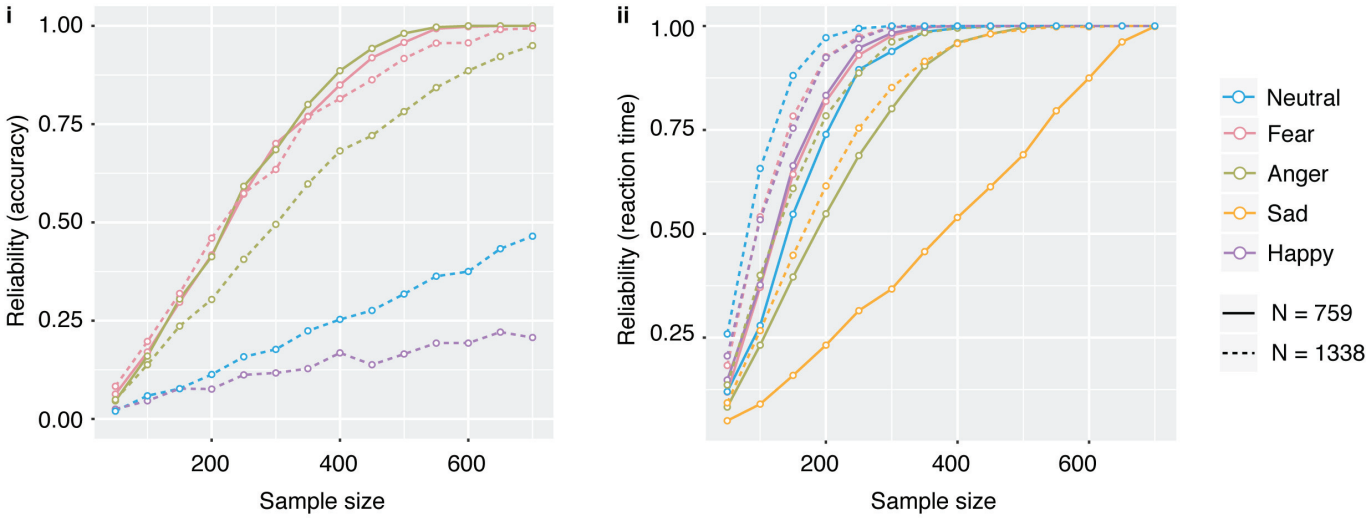
a Age-related changes in emotion identification: accuracy and reaction time (N = 759)



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

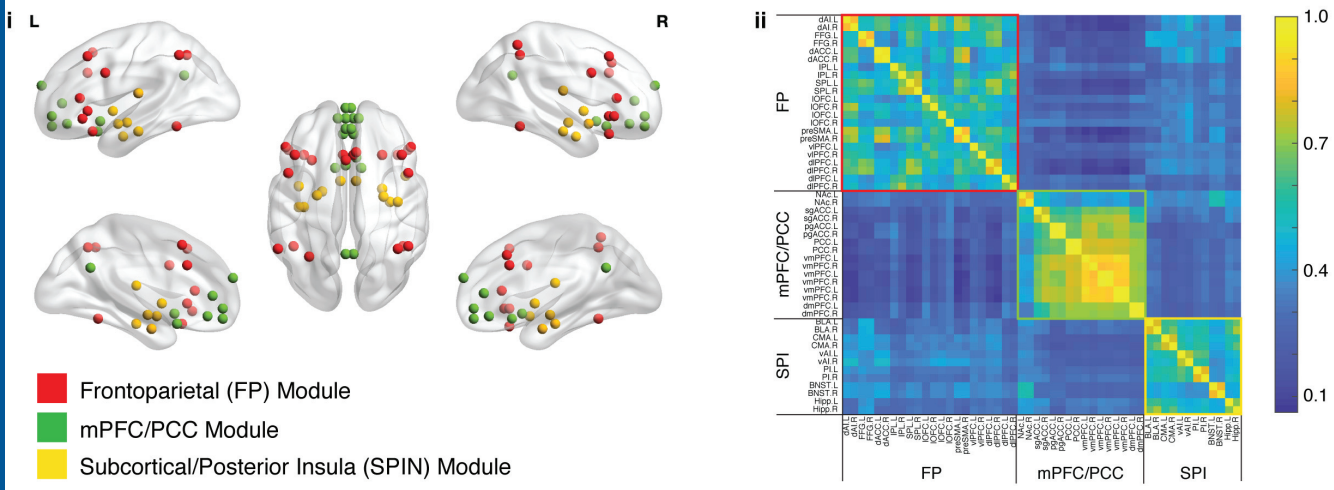


c Effect of sample size on reliability of emotion identification: accuracy and reaction time

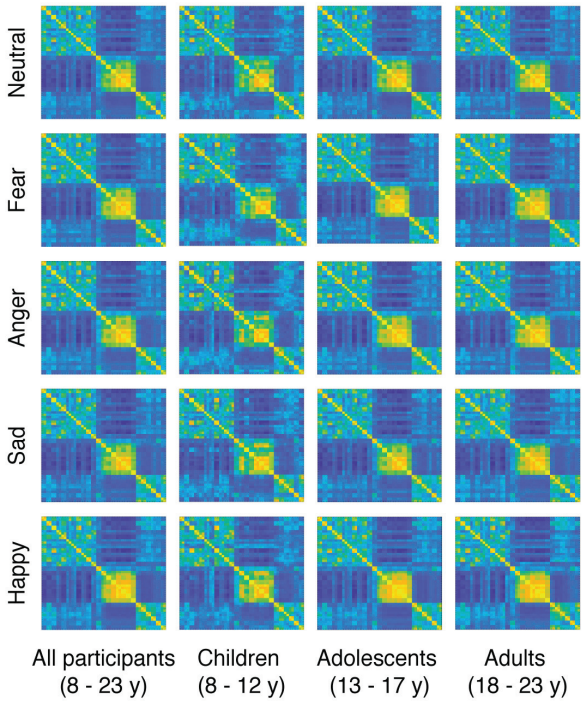


JNeurosci Accepted Manuscript

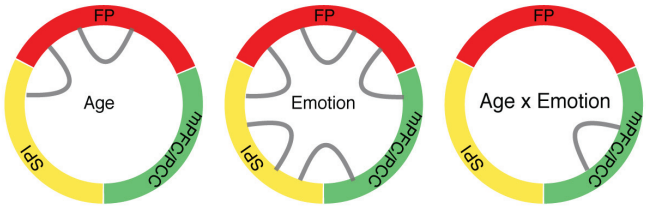
a Modular structure of emotion identification circuits



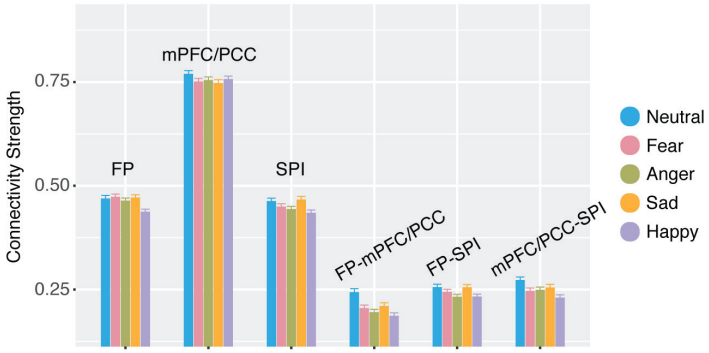
b Modular structure for each emotion category over development



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

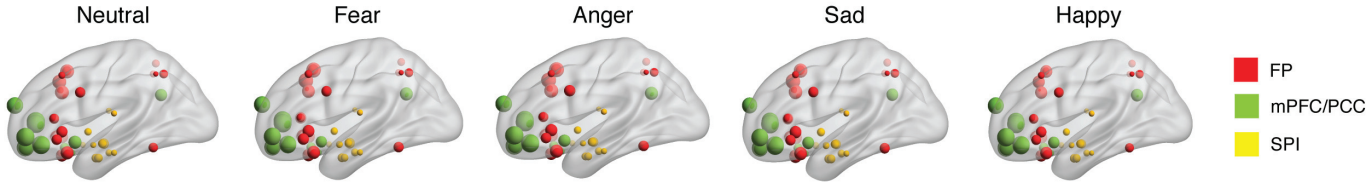


d Intra- and inter-module connectivity across emotions



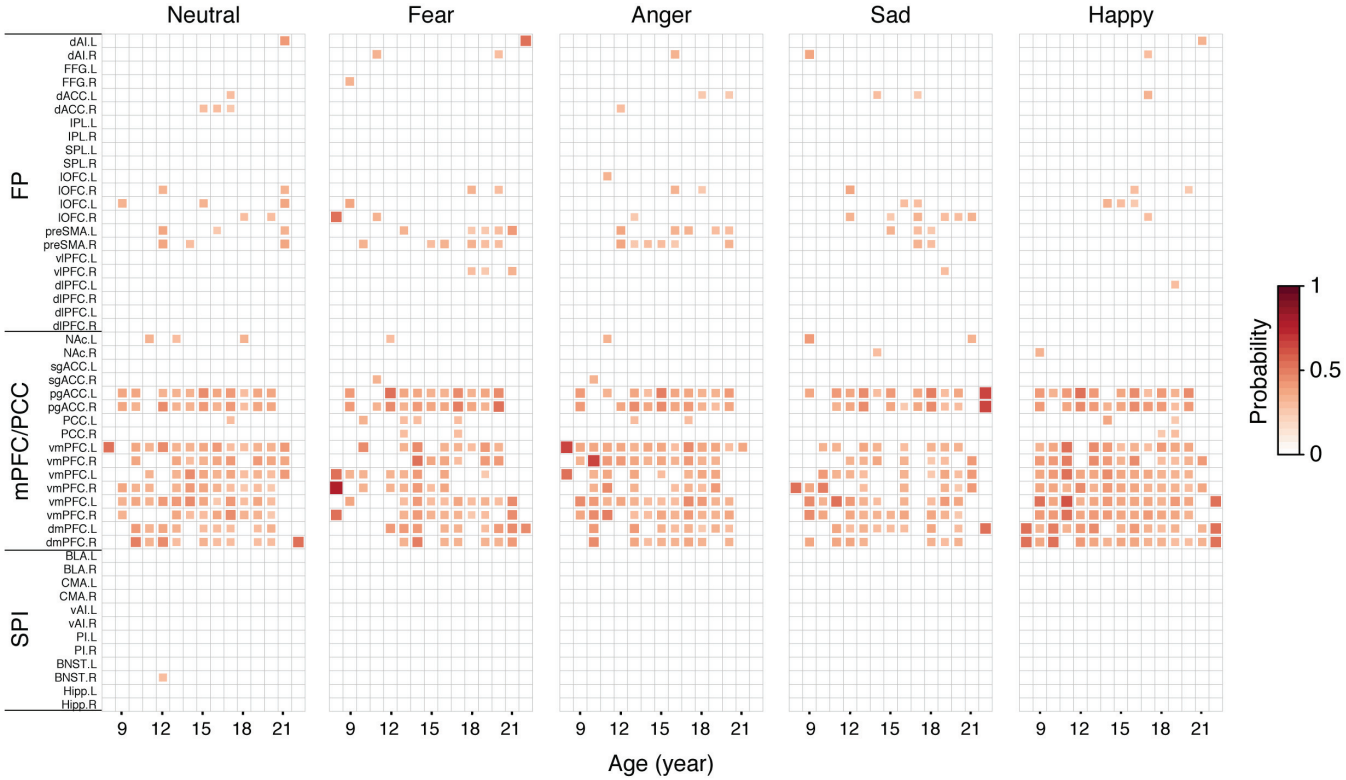
JNeurosci Accepted Manuscript

a Nodal degree of circuit connectivity for each emotion category



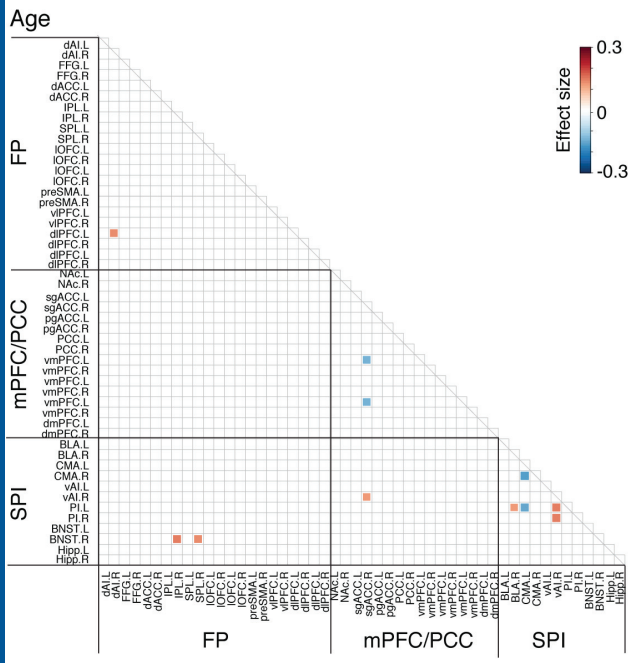
b Hub probability for each emotion category over development

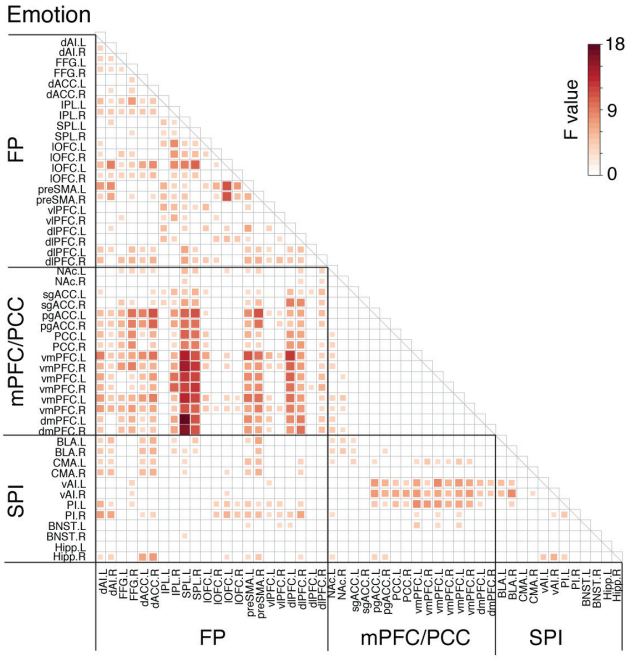
c Developmentally stable hubs in mPFC/PCC



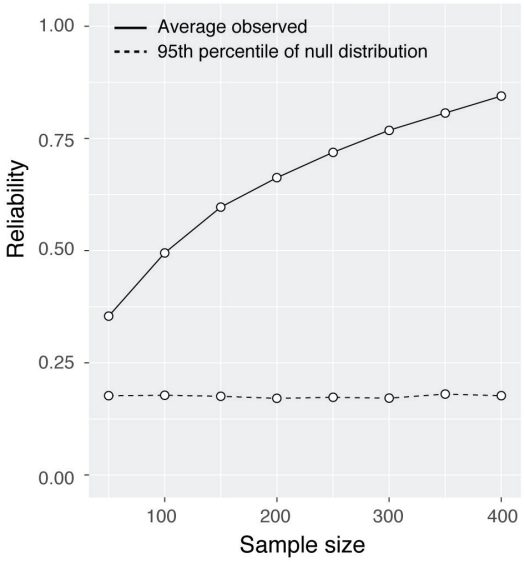
JNeurosci Accepted Manuscript

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

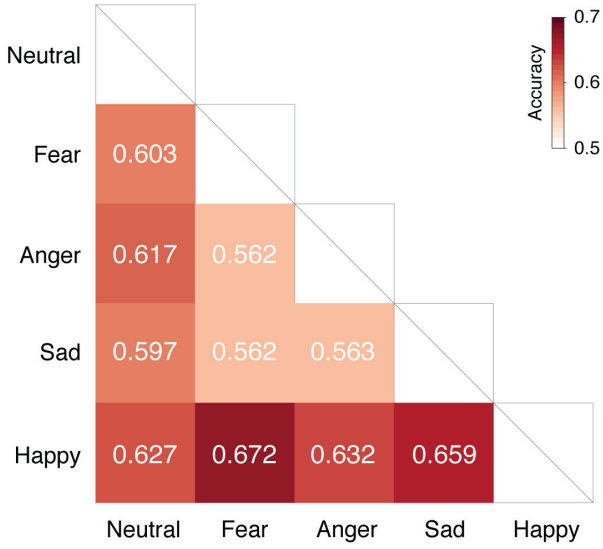




b Effect of sample size on reliability of age-related changes

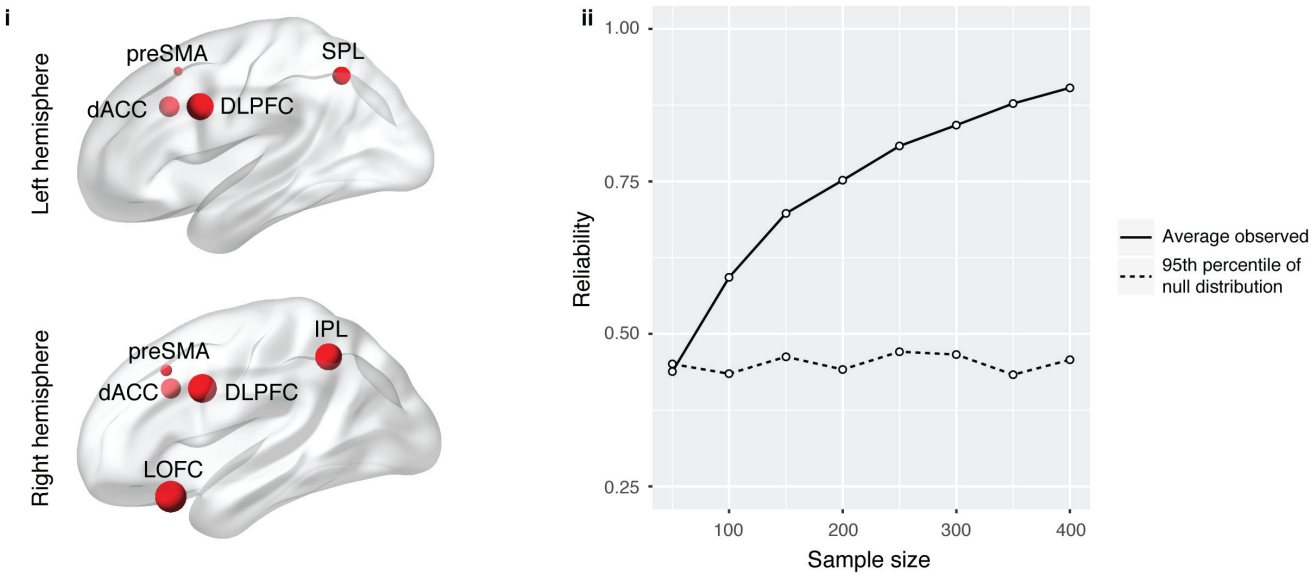


c Classification accuracy for differentiating connectivity patterns of emotions

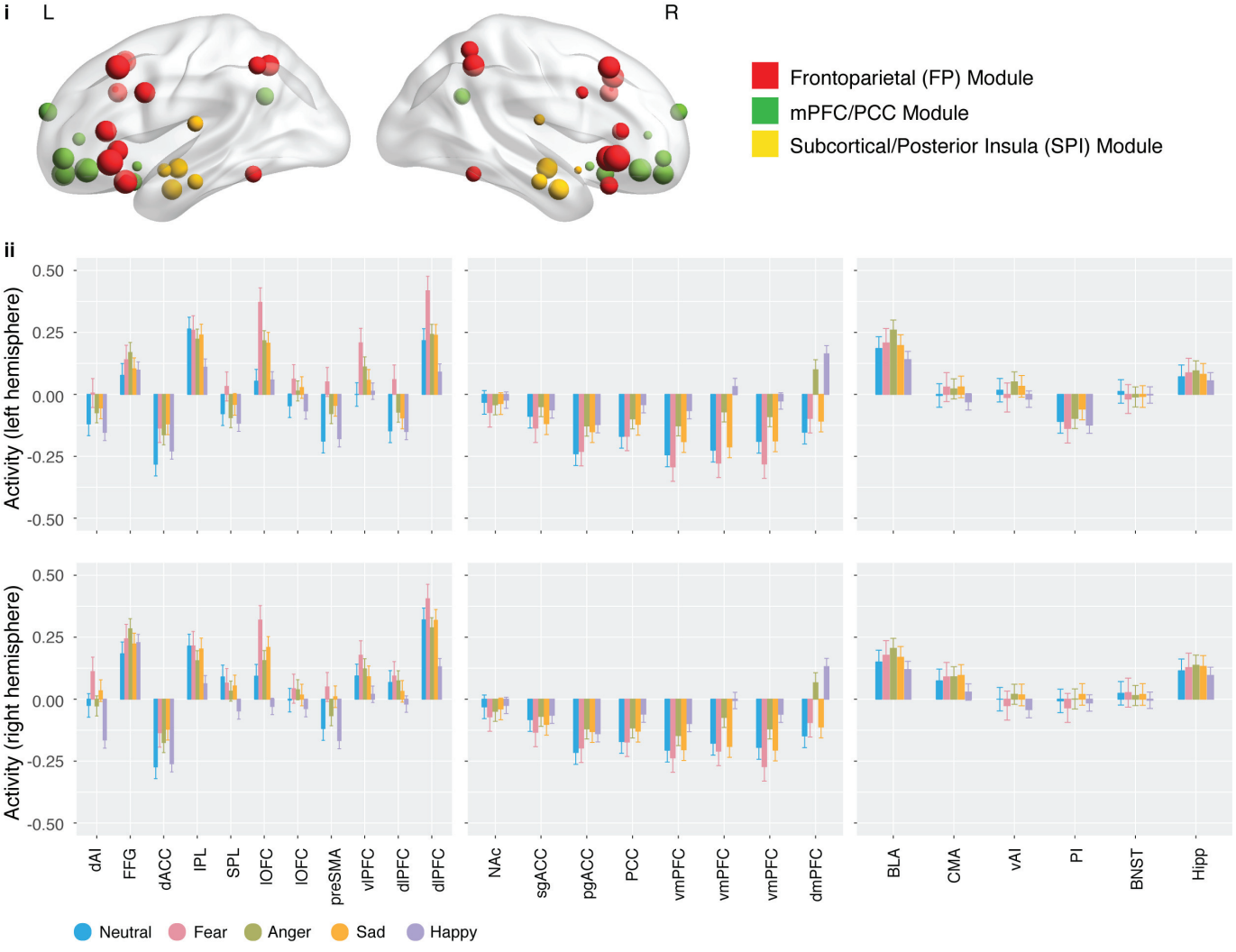


JNeurosci Accepted Manuscript

a Age-related changes in regional activity and reliability



b Effect of emotion on regional activity



JNeurosci Accepted Manuscript

Tables

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

Study	Sample Size	Age (years)	Contrast and #trials	Task	Findings
Gee et al., 2013	45 (19F)	4 – 22	Fearful face vs. Fixation (#trials = 24)	Participants viewed fearful faces with interspersed with neutral faces in one run, and viewed happy faces interspersed with neutral faces in the other run	Amygdala activity decreased with age; amygdala-vmPFC connectivity decreased with age
Kujawa et al., 2016	61 (35F) TD 57 (34F) AD	7 – 25	Emotional faces (happy, angry or fearful) vs. Shapes (#trials per emotion category = 12)	Participants were instructed to match emotion (happy, angry or fearful) of the target face in face blocks and match shapes in shape blocks	No age-related effect in amygdala activity; amygdala-dACC connectivity decreased with age in TD but increased with age in AD for all emotions
Wu et al., 2016	61 (35F)	7 – 25	Emotional faces (angry, fearful or happy) vs. Shapes (#trials per emotion category = 12)	Participants were instructed to match emotion (angry, fearful or happy) of the target face in face blocks and match shapes in shape blocks	No age-related effect in amygdala activity; amygdala-ACC/mPFC connectivity decreased with age for all emotions
Vink et al., 2014	60 (28F)	10 – 24	Positive vs. Neutral (#trials = 32), Negative vs. Neutral (#trials = 32)	Participants viewed each picture selected from the IAPS and label the valence (negative, neutral, or positive)	Amygdala activity decreased with age; amygdala-mOFC connectivity increased with age

JNeurosci Accepted Manuscript

Wolf & Herringa, 2016	24 (13F) TD 24 (16F) PTSD	8 – 18	Threat vs. Neutral (#trials = 16)	Participants viewed threat and neutral images selected from the IAPS and label the valence	No age-related effect in amygdala activity; amygdala-mOFC connectivity increased with age in TD but decreased with age in PTSD

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 & Herrniga (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.

JNeurosci Accepted Manuscript

Table 2. Participant characteristics and task performance.

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

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.

JNeurosci Accepted Manuscript

Table 3. Anatomical location and MNI coordinates of emotion circuit-related nodes.

Region/Abbreviation	MNI Coordinates			Brodmann Area
	x	y	z	
Left basolateral amygdala (BLA.L)	-20	-6	-20	
Right basolateral amygdala (BLA.R)	26	-4	-20	
Left centromedial amygdala (CMA.L)	-24	-9	9	
Right centromedial amygdala (CMA.R)	27	-10	-9	
Left subgenual anterior cingulate cortex (sgACC.L)	-2	14	-16	BA 25
Right subgenual anterior cingulate cortex (sgACC.R)	6	20	-12	BA 25
Left pregenual anterior cingulate cortex (pgACC.L)	-2	42	6	BA 32
Right pregenual anterior cingulate cortex (pgACC.R)	2	42	8	BA 32
Left dorsal anterior cingulate cortex (dACC.L)	-4	22	30	BA 32
Right dorsal anterior cingulate cortex (dACC.R)	4	22	30	BA 32
Left posterior cingulate cortex (PCC.L)	-2	-54	28	BA 31
Right posterior cingulate cortex (PCC.R)	4	-54	28	BA 31
Left ventromedial prefrontal cortex (vmPFC.L)	-8	50	-4	BA 10
Right ventromedial prefrontal cortex (vmPFC.R)	8	50	-4	BA 10
Left ventromedial prefrontal cortex (vmPFC.L)	-2	50	-12	BA 32
Right ventromedial prefrontal cortex (vmPFC.R)	4	50	-12	BA 32
Left ventromedial prefrontal cortex (vmPFC.L)	-4	38	-10	BA 32
Right ventromedial prefrontal cortex (vmPFC.R)	4	38	-10	BA 32
Left dorsomedial prefrontal cortex (dmPFC.L)	-2	58	20	BA 9
Right dorsomedial prefrontal cortex (dmPFC.R)	2	58	20	BA 9
Left pre-supplementary motor area (preSMA.L)	-2	18	46	BA 6
Right pre-supplementary motor area (preSMA.R)	6	24	38	BA 32
Left lateral orbitofrontal cortex (lOFC.L)	-46	26	-6	BA 47
Right lateral orbitofrontal cortex (lOFC.R)	46	26	-4	BA 47
Left lateral orbitofrontal cortex (lOFC.L)	-28	18	-16	BA 47
Right lateral orbitofrontal cortex (lOFC.R)	26	22	-18	BA 47
Left ventrolateral prefrontal cortex (vlPFC.L)	-48	28	10	BA 45
Right ventrolateral prefrontal cortex (vlPFC.R)	48	28	10	BA 45
Left dorsolateral prefrontal cortex (dlPFC.L)	-46	8	30	BA 9
Right dorsolateral prefrontal cortex (dlPFC.R)	46	8	30	BA 9
Left dorsolateral prefrontal cortex (dlPFC.L)	-41	22	43	BA 8
Right dorsolateral prefrontal cortex (dlPFC.R)	41	22	43	BA 8
Left inferior parietal lobule (IPL.L)	-54	-50	44	BA 40
Right inferior parietal lobule (IPL.R)	46	-48	44	BA 40

JNeurosci Accepted Manuscript

Left superior parietal lobule (SPL.L)	-30	-56	44	BA 7
Right superior parietal lobule (SPL.R)	38	-50	52	BA 7
Left fusiform gyrus (FFG.L)	-42	-48	-20	BA 37
Right fusiform gyrus (FFG.R)	42	-48	-20	BA 37
Left dorsal anterior insula (dAI.L)	-34	22	0	BA 13
Right dorsal anterior insula (dAI.R)	38	22	-4	BA 13
Left ventral anterior insula (vAI.L)	-38	-2	-10	BA 13
Right ventral anterior insula (vAI.R)	44	6	-10	BA 13
Left posterior insula (PI.L)	-38	-18	14	BA 13
Right posterior insula (PI.R)	38	-14	16	BA 13
Left bed nucleus stria terminalis (BNST.L)	-6	2	0	
Right bed nucleus stria terminalis (BNST.R)	6	2	0	
Left nucleus accumben (NAc.L)	-10	12	-8	
Right nucleus accumben (NAc.R)	10	12	-8	
Left hippocampus (Hipp.L)	-34	-18	-16	
Right hippocampus (Hipp.R)	32	-14	-16	

JNeurosci Accepted Manuscript

Table 4. MNI Coordinates of amygdala and prefrontal cortex regions used in previous developmental studies of face emotion perception.

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

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.

JNeurosci Accepted Manuscript

Table 5. Age-related changes in amygdala activity in the PNC sample using coordinates from previous studies.

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

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.

JNeurosci Accepted Manuscript

Table 6. Age-related changes in amygdala connectivity in the PNC sample using coordinates from previous studies.

Connectivity	Emotion	*F (1, 754)*	*p*	*Effect Size*
AMY.R-vmPFC.R in Gee et al. (2013)↓	Neutral	0.210	0.647	0.017
	Fear	3.497	0.062	0.069
	Anger	1.880	0.171	0.051
	Sad	0.963	0.327	0.036
	Happy	0.288	0.591	0.020
AMY.L-dACC.L in Kujawa et al. (2016)↓	Neutral	0.008	0.929	-0.003
	Fear	8.921	0.003	0.109
	Anger	2.569	0.109	0.059
	Sad	0.388	0.533	0.023
	Happy	4.596	0.032	0.079
AMY.R-dACC.R in Kujawa et al. (2016)↓	Neutral	0.214	0.644	0.017
	Fear	6.602	0.010	0.094
	Anger	7.661	0.006	0.101
	Sad	4.420	0.036	0.077
	Happy	3.232	0.073	0.066
AMY.L-mPFC.L in Wu et al. (2016)↓	Neutral	0.012	0.914	-0.004
	Fear	4.139	0.042	0.075
	Anger	1.198	0.274	0.040
	Sad	0.003	0.958	0.002
	Happy	1.199	0.274	0.040
AMY.R-mPFC.L in Wu et al. (2016)↓	Neutral	0.334	0.563	0.021
	Fear	3.712	0.054	0.071
	Anger	7.566	0.006	0.101
	Sad	1.333	0.249	0.004
	Happy	2.743	0.098	0.061
AMY.L-ACC.L in Wu et al. (2016)↓	Neutral	0.019	0.890	0.005
	Fear	4.351	0.037	0.076
	Anger	3.000	0.084	0.064
	Sad	0.602	0.438	0.028
	Happy	2.769	0.097	0.061
AMY.R-ACC.R in Wu et al. (2016)↓	Neutral	0.013	0.909	0.004
	Fear	4.551	0.033	0.078

JNeurosci Accepted Manuscript

Anger	5.901	0.015	0.089
Sad	4.496	0.034	0.078
Happy	1.261	0.262	0.041

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.

JNeurosci Accepted Manuscript

Table 7. Replicability of previous developmental findings in amygdala activity and connectivity underlying face emotion perception.

Study	Findings	Replicability	
		amygdala activity	amygdala connectivity
Gee et al., 2013	Amygdala activity decreased with age; amygdala-vmPFC connectivity decreased with age	No	No
Kujawa et al., 2016	No age-related effect in amygdala activity; amygdala-dACC connectivity decreased with age in TD but increased with age in AD for all emotions	Partially Yes	No
Wu et al., 2016	No age-related effect in amygdala activity; amygdala-ACC/mPFC connectivity decreased with age for all emotions	Partially Yes	No