nature neuroscience

Article

https://doi.org/10.1038/s41593-023-01358-9

Large-scale encoding of emotion concepts becomes increasingly similar between individuals from childhood to adolescence

Received: 12 June 2022

Accepted: 12 May 2023

Published online: 08 June 2023

Check for updates

M. Catalina Camacho[1] ✉, Ashley N. Nielsen[2], Dori Balser[3], Emily Furtado[2], David C. Steinberger[3], Leah Fruchtman[3], Joseph P. Culver[1,4,5,6,7], Chad M. Sylvester[1,2] & Deanna M. Barch[1,2,3]

Humans require a shared conceptualization of others' emotions for adaptive social functioning. A concept is a mental blueprint that gives our brains parameters for predicting what will happen next. Emotion concepts undergo refinement with development, but it is not known whether their neural representations change in parallel. Here, in a sample of 5–15-year-old children (n = 823), we show that the brain represents different emotion concepts distinctly throughout the cortex, cerebellum and caudate. Patterns of activation to each emotion changed little across development. Using a model-free approach, we show that activation patterns were more similar between older children than between younger children. Moreover, scenes that required inferring negative emotional states elicited higher default mode network activation similarity in older children than younger children. These results suggest that representations of emotion concepts are relatively stable by mid to late childhood and synchronize between individuals during adolescence.

A critical social skill developed in childhood is the ability to discern and interpret emotions of other people. To aid in emotion perception and inference in real time, it is theorized that our brains use established concepts—blueprints for how specific interactions typically unfold—for prediction[1–4]. Behavioral work has indicated that concepts for how to glean emotion-related cues from the environment begin developing shortly after birth[5], although it is unclear how the brain develops these concepts or to what extent such concepts are shared across individuals. Indeed, separate studies have elucidated changes in brain structure and function across childhood and adolescence[6–8] and changes in activation to different emotions[9,10], none of which use contextualized stimuli that would activate naturalistic emotion concepts. Thus, to understand how children develop real-world emotion concepts, we must investigate how the brain encodes emotional cues with complete narrative/social context. Several questions remain. How are emotion concepts represented in the brain? Does this representation change across childhood and adolescence as children develop and refine their emotion reasoning skills? Identifying how emotion concept development and neurodevelopment are linked would provide fundamental insight into not only how emotional development occurs in conjunction with neuroplasticity but also when we may best intervene to improve health outcomes in children at risk for common psychiatric disorders that are associated with dysfunctions in emotion perception and inference.

Emotion perception and inference includes complex processes: detecting information or pertinent cues, attending to important or

[1]Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, MO, USA. [2]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [3]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA. [4]Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA. [5]Department of Physics, Washington University in St. Louis, St. Louis, MO, USA. [6]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA. [7]Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA. ✉e-mail: camachoc@wustl.edu

salient cues and interpreting and contextualizing those cues within our current state and past experiences[11,12]. Concepts are theorized to influence these processes by providing computational shortcuts, allowing our brains to make quick assessments and accurate predictions. When one considers what cognitive processes are implicated in real-world social interactions, we would expect that nearly every cognitive network would be necessary for these computations. For instance, consider the common experience of hearing an acquaintance tell you a fond story about your mutual friend. This interaction is expected to recruit primary sensory regions for visual and auditory processing[13,14]; the saliency, cingulo-opercular and attentional networks for alertness, cue detection and attention[15–17]; the default mode network for memory, self and relational processing[18,19]; and the frontoparietal network for working memory and behavioral organization[20]. If this is indeed how emotional cues are processed in the real world, we would expect there to be distinct brain-wide patterns for functionally distinct emotion concepts (for example, anger versus happiness) as they require distinct behavioral responses. Rather than finding patterns of activation to external emotion cues throughout the brain in line with this theory, small sample work examining how static emotional pictures[21,22] and positive and negative video content[23,24] are encoded in the brain has largely identified discrete regions or networks. Recent work using single-emotion movie trailers or film clips have found representations of emotion-specific information in primary sensory regions[22,25]. Real-world processing is more complex than single-valence clips can capture, however. Thus, a formal test of how the brain supports real-world emotion processing—using stimuli that are rich in both content and narrative, with multiple emotions presented at once across different characters—has yet to be conducted. For that, more narrative and dynamic stimuli are needed to activate real-world emotion concepts.

Behavioral work suggests that children develop emotion concepts in a broad to specific pattern rapidly across early childhood, learning to parse cues for sadness, anger, fear and happiness by around ages 5–8 years[26]. Alhough at a slower rate of change, there is evidence that emotion concepts continue to change across childhood and early adolescence, with rates of change in emotion inference abilities slowing markedly at around ages 15–17 years[27–29]. Most previous work examining developmental differences in neural activation to emotional stimuli have fewer than 30 individuals in their child group[30–36] and/or fewer than 10 individuals on average per year of age[37–39], and many report only amygdala region-of-interest analyses (for a review, see ref. 40). Although adults and older adolescents readily conjure what static emotional pictures represent, recent work has demonstrated that children rely upon social context and learned labels to be able to verbalize and identify emotions[41,42], consistent with the broader literature demonstrating that children rely on social information for learning[43]. Movie stimuli can, therefore, be used to provide full narrative context, enhancing ecological validity[44–46]. Although movies still differ from reality in important ways (for example, movies have scene cuts, follow a coherent narrative and can use scenery or lighting to add to the emotional cues), they offer a considerably more naturalistic index of emotion processing over single-valence clips or static images. Only two studies have examined activation to emotional stimuli using video stimuli in children, both in relation to emotional valence[23,24]. Both samples were small (fewer than 10 children on average per year), and neither examined activation to functionally distinct emotions, making it difficult to connect these studies to behavioral emotion development. Thus, the potential for using movies to examine emotion concept neurodevelopment has yet to be explored.

In this study, we leveraged a large cross-sectional pediatric dataset to characterize the brain network activation modulated by emotions and how they differ across development. Specifically, we aimed to characterize (1) how broad (positive and negative) and specific (angry, happy, sad, fearful and excited) emotion concepts are represented in the brain; (2) how these representations differ across development; and (3) what specific emotional contexts elicit shared and diverse activation responses with respect to developmental stage. Furthermore, we used a discovery and replication approach in all analyses to ensure generalization of our findings. We predict that emotion categories will have distinct representations in the brain, reflecting a shared functional understanding of these complex socio-emotional cues. Based on the behavioral emotion development literature, we expect changes across adolescence to reflect refinement of emotion processing, which would likely be reflected in change in higher-order cognitive networks that integrate sensory signals. It is possible that emotion processing changes in a manner that is not well captured by linear or curvilinear processes. For instance, activation may stabilize across development, resulting in older children having more similar activation patterns (convergence on a shared concept), or activation may diverge across development, reflecting individual differences in experiences or expectations. Thus, we examined the viability of converging or diverging models of development using similarity analysis in addition to examining linear and curvilinear patterns. Data were pre-processed using the Human Connectome Project minimal processing pipeline[47], and post-processing and analysis were carried out in Python (code available at https://github.com/catcamacho/hbn_analysis).

Results

Contextualized emotions are represented throughout the brain

Videos were processed using the EmoCodes system[48], a standardized coding system that we developed to characterize emotional/non-emotional and low-level/high-level information in cartoon videos, which cannot be characterized using existing algorithms designed for live-action stimuli. Using this system, we derived traces (adult ratings) of general (positive and negative) and specific (angry, happy, sad, fearful and excited) emotion content as well as potential low-level covariates (see Appendix A for full video characterization). We estimated activation to each contextualized emotion by convolving each emotion-specific timeseries (rescaled to a range of 0–1) with the hemodynamic response function and entering them into a general linear model (GLM) predicting each parcel's blood-oxygen-level-dependent (BOLD) signal. The general emotions activation model included positive emotion, negative emotion, brightness, loudness, spoken words and written words traces. The specific emotions activation model included angry, happy, sad, fearful and excited emotions as well as brightness, loudness, spoken words and written words traces. Presence of each of these video features varied widely across the videos, from 2.3% to 75.2% of the run time (Appendix A). Mean emotion activation maps (beta weights for each model) are shown in Fig. 1, and difference maps are in Extended Data Fig. 1. Parcel-level subject-level maps (encompassing 394 regions from 12 cortical networks and eight subcortical structures) were used as features in the support vector machine analyses.

To test if activation patterns were dissociable across emotions, we performed support vector classification analysis on the subject-level parcel activation maps (one per emotion per video) predicting emotion data labels. We trained each model on parcel-wise activation maps from the Discovery sample (n = 424 participants) with 10-fold cross-validation and tested on the Replication sample (n = 399 participants). Across both the general and specific emotion whole-brain models, activation to emotions was classified with high accuracy (general: 88% (chance = 50%) and specific: 73% (chance = 20%)), indicating that activation patterns to specific emotions are highly dissociable. Models performed significantly above chance even when non-emotion feature activation maps (brightness, loudness, spoken words and written words) were included alongside emotion activation maps for model fitting and testing (general: 76% accuracy (chance = 20%) and specific: 66% accuracy (chance = 11%)). Full model statistics are reported in Table 1. Follow-up permutation-based testing of the emotion-only models revealed that several networks were

https://doi.org/10.1038/s41593-023-01358-9

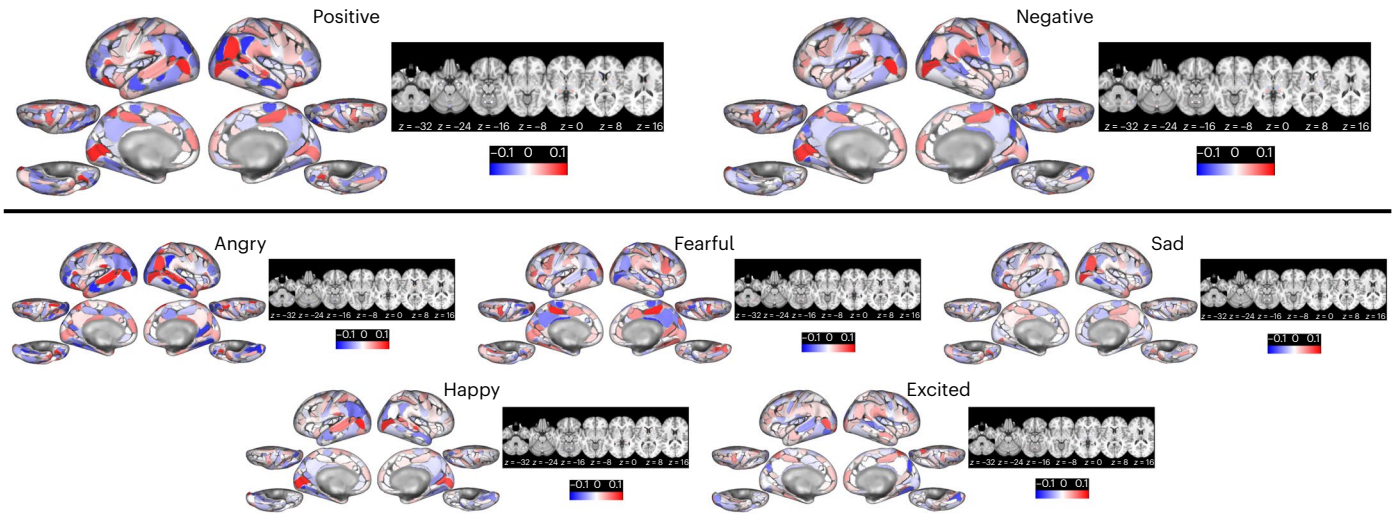


Fig. 1 | Mean activation (model coefficients or betas) to each emotion category across both videos. No statistical test was performed on these data directly. Top: General emotions. The general emotions were modeled as regressors in a GLM predicting BOLD signal alongside brightness, loudness, spoken words and written words. Bottom: Specific emotions. The specific emotions were modeled as regressors alongside brightness and loudness.

Table 1 | Emotion activation classification

Data labels predicting	Chance accuracy	Train accuracy	Train 95% CI	Test accuracy	Test 95% CI	Permuted *P* value
General emotion classification						
Positive and negative	50%	89%	(81%, 94%)	88%	(86%, 89%)	<0.001
Positive, negative, loudness, brightness, spoken words and written words	17%	81%	(73%, 84%)	76%	(74%, 77%)	<0.001
Specific emotion classification						
Angry, fearful, sad, happy and excited	20%	76%	(68%, 81%)	73%	(72%, 75%)	<0.001
Angry, fearful, sad, happy, excited, brightness, loudness, spoken words and written words	11%	66%	(64%, 76%)	66%	(65%, 67%)	<0.001

For each analysis, activation maps across the two videos were pooled. Models were trained on data from the RUBIC site (Discovery) using 10-fold cross-validation with participant ID entered as a grouping variable and tested on unseen data from the CBIC site (Replication). Support vector classification was highly accurate in distinguishing activation to each general and specific emotions across children, suggesting consistency in how activation to each concept is represented across 5–15 year olds.

informative to the model, with brain regions spanning primary sensory and higher-level cortex. Specifically, for the model predicting general emotion activation labels (positive and negative), permuting the Visual, Default Mode, Somatomotor-Hand and Cingulo-Opercular networks each affected the model classification accuracy more than random parcels of the same size of the other networks (Fig. 2a), suggesting that these networks have unique emotion information relative to other brain regions/networks. Interestingly, for the specific emotion model (angry, excited, fearful, happy and sad), the Visual network had the most unique emotion-specific information above and beyond random regions, followed by the Dorsal Attention, Ventral Attention, Cingulo-Opercular, Somatomotor-Hand, Cerebellum, Fronto-Parietal, Temporo-ventromedial prefrontal cortex (VMPFC), Auditory and Default Mode networks (Fig. 2b).

Systematically permuting data from each of these regions/networks did not lower accuracy to chance levels, however, suggesting that there was still sufficient information throughout the rest of the brain to make accurate classifications. To directly examine which networks/regions contained sufficient information to classify emotion activation maps at above chance levels, we systematically limited Training and Testing data to single networks/regions at a time and computed confidence intervals (CIs) using a bootstrapping approach. When the activation data were systematically limited to each region/network, we found that most of the brain had some emotion information represented, with parcel activation from 15 of 21 regions/networks able to predict general emotion activation labels significantly above chance: Visual, Temporo-VMPFC, Fronto-Parietal, Auditory, Ventral Attention, Default, Dorsal Attention, Cingulo-Opercular, Cerebellum, Somatomotor-Hand, Medial Parietal, Caudate, Somatomotor-Mouth, Parietal-Occipital and the Thalamus. For the model predicting specific emotion activation labels, activation from each of 14 of 21 networks/regions alone was able to classify emotion activation labels better than chance: Visual, Temporo-VMPFC, Cingulo-Opercular, Ventral Attention, Dorsal Attention, Auditory, Default Mode, Somatomotor-Hand, Fronto-Parietal, Cerebellum, Medial Parietal, Somatomotor-Mouth and Salience. Together, these results indicate that emotional content is processed throughout the brain in children, with the most unique information represented in primary sensory visual and auditory regions as well as higher-order associative networks that span the frontal, temporal and parietal lobes. As a follow-up, we repeated these analyses using data for each video separately. Those results are reported in Appendix B and largely replicate the main analyses.

Activation patterns are relatively stable across development

To test if activation to general or specific emotions differed across maturity, we performed support vector regression predicting chronological

https://doi.org/10.1038/s41593-023-01358-9

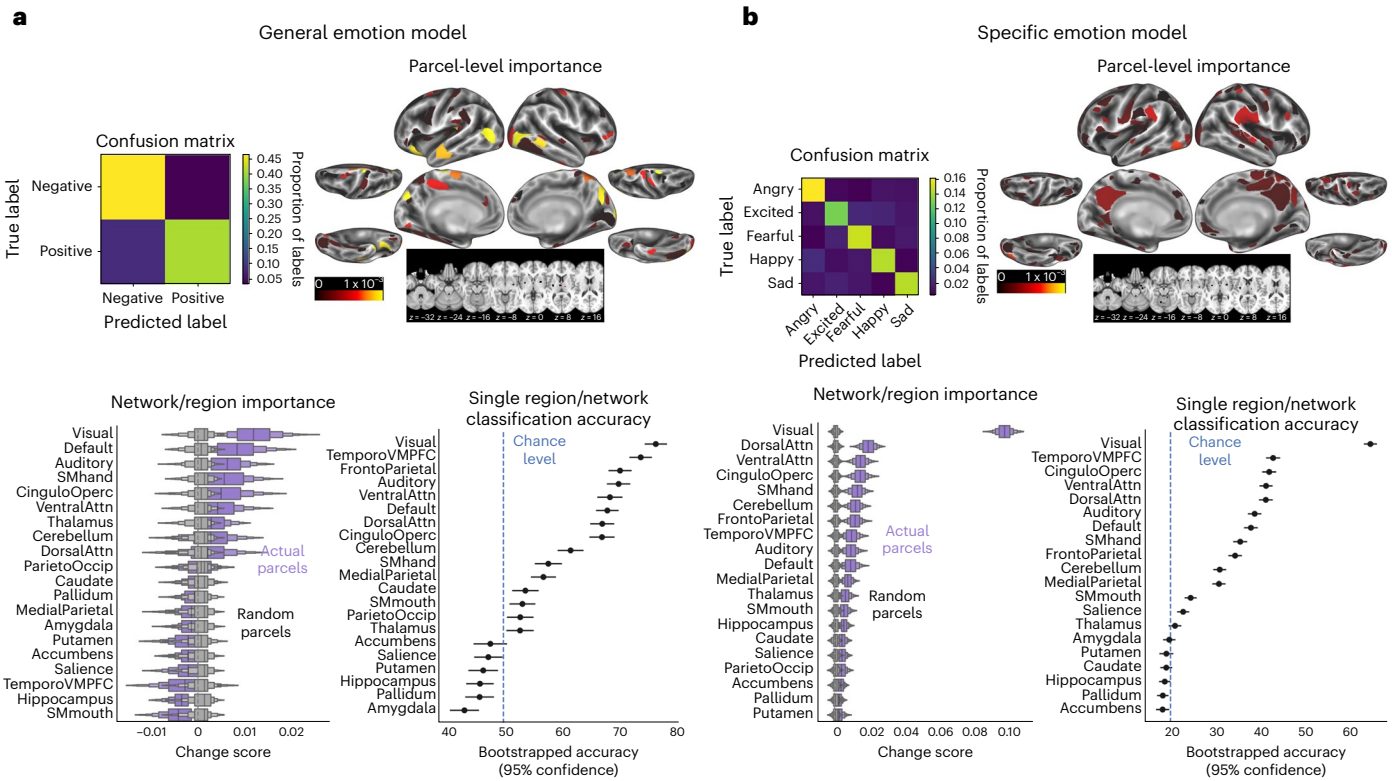


Fig. 2 | Activation classification analysis results. Activation classification analysis results for general emotions (Negative and Positive, **a**) and for specific emotions (Angry, Happy, Sad, Fearful and Excited, **b**). Top left: confusion matrix indicating the proportion of correct data labels and incorrect data labels for each emotion category in the testing (unseen) dataset. Top right: Parcel-level mean importance scores derived from permutation-based analysis. Higher values indicate that permuting that data label resulted in a decrease in model accuracy. Note that no one parcel permutation causes the model accuracy to drop below chance. Bottom left: network/region-level permutation-based importance testing results. Null distributions were generated by permutation random parcels of the same *n* as the target network/region. Boxen plots include the median as the center of the distribution with each successive level outwards including 50% of the remaining data in the distribution. Bottom right: mean classification accuracy when train/test data are limited to a single region/network. Limiting data did not improve model performance for either General or Specific emotion classification. The 95% CIs were estimated using a pseudo-bootstrap method in which a subset of the testing data was used to estimate accuracy. This procedure was repeated 10,000 times, and the resulting distribution was used to estimate the CI. SM, Somatomotor.

age and puberty scores from each emotion activation map. Across all models, test sample accuracy was poor as indicated by high mean squared error (MSE; 9.27–12.16 for age and 22.21–27.60 for puberty) and modest correlations between actual and predicted labels (0.07–0.19), suggesting a weak linear association between maturity indices and activation, with no meaningful differences between age and puberty model performance. Interestingly, we found similar model fits for activation to non-emotional content (brightness, loudness, spoken words and written words), suggesting modest differences in global processing across development that is not specific to a single cue modality. Full results are reported in Supplementary Table SC2. Using limited data or a curvilinear kernel did not meaningfully improve model performance (see Supplementary Table SC2 for full results).

Post hoc examination of parcel-wise correlations with maturity confirmed a modest linear association between maturity and activation (Extended Data Fig. 2), with maturity explaining, at most, 3.8% of the variance in parcel-level activation (Age-Activation r^2 range: Negative 0–0.017, Positive 0–0.030, Angry 0–0.020, Happy 0–0.026, Sad 0–0.023, Excited 0–0.028, Fearful 0–0.023; Puberty-Activation r^2 range: Negative 0–0.029, Positive 0–0.030, Angry 0–0.024, Happy 0–0.038, Sad 0–0.014, Excited 0–0.030, Fearful 0–0.021).

Synchronization of emotion activation across development

It is possible that linear/curvilinear models are not able to fully capture the changes associated with processing complex stimuli (for example, if activation variability changes across development rather than magnitude of signal); thus, we also tested if we could identify emotion processing-related differences across maturity using a model-free approach. Specifically, we sought to test if (1) there were differences across maturity not captured by linear/curvilinear models and (2) if specific scenes elicited lesser or greater similarity within older/younger children. To accomplish this, we used inter-subject representational similarity and tested if similarity in activation across the video (inter-subject correlation (ISC)) corresponded to one of three different models of cognitive affective development: nearest neighbor (are children proximal in maturity processing the video similarly?), convergence (do children process the video more similarly as they mature?) and divergence (do children process the video less similarly as they mature?). We found the most evidence for convergence in parcel-wise activation across both age and puberty, with age fitting the data better than puberty as indicated by more significant parcels and higher similarity coefficients. Specifically, for each *Despicable Me* and *The Present*, the convergence models (that is, that activations among older children exhibit more similarity than activations among younger children) had the most significant parcels and higher similarity coefficients after false discovery rate (FDR) correction (*Despicable Me* Age: 176 parcels, 0.01–0.16 rho; *The Present* Age: 150 parcels, 0.01–0.13 rho; *Despicable Me* Puberty: 38 parcels, 0.02–0.08 rho; *The Present* Puberty: 84 parcels, 0.02–0.10 rho). Regions highest in similarity were primarily found in occipital, lateral parietal and temporal cortex, similarly

https://doi.org/10.1038/s41593-023-01358-9

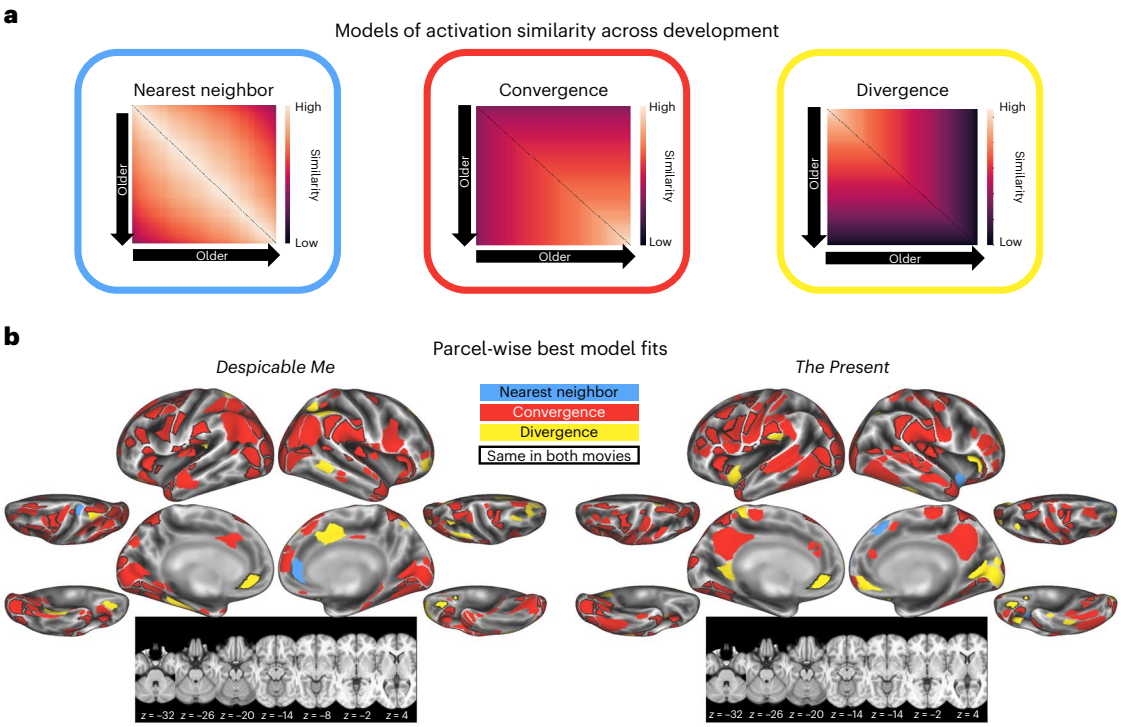


Fig. 3 | Similarity analysis results. a, Similarity models of development examined for each movie and for each maturity metric. **b**, Best fit models for each parcel using chronological age as the metric of maturity. Parcels were considered significant if the similarity coefficient (Spearman *r*, one-sided) was FDR-corrected $P < 0.05$ in both the Discovery and Replication samples. Model fits were determined by generating a distribution of similarity coefficients (through bootstrapping) and comparing distributions for each model, pairwise (two-sided *t*-test with a permutation-based *P* value). For each parcel, the model with the highest ranking of similarity coefficients was determined to be the best fit. Outlined parcels are those that were the same model designation across both movies.

to the group-level ISC maps for each video (Extended Data Fig. 3a). Similarity statistical maps are shown in Extended Data Fig. 3b–d. We compared model fits for each parcel using a bootstrapping approach. The results of this analysis are shown in Fig. 3. Parcels that replicated across samples for each movie are shown with parcels that replicated across movies outlined. Across significant parcels, the convergence model was predominantly the best fit, indicating that older children had more similar activation patterns to each other when viewing the videos than did younger children.

Emotional intensity is associated with increased synchrony

We next sought to examine what specific scenes or content elicited increased synchrony in the older age group. Because the convergence model of age best fit most parcels in the brain, we then followed up this analysis by examining dynamic synchrony (inter-subject phase synchrony (ISPS)) both across the full sample and within the oldest children (top 20% by age). Significant parcels were grouped into networks that were found to encode emotion-specific information (visual, auditory, dorsal attention, ventral attention, cingulo-opercular, default mode and fronto-parietal) and averaged across participants for group-level peak analysis. We identified scenes at least 5 s long of high network-level synchrony across the sample (Fig. 4a and Extended Data Fig. 4). Differences in high and low synchrony scenes were examined, revealing that scenes higher in emotional intensity elicited increased synchrony across the sample and across movies in the visual, ventral attention and default mode networks.

Default mode network synchronization across development

A side-by-side comparison of full group dynamic analysis results and the results from repeating the analysis in the oldest children only for each video indicated yielded three observations. First, although the overall range of synchrony was higher for the oldest children, the specific scenes that elicited higher synchrony largely overlapped with scenes derived from the full sample dynamic synchrony traces for the visual, auditory, dorsal attention, ventral attention and cingulo-opercular networks. The same number of scenes elicited high synchrony in the oldest children as in the full sample for both videos in the fronto-parietal network, although the specific scenes did not overlap. Because few parcels from the frontoparietal network were significant from the inter-subject representational similarity analysis (IS-RSA), we did not interpret results from this network further. Finally, the default mode network demonstrated the largest relative difference in synchrony between the oldest children and the full sample. Specifically, three additional scenes evoked higher synchrony in the oldest children that did not appear in the full sample analysis. Results from this analysis are shown in Fig. 5 and Extended Data Figs. 5 and 6.

Quantitative analysis of the video features under the peaks and outside of the peaks indicated that increased synchrony across the whole sample and the oldest children was induced during scenes depicting negative emotion. Qualitative analysis (Fig. 6) of the scenes indicated that the three *Despicable Me* scenes that induced increased synchrony in the older children, and were not identified in the full sample, were those with multiple layers of implicit emotional information. For example, the scene with the highest peak synchrony of the three is during the rocket launch, when the minions place a ticket to a ballet recital in Gru's pocket, and Gru quickly gets upset with the minions in response. To understand why Gru reacts so strongly in this scene, the viewer would have to have understood the inner turmoil that Gru experienced in the previous scenes. Thus, it is possible that this analysis is capturing development in the interpretation and integration stages of emotion processing (from the social information processing model[11]) represented in converging activation of the default mode network.

https://doi.org/10.1038/s41593-023-01358-9

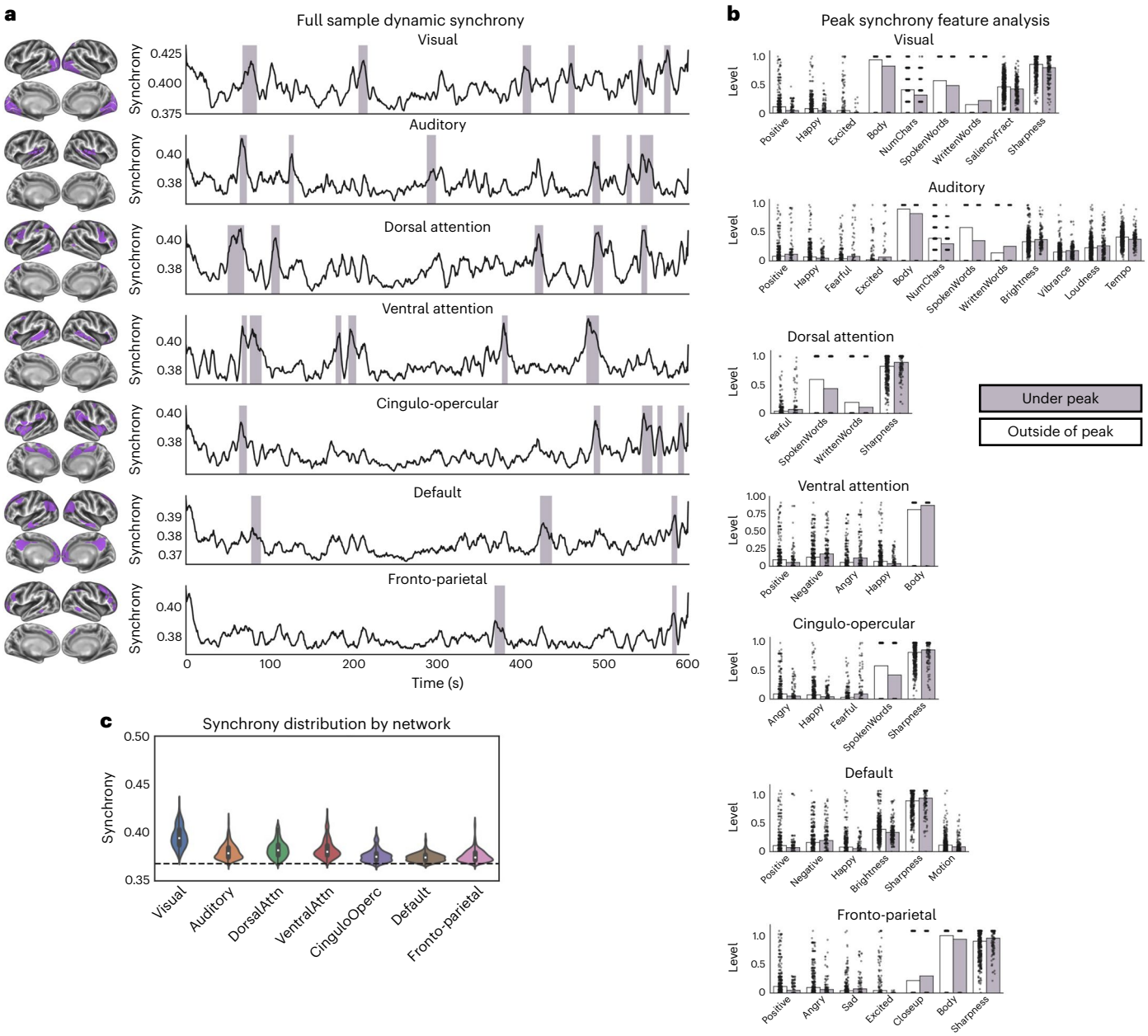


Fig. 4 | Group-level dynamic synchrony results for *Despicable Me*. **a**, Dynamic synchrony across the full sample with replicating peaks in synchrony shaded purple. Replicating peaks were defined as peaks at least 5 s wide and with a prominence higher than the 95th percentile value for that network (permutation-based $P < 0.001$) appearing in both Discovery and Replication samples. Parcels were limited to those that were significantly correlated across the sample at the group level after FDR correction. **b**, Video feature means within the peaks were compared to features outside of the peaks using a two-sided *t*-test to test if specific video features elicited increased synchronization. Plotted features were those that were significantly different (two-sided *t*-test, FDR-corrected $P < 0.05$). Bar plots indicate mean values with overlaid dots of individual data points. **c**, Violin plots of mean synchrony values by network. White dots indicate median value; box indicates the 50% interquartile range; and whiskers indicate each the upper and lower 25%. The dashed horizontal line indicates the value at permuted $P < 0.05$ after FDR correction. NumChars, number of characters; SaliencyFract, fraction of frame containing highly salient content.

Discussion

In this study, we provide, to our knowledge, the first empirical evidence that contextualized emotion processing—which activates emotion concepts—is represented throughout the brain in middle childhood through mid-adolescence. We demonstrate that activation to each general emotion (positive and negative) and specific emotion (angry, happy, sad, fearful and excited) cues are highly dissociable, indicating well-established concept representation of these emotions in this age range. We demonstrate through multiple approaches that concept representation shifts modestly across age and puberty, with some evidence that older children show more similar patterns of activation to video stimuli to each other than do younger children. Dynamic analysis revealed that this effect was most profound for the default mode network, with scenes depicting negative emotions evoking increasingly similar activation patterns with age. Taken together, our work suggests that, across late childhood and early adolescence, there is fine-tuning of pre-existing emotion concepts. These findings have important implications for not only how our brain processes real-world emotions but also how to conceptualize risk for emotion dysregulation and the design of psychosocial interventions.

https://doi.org/10.1038/s41593-023-01358-9

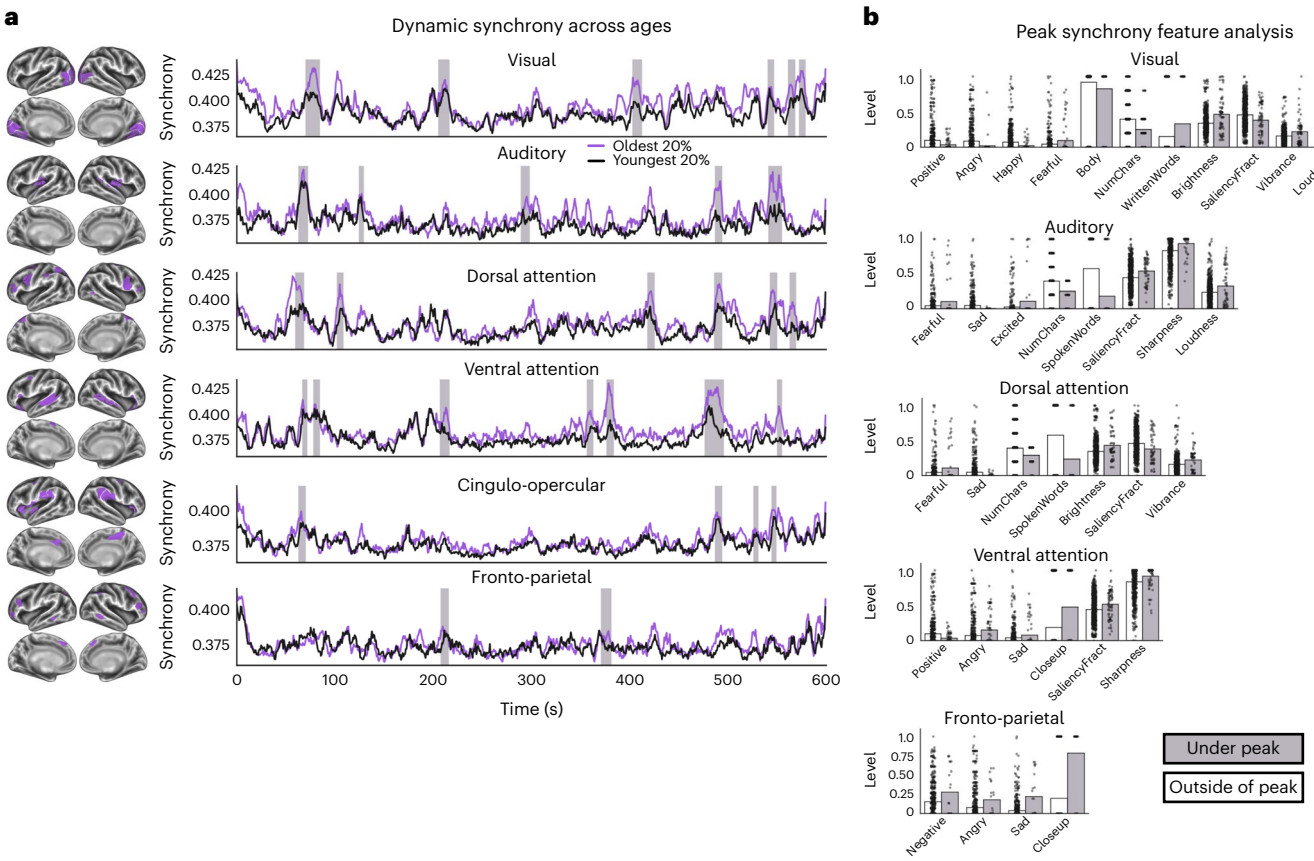
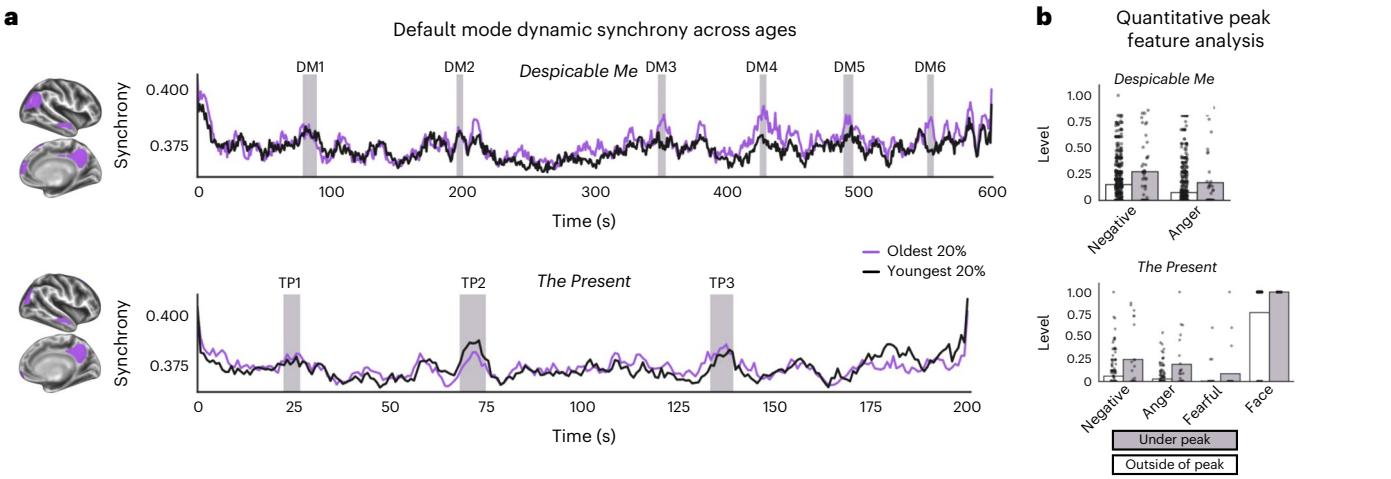


Fig. 5 | Dynamic synchrony analysis results in the oldest children for *Despicable Me*. **a**, Network dynamic activation similarity (synchrony) for each of the oldest and youngest children in the sample. Included parcels are shown to the left of each trace. Shaded areas denote significant increases in synchrony in the oldest children (1.5 s.d. above the mean). **b**, Bar plots of the results from the video feature analysis comparing portions of the video within peaks of inter-subject synchrony to outside the peaks using a two-sided *t*-test. Bar plots are of the mean with individual datapoints overlaid. Only features that significantly differed (two-sided FDR-corrected $P < 0.05$) are plotted.

We found evidence that contextualized emotion-specific cues are encoded throughout the brain. We were able to classify specific and broad emotion concepts with high levels of precision. This suggests that there are both shared and distinct activation features of each broad and specific emotions, likely capturing shared and distinct functions. Notably, every cortical network had some level of emotion-specific activation, supporting the social information processing model of emotion processing presented in the introduction[11] as well as extensions of this framework[49]. These frameworks use cognitive models of emotion processing[50,51], which propose that emotions are interpreted as context-dependent concepts (for example, anger may be indicated by any number of signals) and are, therefore, not localized to any one network or region. We observed more widespread activation pattern differences than previous work conducted using single-emotion/valence narrative stimuli[22,25,52], replicating and extending their findings. This is likely due to the additional cognitive processes involved in understanding contextualized emotions, resulting in a broader range of moment-to-moment possibilities that the brain may predict. Thus, it is possible that we are capturing this complex cognition necessary for real-world identification of specific emotions in others in our results. Our results can also be interpreted to support discrete or basic emotion theories, which posit that there is a finite number of core emotions each with unique attributes (see ref. 53 for a synthesis). One way of interpreting these theories in the context of neural computation would be that each core emotion has a unique expression and, therefore, would evoke a unique activation pattern regardless of the context. Although we did find unique activation patterns to each emotion that we examined, we did not define our emotion regressors assuming that emotions have unique behavioral cues. Instead, we used functional emotion definitions that emphasize context in combination with behavioral cues. More systematic examination of real-world emotion perception is needed to fully disentangle context-dependent versus context-independent activation to emotion cues. Another interpretation is that discrete or 'basic' emotions exist at a level in brain circuitry that is shared across species[54]. By this logic, expression of these emotions would have distinct presentations, resulting in distinct activation of sensory cortex. Although we did observe strong emotion-specific signals in sensory cortex in our results, we also observed differences across the rest of the brain, suggesting that emotion context is not only a difference of sensory input. Emotion-specific information is also modulating activation of higher-order cognitive networks, in line with the notion of constructed emotion concepts rather than of discrete emotion configuration.

Another interpretation of our results is that it is possible that activation to naturalistic emotion cues is innate to some degree. Specifically, the fact that these activation patterns were so consistent across the sample with relatively subtle changes across age suggests that the basic architecture for processing social emotional cues may be present very young, refining quickly across childhood. More research is needed to tease apart if our results represent more constructionist notions of emotion, discrete emotion theory or something in between. Specifically, future work could compare annotations of several longer video clips based on functional definitions (as used here) and discrete expressions and systematically test which coding schemes capture

https://doi.org/10.1038/s41593-023-01358-9



c High synchrony scene descriptions

DM1. Gru balks at the book the girls wanted him to read to them before acquiescing.

DM2. The girls lie down to sleep deciding that they like Gru after all right before we see Gru in the hall, thinking about his feelings toward the girls.

DM3. Dr. Nefario walks up behind Gru right after Miss Hattie arrived to take the girls. Dr. Nefario nods to Gru who, resigned, allows Miss Hattie to take the girls away.

DM4. A montage of Gru missing the girls and the girls appearing scared/sad at the orphanage set to sad music.

DM5. Gru gets suddenly upset at the minions who placed a ballet recital ticket in his pocket right before the rocket launch.

DM6. A civilian bystander looks up in awe. Vector waits for the squid grappling hook that he just fired to pull him up into the sky to try to thwart Gru's mission.

TP1. A boy plays video games as a woman walks into the house holding a large box and says, 'Hey sweetie!'. The boy does not immediately react or respond.

TP2. The boy's face falls into scowl as he sees that the puppy has a short leg. He throws the puppy down and smacks the box aside, visibly upset.

TP3. The puppy frowns after tossing the box off and then turns and barks at the empty box. The boy chuckles and then abruptly switches back to being upset.

Fig. 6 | Qualitative examination of the scenes eliciting higher synchrony in the oldest children for each video. a, Dynamic synchrony in the oldest and youngest children for each video. Shaded areas are peaks of increased synchrony in the oldest children. **b**, Bar plots of the results from the video feature analysis comparing portions of the video within peaks of inter-subject synchrony to outside the peaks using a two-sided *t*-test. Bar plots are of the mean with individual datapoints overlaid. Only features that significantly differed (two-sided FDR-corrected $P < 0.05$) are plotted. **c**, Descriptions of each scene, all of which included negative emotional inference.

more variance in the activation data across ages. Matching these data to individual differences in emotion reasoning would also help clarify how emotion cues are represented in the brain across development.

We found evidence for only a modest change in activation across age, suggesting that the core functional architecture of how contextualized emotion cues are interpreted is already present by the time children enter school. Furthermore, similarity analysis revealed that the modest differences across age observed were best characterized as converging similarity, with older children having more similar activation patterns than did younger children. This pattern is consistent with behavioral work, which has shown that, although children continue to develop a shared understanding of what isolated faces represent well into late childhood and early adolescence[40], young children are able to readily distinguish happy, angry, sad and scared emotional states in others when provided narrative context[41,55–57]. It is, therefore, likely then that the convergence pattern in the current study indicates that existing shared emotion concepts are refined with increasing age. Notably, the fact that we observed this convergence across regions of nearly every cortical network suggests that refinement may not be limited to single facets of emotion stimuli but, rather, the overall gestalt. Interestingly, puberty was not as strongly associated with the neural activation data. Previous work found an association between pubertal status and emotion processing[29,58,59]; however, this research did not find a consistent association between puberty and activation to specific emotions or an association that holds above and beyond associations with age[60]. Thus, our data suggest that activation to emotional content in videos is likely capturing the effects of accumulation of life experience on emotion processing development rather than pubertal influences. Our findings provide support for the constructionist theory of emotional development, which posits that children develop shared concepts of socially embedded emotions through socialization, language learning and improved integration of multi-sensory information in the brain[61], which enhances children's abilities to anticipate the consequences of emotions in the real world. Because we used functional definitions when identifying emotional content and found consistent patterns of activation of early-learned emotion concepts across age (and less change in relation to puberty than age), our results are in line with this theory.

We used adult ratings of functional emotion cues for these data. This was done following standard procedures for examining emotion reasoning development, where studies typically use the adult judgments as the benchmark against which to judge child judgments[5,26]. Although this methodological decision enables us to use this literature to interpret our findings, we would expect individual differences in emotion cue perception to color how these cues are encoded. For instance, there is evidence to suggest that school-age children (who make up the younger end of our age range) tend to accurately label emotion cues consistently, assigning a single emotion label to a given situation[26]. Across late childhood and adolescence, however, children are increasingly likely to assign multiple possible emotion labels to the same set of cues, reflecting a better understanding of individual differences in experience[27]. Although work in adolescents could be conducted to tease apart these differences in youth versus adult judgments, it would be difficult to examine age-related differences in emotion cue labeling in young children who may not be as adept at talking about emotions at an abstracted level (what emotion is communicated versus what the child feels, for example). Further work is needed to develop ways to acquire these data in young children to test differences in activation related to adult-defined labels versus individual-defined labels.

Similarity analysis from the longer video clip, *Despicable Me*, revealed a striking difference in which scenes elicited greater similarity

in default mode activation between the full sample and in the oldest children. Intriguingly, scenes that required negative emotional inference along a longer timescale (that is, remembering previously inferred emotional states) induced more similar activation in older children. Our findings provide a theoretical extension of previous work examining mentalizing network—a subnetwork of the default mode network[18]—development in 3–12-year-old children. Specifically, previous work has shown that default mode network activation to movie scenes that require inference about a character's emotional state increases with age and with mentalizing skills[62]. Across childhood and adolescence, the default mode network becomes increasingly modular, with increasing intra-network connectivity and decreasing connectivity with other networks[62,63]. Previous work examining a large sample of 5–12-year-old children found that this change in network connectivity was driven by increasing connectivity between the parietal/temporal network nodes and the prefrontal nodes[63], major hubs of not only the default mode network but also of the brain more broadly[64]. Within the context of work conducted in adults[64–67], it is likely that this change in default mode network modularity across development is indicative of improved integration of multi-modal information across the brain. Taken together, it is likely that differences across age in activation to contextualized emotional stimuli that we observed in the default mode network reflect a combination of increased efficiency in multi-modal inferences across development and cognitive improvements in emotion inference skills. Considering how commonly default mode network dysfunction is observed across psychiatric disorders[68–71] and the default mode network's proposed role in emotion experience[19], fully mapping how changes in default mode network activation and connectivity shift with emotional development would provide valuable insight into how emotion dysregulation emerges. Although we did not have measures of subjective experience in this study, previous work in adults has found that activation patterns in default mode network encodes subjective ratings of valence and arousal[19,72,73], whereas studies examining discrete emotion states have found brain-wide differences in activation during discrete states[74–76]. This suggests that some of the variance in brain-wide activation that we observed is likely individual differences in how these emotion cues made the children feel. Using facial expressions rather than relying solely on individual ratings, a recent study identified the VMPFC as a region capturing individual experiences during movie watching[77]. Interestingly, this region is also one of the few regions that we observed to become less synchronized in activation in older children compared to younger children. Thus, it is possible that the decrease in synchronization across age reflects a greater influence of individual differences in affective experience during the videos in older children. Further work is needed to tease apart how individual differences in subjective experience during movie watching relate to how emotion cues are encoded in the brain across development.

This study has several major strengths, including (1) a large sample size, which is critical for assessing general trends across development; (2) use of discovery and replication cohorts in addition to FDR correction to maximize replicability and generalization; (3) use of multiple video stimuli; and (4) careful video analysis and reporting, including of both high-level (emotions, objects, etc.) and low-level (brightness, loudness, etc.) video features. A critical limitation of this study is that it is not longitudinal; thus, these findings must be interpreted as associations with maturity, not development. Although we speculate that associations with maturity are reflecting developmental processes, emotional development is likely highly specific to the individual because our previous experiences shape how we interact with others in the future. Thus, these trends should be confirmed in large, longitudinal and richly characterized samples to better map how humans experience affective neural activation to real-world emotions. Additionally, although cartoons are more naturalistic stimuli than single-valence clips or pictures, they are imperfect proxies for real-life emotion perception. Specifically, cartoon videos often exaggerate emotion states in characters or manipulate lighting and color palettes to evoke a particular emotion, neither of which is consistent with reality. Nonetheless, this work provides a substantial step forward to understanding complex and contextualized emotion perception. Another important note is that, although there were no associations between non-emotional video features and emotional content, we did not systematically vary audio-visual features and emotional content. We, therefore, cannot completely disentangle non-emotional sensory processing from emotion perception. Finally, we do not have measures of the children's subjective experiences or behaviors during the videos. Future work should seek to include measures of subjective experience to better understand how each child's experiences color their perception of external emotion cues.

We present empirical evidence that specific contextualized emotion cues are encoded throughout the cortex, cerebellum and caudate. We showed that activation to contextualized emotion stimuli was fairly stable across development, with modest changes suggesting improved integration of multi-modal information as children refine their shared conceptual understanding of emotion cues. This work has important implications for models of affective neurodevelopment. Further work is needed to longitudinally confirm these trends and to map how they may go awry in individuals with socio-emotional dysregulation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-023-01358-9.

References

1. Alba, J. W. & Hasher, L. Is memory schematic? *Psychol. Bull.* **93**, 203–231 (1983).
2. Masís-Obando, R., Norman, K. A. & Baldassano, C. Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *eLife* **11**, e70445 (2022).
3. van Kesteren, M. T. R., Ruiter, D. J., Fernández, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
4. Darley, J. M. & Fazio, R. H. Expectancy confirmation processes arising in the social interaction sequence. *Am. Psychol.* **35**, 867–881 (1980).
5. Ruba, A. L. & Pollak, S. D. The development of emotion reasoning in infancy and early childhood. *Annu. Rev. Dev. Psychol.* **2**, 503–531 (2020).
6. Mills, K. L. et al. Structural brain development between childhood and adulthood: convergence across four longitudinal samples. *Neuroimage* **141**, 273–281 (2016).
7. Giedd, J. N. & Rapoport, J. L. Structural MRI of pediatric brain development: what have we learned and where are we going? *Neuron* **67**, 728–734 (2010).
8. Grayson, D. S. & Fair, D. A. Development of large-scale functional networks from birth to adulthood: a guide to the neuroimaging literature. *Neuroimage* **160**, 15–31 (2017).
9. Malsert, J., Palama, A. & Gentaz, E. Emotional facial perception development in 7, 9 and 11 year-old children: the emergence of a silent eye-tracked emotional other-race effect. *PLoS ONE* **15**, e0233008 (2020).
10. Batty, M. & Taylor, M. J. The development of emotional face processing during childhood. *Dev. Sci.* **9**, 207–220 (2006).
11. Lemerise, E. A. & Arsenio, W. F. An integrated model of emotion processes and cognition in social information processing. *Child Dev.* **71**, 107–118 (2000).

12. Crick, N. R. & Dodge, K. A. A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychol. Bull.* **115**, 74–101 (1994).
13. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417 (1983).
14. Dehaene-Lambertz, G., Hertz-Pannier, L. & Dubois, J. Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends Neurosci.* **29**, 367–373 (2006).
15. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 215–229 (2002).
16. Sadaghiani, S. & D'Esposito, M. Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cereb. Cortex* **25**, 2763–2773 (2015).
17. Barrett, L. F. & Satpute, A. B. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Curr. Opin. Neurobiol.* **23**, 361–372 (2013).
18. Buckner, R. L. & DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).
19. Satpute, A. B. & Lindquist, K. A. The default mode network's role in discrete emotion. *Trends Cogn. Sci.* **23**, 851–864 (2019).
20. Marek, S. & Dosenbach, N. U. F. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci.* **20**, 133–140 (2018).
21. Lindquist, K. A. & Barrett, L. F. A functional architecture of the human brain: emerging insights from the science of emotion. *Trends Cogn. Sci.* **16**, 533–540 (2012).
22. Kragel, P. A., Reddan, M. C., LaBar, K. S. & Wager, T. D. Emotion schemas are embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358 (2019).
23. Park, A. T. et al. Early stressful experiences are associated with reduced neural responses to naturalistic emotional and social content in children. *Dev. Cogn. Neurosci.* **57**, 101152 (2022).
24. Gruskin, D. C., Rosenberg, M. D. & Holmes, A. J. Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. *Neuroimage* **216**, 116217 (2020).
25. Camacho, M. C., Karim, H. T. & Perlman, S. B. Neural architecture supporting active emotion processing in children: a multivariate approach. *Neuroimage* **188**, 171–180 (2019).
26. Widen, S. C. Children's interpretation of facial expressions: the long path from valence-based to specific discrete categories. *Emot. Rev.* **5**, 72–77 (2013).
27. Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A. & Somerville, L. H. The nonlinear development of emotion differentiation: granular emotional experience is low in adolescence. *Psychol. Sci.* **29**, 1346–1357 (2018).
28. Nook, E. C. et al. Charting the development of emotion comprehension and abstraction from childhood to adulthood using observer-rated and linguistic measures. *Emotion* **20**, 773–792 (2020).
29. Motta-Mena, N. V. & Scherf, K. S. Pubertal development shapes perception of complex facial expressions. *Dev. Sci.* **20**, e12451 (2017).
30. Thomas, K. M. et al. Amygdala response to facial expressions in children and adults. *Biol. Psychiatry* **49**, 309–316 (2001).
31. Wiggins, J. L. et al. Developmental differences in the neural mechanisms of facial emotion labeling. *Soc. Cogn. Affect. Neurosci.* **11**, 172–181 (2016).
32. Marusak, H. A., Carré, J. M. & Thomason, M. E. The stimuli drive the response: an fMRI study of youth processing adult or child emotional face stimuli. *Neuroimage* **83**, 679–689 (2013).
33. Ladouceur, C. D., Schlund, M. W. & Segreti, A.-M. Positive reinforcement modulates fronto-limbic systems subserving emotional interference in adolescents. *Behav. Brain Res.* **338**, 109–117 (2018).
34. Hoehl, S., Brauer, J., Brasse, G., Striano, T. & Friederici, A. D. Children's processing of emotions expressed by peers and adults: an fMRI study. *Soc. Neurosci.* **5**, 543–559 (2010).
35. Haller, S. P. et al. Reliability of neural activation and connectivity during implicit face emotion processing in youth. *Dev. Cogn. Neurosci.* **31**, 67–73 (2018).
36. Lobaugh, N. J., Gibson, E. & Taylor, M. J. Children recruit distinct neural systems for implicit emotional face processing. *Neuroreport* **17**, 215–219 (2006).
37. Guyer, A. E. et al. A developmental examination of amygdala response to facial expressions. *J. Cogn. Neurosci.* **20**, 1565–1582 (2008).
38. Pagliaccio, D. et al. Functional brain activation to emotional and nonemotional faces in healthy children: evidence for developmentally undifferentiated amygdala function during the school-age period. *Cogn. Affect. Behav. Neurosci.* **13**, 771–789 (2013).
39. Hare, T. A. et al. Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biol. Psychiatry* **63**, 927–934 (2008).
40. Somerville, L. H., Fani, N. & McClure-Tone, E. B. Behavioral and neural representation of emotional facial expressions across the lifespan. *Dev. Neuropsychol.* **36**, 408–428 (2011).
41. Widen, S. C. & Russell, J. A. Children acquire emotion categories gradually. *Cogn. Dev.* **23**, 291–312 (2008).
42. Widen, S. C. & Russell, J. A. Children's scripts for social emotions: causes and consequences are more central than are facial expressions. *Br. J. Dev. Psychol.* **28**, 565–581 (2010).
43. Wu, Y., Schulz, L. E., Frank, M. C. & Gweon, H. Emotion as information in early social learning. *Curr. Dir. Psychol. Sci.* **30**, 468–475 (2021).
44. Cantlon, J. F. The balance of rigor and reality in developmental neuroscience. *Neuroimage* **216**, 116464 (2020).
45. Vanderwal, T., Eilbott, J. & Castellanos, F. X. Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. *Dev. Cogn. Neurosci.* **36**, 100600 (2019).
46. Eickhoff, S. B., Milham, M. & Vanderwal, T. Towards clinical applications of movie fMRI. *Neuroimage* **217**, 116860 (2020).
47. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
48. Camacho, M. C. et al. EmoCodes: a standardized coding system for socio-emotional content in complex video stimuli. *Affect. Sci.* **3**, 168–181 (2022).
49. Sander, D., Grandjean, D. & Scherer, K. R. An appraisal-driven componential approach to the emotional brain. *Emot. Rev.* **10**, 219–231 (2018).
50. Pessoa, L. Understanding emotion with brain networks. *Curr. Opin. Behav. Sci.* **19**, 19–25 (2018).
51. Lindquist, K. A. & MacCormack, J. K. Comment: Constructionism is a multilevel framework for affective science. *Emot. Rev.* **6**, 134–135 (2014).
52. Skerry, A. E. & Saxe, R. Neural representations of emotion are organized around abstract event features. *Curr. Biol.* **25**, 1945–1954 (2015).
53. Tracy, J. L. & Randles, D. Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emot. Rev.* **3**, 397–405 (2011).
54. Panksepp, J. & Watt, D. What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emot. Rev.* **3**, 387–396 (2011).
55. Ogren, M. & Johnson, S. P. Factors facilitating early emotion understanding development: contributions to individual differences. *Hum. Dev.* **64**, 108–118 (2020).

56. Ogren, M. & Sandhofer, C. M. Emotion words link faces to emotional scenarios in early childhood. *Emotion* **22**, 167–178 (2022).
57. Camras, L. A. & Allison, K. Children's understanding of emotional facial expressions and verbal labels. *J. Nonverbal Behav.* **9**, 84–94 (1985).
58. Lawrence, K., Campbell, R. & Skuse, D. Age, gender, and puberty influence the development of facial emotion recognition. *Front. Psychol.* **6**, 761 (2015).
59. Keulers, E. H. H., Evers, E. A. T., Stiers, P. & Jolles, J. Age, sex, and pubertal phase influence mentalizing about emotions and actions in adolescents. *Dev. Neuropsychol.* **35**, 555–569 (2010).
60. Dai, J. & Scherf, K. S. Puberty and functional brain development in humans: convergence in findings? *Dev. Cogn. Neurosci.* **39**, 100690 (2019).
61. Hoemann, K., Xu, F. & Barrett, L. F. Emotion words, emotion concepts, and emotional development in children: a constructionist hypothesis. *Dev. Psychol.* **55**, 1830–1849 (2019).
62. Richardson, H., Lisandrelli, G., Riobueno-Naylor, A. & Saxe, R. Development of the social brain from age three to twelve years. *Nat. Commun.* **9**, 1027 (2018).
63. Fan, F. et al. Development of the default-mode network during childhood and adolescence: a longitudinal resting-state fMRI study. *Neuroimage* **226**, 117581 (2021).
64. Gordon, E. M. et al. Three distinct sets of connector hubs integrate human brain function. *Cell Rep.* **24**, 1687–1695 (2018).
65. Brandman, T., Malach, R. & Simony, E. The surprising role of the default mode network in naturalistic perception. *Commun. Biol.* **4**, 79 (2021).
66. da Silva, P. H. R., Rondinoni, C. & Leoni, R. F. Non-classical behavior of the default mode network regions during an information processing task. *Brain Struct. Funct.* **225**, 2553–2562 (2020).
67. Gordon, E. M. et al. Default-mode network streams for coupling to language and control systems. *Proc. Natl Acad. Sci. USA* **117**, 17308–17319 (2020).
68. Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D. & Pizzagalli, D. A. Large-scale network dysfunction in major depressive disorder. *JAMA Psychiatry* **72**, 603 (2015).
69. Vargas, C., López-Jaramillo, C. & Vieta, E. A systematic literature review of resting state network—functional MRI in bipolar disorder. *J. Affect. Disord.* **150**, 727–735 (2013).
70. Broyd, S. J. et al. Default-mode brain dysfunction in mental disorders: a systematic review. *Neurosci. Biobehav. Rev.* **33**, 279–296 (2009).
71. Williams, L. M. Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: a theoretical review of the evidence and future directions for clinical translation. *Depress. Anxiety* **34**, 9–24 (2017).
72. Lettieri, G. et al. Emotionotopy in the human right temporo-parietal cortex. *Nat. Commun.* **10**, 5568 (2019).
73. Lettieri, G. et al. Default and control network connectivity dynamics track the stream of affect at multiple timescales. *Soc. Cogn. Affect. Neurosci.* **17**, 461–469 (2022).
74. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437–1448 (2015).
75. Kragel, P. A., Knodt, A. R., Hariri, A. R. & Labar, K. S. Decoding spontaneous emotional states in the human brain. *PLoS Biol.* **14**, e2000106 (2016).
76. Kragel, P. A. & Labar, K. S. Decoding the nature of emotion in the brain. *Trends Cogn. Sci.* **20**, 444–455 (2016).
77. Chang, L. J. et al. Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* **7**, eabf7129 (2021).

Methods

Study procedures were approved by the Chesapeake Institutional Review Board. For participants under the age of 18 years, written assent was obtained from the participant, and written consent was obtained from the participant's parent or legal guardian. Monetary compensation was provided for time and expenses incurred as a result of participating (for example, travel).

Study description

Analyses included data from 823 participants from the first nine releases of the Healthy Brain Network (HBN). The HBN study is a large, multi-site study of children and young adults ages 5–21 years all collected in the New York area. Recruitment, consent and study procedures are described in detail at https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/index.html and in the data publication[78]. Of the available datasets, 88% were from two of the four sites; thus, the data from the other two sites were excluded. Additionally, because 92% of the data were from children under the age of 16 years, and because previous work suggests that rates of change of emotion inference and reasoning skills sharply decrease at around age 15–18 years[79–84], we limited our analyses to the 5–15 year olds. Thus, datasets from 2,053 participants were considered for analysis. Of those data, two were removed for intracranial anomalies; 522 were removed for incomplete data; and 679 were removed for high motion based on strict data processing standards (see the 'MRI data processing' subsection). Final sample characteristics are included in Supplementary Table SC1, and the distribution plots for age and puberty scores are in Supplementary Fig. SC2. No power analysis was conducted to determine our final sample size, but our sample exceeds those used in previous similar analyses.

Magnetic resonance imaging data (MRI)

MRI data from two sites, the CitiGroup Cornell Brain Imaging Center (CBIC) and the Rutgers University Brain Imaging Center (RUBIC), were considered for analysis. CBIC data were collected on a Siemens 3T Prisma scanner, and RUBIC data were collected on a Siemens 3T Trio scanner. Both sites used a 64-channel head coil. The MRI data included in this analysis were T1-weighted and T2-weighted anatomy scans as well as BOLD functional MRI (fMRI). fMRI data were simultaneous multi-band echo planar imaging sequences with an 800-ms repetition time collected while children watched two video clips during fMRI scanning, a 10-min clip from the movie *Despicable Me* and a 3-min, 20-s short called *The Present*. For datasets with multiple T1-weighted or T2-weighted images, the image with the least artifact/motion was used for processing. Sequence parameters are listed in detail on the HBN project website.

MRI data processing

Data were pre-processed using the Human Connectome Project minimal processing pipeline[47]. Additional processing steps were carried out using custom-written Python (version 3.8) scripts using the numpy[85] version 1.21.6, scipy[86] version 1.7.3, nibabel version 3.2.1 and pandas version 1.1.2 libraries. These scripts are publicly available at https://github.com/catcamacho/hbn_analysis.

In brief, structural T1-weighted and T2-weighted data were aligned and corrected for both gradient and bias field[87] distortions before processing through FreeSurfer's recon-all pipeline[88] version 7. FreeSurfer-generated surfaces were visually inspected for accuracy, and datasets with surface errors were removed from analysis ($n = 336$). Most of the rejected surfaces had motion in either the T1 or T2 image and would require substantive manual editing to correct, potentially introducing barriers to reproducibility. We, therefore, opted to remove these data rather than invest time and resources in manual editing. Surfaces were then aligned to template space using multi-modal surface matching[89] and resampled to 32,000 vertices.

Functional data were corrected for both gradient and bias field[87] distortions, slice time corrected, rigidly realigned, normalized to a global mean and masked in volume space. Volumes were next converted into surface space by aligning the cortical ribbon produced from the anatomical data processing. In surface space, functional data were rescaled to standard units before applying nuisance regression and bandpass filtering (0.008–0.1 Hz). Nuisance regressors included global signal, six motion parameters (translation and rotation in each x, y and z directions), framewise displacement and temporal censoring of volumes with a framewise displacement of 0.9 mm or more, a motion cutoff that is found to be optimal for task-based analyses[90,91]. Although movie stimuli have been shown to enhance task compliance in children[92,93], we employed these strict motion correction procedures to avoid motion contamination of inter-subject synchrony signals. For each functional run, data were considered usable if 80% of the sequence had a framewise displacement of less than 0.9 mm. Finally, the Gordon cortical[94] and Seitzman subcortical and cerebellum[95] atlases were applied to the residuals, resulting in a denoised timeseries for 333 cortical regions representing 12 cognitive networks, 34 subcortical regions and 27 cerebellar regions, for a total of 394 parcels for analysis. Data were split by site into Discovery/Training (RUBIC, $n = 424$) and Replication/Testing (CBIC, $n = 389$) samples for analysis.

Video stimuli analysis

Emotional content characterization. Videos were coded and processed using the EmoCodes system version 1.0 (ref. 96). The EmoCodes system is a tool for characterizing low-level and affective content in movies that is both reproducible and externally validated. Two independent raters coded each video for faces, body parts and written words as well as expressions of negative and positive emotionality of each character (Sørensen–Dice similarity = 0.65–1.00), which were averaged across raters before applying to further analysis. We refer to the 'negative' and 'positive' emotion codes together as 'global' or 'general' emotions rather than valence because valence typically implies that a given cue is either positive or negative, whereas our coding allows a given cue to be both. For most of the video frames, however, the term valence could still be used accurately to describe the coding because it was rare for a given cue to be both negative and positive. These global negative and positive codes for each character were then combined with character affective intensity codes to create a summary measure of overall, frame-by-frame negative and positive emotional intensity per externally validated procedures[16]. Specific emotions that are cognitively basic and learned across childhood were also coded for each character: anger, fear, excitement, sadness and happiness. Specifically, for each frame and each character the facial, physical and verbal expression of each of these emotions were coded as present or not (1 or 0, respectively) and then summed for each frame. Just as with general emotions, this sum was then multiplied by the character's intensity rating and summed across characters to create a frame-by-frame global measure of the intensity of each specific emotion across each video. These ratings are shown for each video in Supplementary Fig. SA1.

Low-level video feature analysis. Low-level video features were extracted using pliers library version 0.4.1 (ref. 97) via EmoCodes Python library version 1.0.10 (ref. 96) and included framewise brightness, visual sharpness, optical flow, fraction of frame containing highly salient content, visual vibrance, loudness and a rolling measure of tempo. We tested if the coded emotion content timeseries were collinear with each of these features using the SummarizeFeatures class in the EmoCodes library. Key figures from this report are included in Supplementary Fig. SA2. Except for brightness in *The Present* (positive $r = 0.39$ and negative $r = -0.66$), none of these features was notably correlated with emotion metrics ($rs < |0.30|$). Notably, none of the variance inflation factors exceeded 3 across both videos, suggesting that, even with the higher correlation between brightness and positive/negative

content, the variables are not so correlated to destabilize a multiple regression model. Thus, activation models included the affective measure (positive and negative content), brightness and loudness (included for thoroughness).

A detailed description of each feature coded in the videos and their covariance is included in Supplementary Table SA1. An analysis of video features in relation to motion metrics is included in Appendix A2 and Fig. SA3.

Estimating activation to emotional content

The positive, negative, brightness and loudness signals were next convolved with a double gamma hemodynamic response function (5-s peak and 12-s undershoot) for within-subject activation analysis. The transformed regressors were next entered into a GLM, predicting the BOLD timeseries for each voxel. Parameter estimate maps from this analysis were used in group-level analysis. Sample average activation maps are shown in Fig. 1, and difference maps are shown in Extended Data Fig. 1. General and specific activated maps were extracted separately: general emotion regressors included negative, positive, brightness, loudness, written words and spoken words, whereas specific emotion regressors included anger, happiness, fear, excitement, sadness, brightness, loudness, written words and spoken words. As expected given the lack of collinearity with the emotion regressors, including the 'words' and 'speaking' regressors in the activation models did not change our main emotion analysis findings. The statsmodels version 0.13.2, scipy version 1.7.3, numpy version 1.21.6, nibabel version 3.2.1 and pandas version 1.1.2 libraries were used for these analyses.

Emotion activation classification

Support vector machine classification on the activation (beta) maps was used to characterize the regions for which activation differentiates emotions across the sample. For each analysis, data were split into a training and testing dataset by collection site. Support vector models were fitted to the training data using 10-fold cross-validation. Specifically, the training data were split into 10 partitions, and training was conducted on nine partitions before being tested on the tenth. This process resulted in 10 models, which were then combined for final performance assessment. Final performance was determined using the left-out, unseen testing data. Accuracy point estimates were obtained based on the correct number of label assignments. A pseudo-bootstrapping procedure was used to compute 95% CIs for each the training cross-validated accuracy and the final test accuracy to provide statistical confidence of these point estimates. This procedure involved taking random subsamples from the dataset (of random size n, which is between 50% and 75% of the full dataset size) and either training or testing the models as appropriate. CIs were estimated from the resulting distribution of scores. Finally, a P value was assigned to the full model performance by permuting the feature set 1,000 times to build a null distribution to compare against the accuracy point estimates. This procedure was done separately for the general emotion labels (positive and negative) and the specific emotion labels (angry, happy, sad, fearful and excited). Results were practically identical regardless of which site was used for training or testing. The scikit-learn version 0.24.2, numpy version 1.21.6, nibabel version 3.2.1, pandas version 1.1.2 and scipy version 1.7.3 libraries were used for these analyses, and seaborn version 0.11.1 and matplotlib version 3.4.2 were used for visualization.

Characterize regions that are critical for differentiating affective content. Specific parcels, regions or networks that are critical for the models to identify emotion labels were systematically permuted, and the change in accuracy (original minus new) was recorded. Specifically, activation for either each parcel (for example, L_visual_1 and L_putamen) or each network or subcortical region label (for example, visual and putamen) was permuted by shuffling the values associated with that label or labels among datasets, and the model was retrained 1,000 times to create a distribution of accuracy change scores, with positive numbers indicating that the permuted model did a poorer job of identifying the emotion labels. As a control for network/region-level approach, we also repeated this procedure permuting random labels of the same number of parcels as were included in the previously described analysis, because each cognitive network and subcortical region set are not of equal sizes (for example, visual network is composed of 39 parcels, whereas the pallidum has only two parcels, right and left). In other words, we sought to differentiate changes in accuracy due to missing specific information (the network or region) versus missing that amount of information (the n parcels permuted). Both distributions were plotted, and regions/networks with a mean accuracy change score that did not overlap with the null distribution was considered 'important' for model accuracy.

Characterize which networks are sufficient for differentiating affective content. Another important question is whether a specific network or region is sufficient for predicting emotion labels or if the multivariate information is critical for classification. To test this, we also completed the model fitting procedures described earlier on datasets limited to each network or region.

Identify inter-subject similarities in activation to the videos

ISC analysis. We computed inter-subject activation similarity across each video by computing the ISC following standard procedures[98,99]. Raw P values were assigned from a null distribution of 10,000 ISC values derived from shuffled raw data. Bonferroni FDR correction was applied, and parcels significant across both the Discovery and Replication samples were considered regions of similarity in activation across the sample. Significant parcels from this analysis are reported in Extended Data Fig. 3. This analysis was performed to compare against the maturity inter-subject representational analysis results.

Dynamic similarity analysis. Dynamic inter-subject similarity was computed using ISPS analysis[100,101] to identify which scenes induced similar changes in activation across the sample (that is, increased inter-subject synchrony). ISPS was computed using previously established procedures[100]. For each parcel and for each participant, activation timeseries were z-scored, and phase angle was computed using a Hilbert transformation. This approach emphasizes the change in signal over the absolute magnitude of signal at each timepoint, allowing for comparisons in dynamic processing without the confound of differences in arbitrary signal units. Next, pairwise phase synchrony was computed as 1 minus the sine of the difference in phase angle between the two participants, resulting in a measure between 0 and 1 for each pair at each timepoint, with 1 indicating perfect synchrony. To identify specific timepoints when children were more in sync in activation, pairwise ISPS was averaged across networks and then across participants. Raw P values were assigned to each synchrony value by comparing the value to a distribution of null synchrony values created by permuting the original pre-processed data. To be considered a significant peak, the peak had to be at least 5 s wide with a prominence of Bonferroni FDR-corrected $P < 0.05$ detected using the scipy find_peaks function and had to appear in both the Discovery and Replication samples. To ensure that the peaks being detected were indeed the highest synchrony, if the synchrony threshold at FDR-corrected $P < 0.05$ was less than the value at the 95th percentile, the latter was used to define the peak prominence threshold instead.

Characterize linear and curvilinear activation differences across development

Support vector regression analysis was used to test if activation to specific emotion categories differs across development (either chronological age or pubertal stage). Pubertal stage scores were derived from the Peterson Puberty Scale[102] using standard scoring procedures,

resulting in scores ranging from 5 to 20. As expected, age and puberty scores were highly correlated ($r = 0.76$, $P < 0.001$). Just as with the support vector classification analyses, data were split into a training and a testing dataset by collection site; models were fitted to the training data using 10-fold cross-validation; and final performance was determined using the left-out, unseen testing data. We computed parametric and non-parametric correlations as well as the MSE between actual and predicted data labels as model performance metrics. CIs were assigned using the same procedures as used in the classification models. MSE *P* values were assigned using the permutation-based approach described in the previous section. *P* values for the correlation statistics were computed per standard procedures. Analyses were completed using a linear support vector kernel and a curvilinear kernel (radial basis function kernel) to test for each association between activation and maturity.

Test if maturity can be predicted from activation. Activation maps from each emotion (positive, negative, angry, happy, sad, fearful and excited) were used as separate feature sets to predict each maturity measure (age and puberty) for a total of 14 emotion models. Like the classification analyses, first whole-brain activation was used to predict each chronological age and puberty scores, followed by post hoc permutation-based importance testing. Model-fitting procedures were then repeated, limiting the dataset to a specific region/network to test if information from each is sufficient for predicting maturity.

Specificity analyses and control analyses. Activation maps from each low-level (brightness and loudness) and non-emotion higher-level (spoken words and written words) feature included in the first-level activation models were used as separate features sets to predict each maturity metric using the same procedures as in the emotion analyses. Additionally, we tested if we could predict motion (mean framewise displacement) from each emotion (negative, positive, angry, fearful, excited, sad and happy) activation map.

Inter-subject similarity across development

We also tested if maturity was best captured using a nonlinear, model-free approach through similarity analysis. This approach complements the activation classification analysis; in the classification analysis, we used video features to extract activation, whereas, in this similarity analysis, we used activation patterns to identify what video content induces maturity-related differences in activation. To do this, we examined the association between the inter-subject similarity in activation across each movie (ISC) and the below maturity models using IS-RSA[103]. Maturational similarity was computed in three ways, each testing a different hypothesis of how cognitive affective processing changes with age or puberty:

(1) Nearest neighbor: children of similar maturity will process complex socio-emotional stimuli similarly (metric: sample maximum minus pair absolute difference)
(2) Divergence: younger children process complex socio-emotional stimuli more similarly, with divergence across development representing effects of individual differences in experience (metric: sample maximum minus pair average)
(3) Convergence: older children process complex socio-emotional stimuli more similarly, representing convergence on a shared social concept (metric: pair minimum)

IS-RSA of ISC data. We computed ISC across each video following standard procedures[98,99]. First, we computed the pairwise ISC in activation timeseries on a parcel-wise basis. Next, the maturity similarity ratings were each correlated with the parcel-wise ISCs. *P* values were assigned using a Mantel test[104] with FDR correction. Specifically, the maturity ratings were systematically shuffled before analysis 10,000 to create a null distribution. *P* values were estimated from this permuted distribution, and the significance threshold was determined using a Bonferroni FDR for two independent samples. The exact same procedures were conducted for each video, with only parcels that were significant for both the Discovery and Replication samples reported in the results.

We then compared model fits for each parcel using a bootstrapping approach. Specifically, we ranked the model point estimates for each parcel. If the parcel was significant for more than one model, we used a bootstrapping method to generate a distribution of similarity coefficients for each of the two models and a *t*-test to compare estimates. If one model's point estimates were significantly higher than the other, that model was assigned for that parcel. We also compared the winning models for each parcel between videos and additionally report those that were consistent.

IS-RSA of ISPS data. ISPS was computed using previously established procedures[100,101]. In brief, activation timeseries were *z*-scored, and phase angle was computed using a Hilbert transformation for each parcel and each participant. This approach emphasizes the change in signal over the absolute magnitude of signal at each timepoint, allowing for comparisons in dynamic processing without the confound of differences in arbitrary signal units. Next, pairwise phase synchrony was computed as 1 minus the sine of the difference in phase angle between the two participants, resulting in a measure between 0 and 1 for each pair at each timepoint, with 1 indicating perfect synchrony.

Dynamic IS-RSA was guided by the ISC IS-RSA results. We limited the parcels to those that were significant for the prevailing developmental model. Because there were numerous cortical parcels spanning cognitive networks and no significant subcortical or cerebellar regions, we then grouped the significant parcels based on cognitive network and averaged for further analysis. Furthermore, because age captured variance in the data better than puberty based on the total number of parcels and magnitude of IS-RSA similarity, we did not examine puberty further. Next, we grouped the samples based on the prevailing developmental model to capture developmental differences in activation similarity across the sample. Because the Convergence model was the best fit across the majority of significant parcels, we, therefore, examined dynamic changes in cognitive network synchrony in the oldest children (top 20% by age) to compare to the full sample results for each sample and each video (*Despicable Me* Discovery = 12.63–15.90 years; *Despicable Me* Replication = 13.85–15.97 years; *The Present* Discovery = 12.95–15.99 years; *The Present* Replication = 13.46–15.97 years). ISPS data were averaged across the oldest children, and the same peak detection procedures used in the full sample were applied (minimum peak width of 5 s, prominence of FDR-corrected $P < 0.05$). Nearly the full timeseries was above the peak threshold in the oldest children, however, so we adjusted the minimum to 1.5 s.d. above the mean for each network to identify scenes of relatively high synchrony. As with the full sample, peaks must be present in both the Discovery and Replication samples to be considered significant.

Scene analysis

To determine if specific video features elicited greater similarity in children of similar ages, we conducted *t*-tests comparing coded video features from video segments identified in the previous section compared to other sections. Features were time-shifted six repetition times (4.8 s) to account for the delay in the hemodynamic response before conducting *t*-tests comparing the ratings within the identified peaks to ratings outside of the peaks. Only features that were significantly different (FDR-corrected[105] $P < 0.05$) are reported and plotted.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

MRI data: The first nine releases of the Healthy Brain Network Biobank, an open dataset, were used in these analyses. These data are available at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/sharing_neuro.html.
Video codes: The codes for the videos obtained using the EmoCodes system are available at https://emocodes.org/datasets/.

Code availability

Pre-processing was carried out using the Human Connectome Project minimal processing pipeline (available at https://github.com/Washington-University/HCPpipelines). Additional processing, analyses and plotting were carried out using custom scripts written in Python 3.7.2 (available at https://github.com/catcamacho/hbn_analysis) using the numpy version 1.21.6, scipy version 1.7.3, nibabel version 3.2.1 and pandas version 1.1.2 libraries. Analyses were carried out using the pliers version 0.4.1, statsmodels version 0.13.2 and scikit-learn version 0.24.2 libraries. Plotting was carried out using the seaborn version 0.11.1 and matplotlib version 3.4.2 libraries.

References

78. Alexander, L. M. et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 170181 (2017).
79. Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A. & Somerville, L. H. The nonlinear development of emotion differentiation: granular emotional experience is low in adolescence. *Psychol. Sci.* **29**, 1346–1357 (2018).
80. Nook, E. C. et al. Charting the development of emotion comprehension and abstraction from childhood to adulthood using observer-rated and linguistic measures. *Emotion* **20**, 773–792 (2020).
81. Motta-Mena, N. V. & Scherf, K. S. Pubertal development shapes perception of complex facial expressions. *Dev. Sci.* **20**, e12451 (2017).
82. Thomas, L. A., De Bellis, M. D., Graham, R. & LaBar, K. S. Development of emotional facial recognition in late childhood and adolescence. *Dev. Sci.* **10**, 547–558 (2007).
83. Durand, K., Gallay, M., Seigneuric, A., Robichon, F. & Baudouin, J.-Y. The development of facial emotion recognition: the role of configural information. *J. Exp. Child Psychol.* **97**, 14–27 (2007).
84. Wu, M. et al. Age-related changes in amygdala-frontal connectivity during emotional face processing from childhood into young adulthood. *Hum. Brain Mapp.* **37**, 1684–1695 (2016).
85. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
86. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
87. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
88. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
89. Robinson, E. C. et al. Multimodal surface matching with higher-order smoothness constraints. *Neuroimage* **167**, 453–465 (2018).
90. Siegel, J. S. et al. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* **35**, 1981–1996 (2014).
91. Siegel, J. S. et al. Data quality influences observed links between functional connectivity and behavior. *Cereb. Cortex* **27**, 4492–4502 (2017).
92. Greene, D. J. et al. Behavioral interventions for reducing head motion during MRI scans in children. *Neuroimage* **171**, 234–245 (2018).
93. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C. & Castellanos, F. X. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage* **122**, 222–232 (2015).
94. Gordon, E. M. et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26**, 288–303 (2016).
95. Seitzman, B. A. et al. A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *Neuroimage* **206**, 116290 (2020).
96. Camacho, M. C. et al. EmoCodes: a standardized coding system for socio-emotional content in complex video stimuli. *Affect. Sci.* **3**, 168–181 (2022).
97. McNamara, Q., De La Vega, A. & Yarkoni, T. Developing a comprehensive framework for multimodal feature extraction. In *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (eds Matwin, S. et al.) 1567–1574 (Association for Computing Machinery, 2017).
98. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 669–687 (2019).
99. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).
100. Aviyente, S., Bernat, E. M., Evans, W. S. & Sponheim, S. R. A phase synchrony measure for quantifying dynamic functional integration in the brain. *Hum. Brain Mapp.* **32**, 80–93 (2011).
101. Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P. & Sams, M. Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connect.* **2**, 91–101 (2012).
102. Petersen, A. C., Crockett, L., Richards, M. & Boxer, A. A self-report measure of pubertal status: reliability, validity, and initial norms. *J. Youth Adolesc.* **17**, 117–133 (1988).
103. Finn, E. S. et al. Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* **215**, 116828 (2020).
104. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
105. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Acknowledgements

We thank A. Witherspoon for assistance with video coding, the families who participated in the Healthy Brain Network study and the Child Mind Institute for sharing the data publicly. This work was funded by the National Science Foundation (DGE-1745038 to M.C.C.) and the National Institutes of Health (HD102156 to M.C.C. and MH109589 to D.M.B.).

Author contributions

M.C.C.: conceptualization, methodology, software, validation, formal analysis, data curation, writing—original draft, writing—review and editing, visualization and project administration. A.N.N.: conceptualization, methodology, supervision and writing—review and editing. D.B.: conceptualization, investigation and writing—review and editing. E.F.: conceptualization, writing—original draft and writing—review and editing. D.C.S.: conceptualization, investigation and writing—review and editing. L.F.: conceptualization, investigation and writing—review and editing. J.P.C.: conceptualization, supervision and writing—review and editing. C.M.S.: conceptualization, supervision and writing—review and editing. D.M.B.: conceptualization, methodology, resources, supervision, writing—review and editing and funding acquisition.

Competing interests
The authors declare no competing interests.

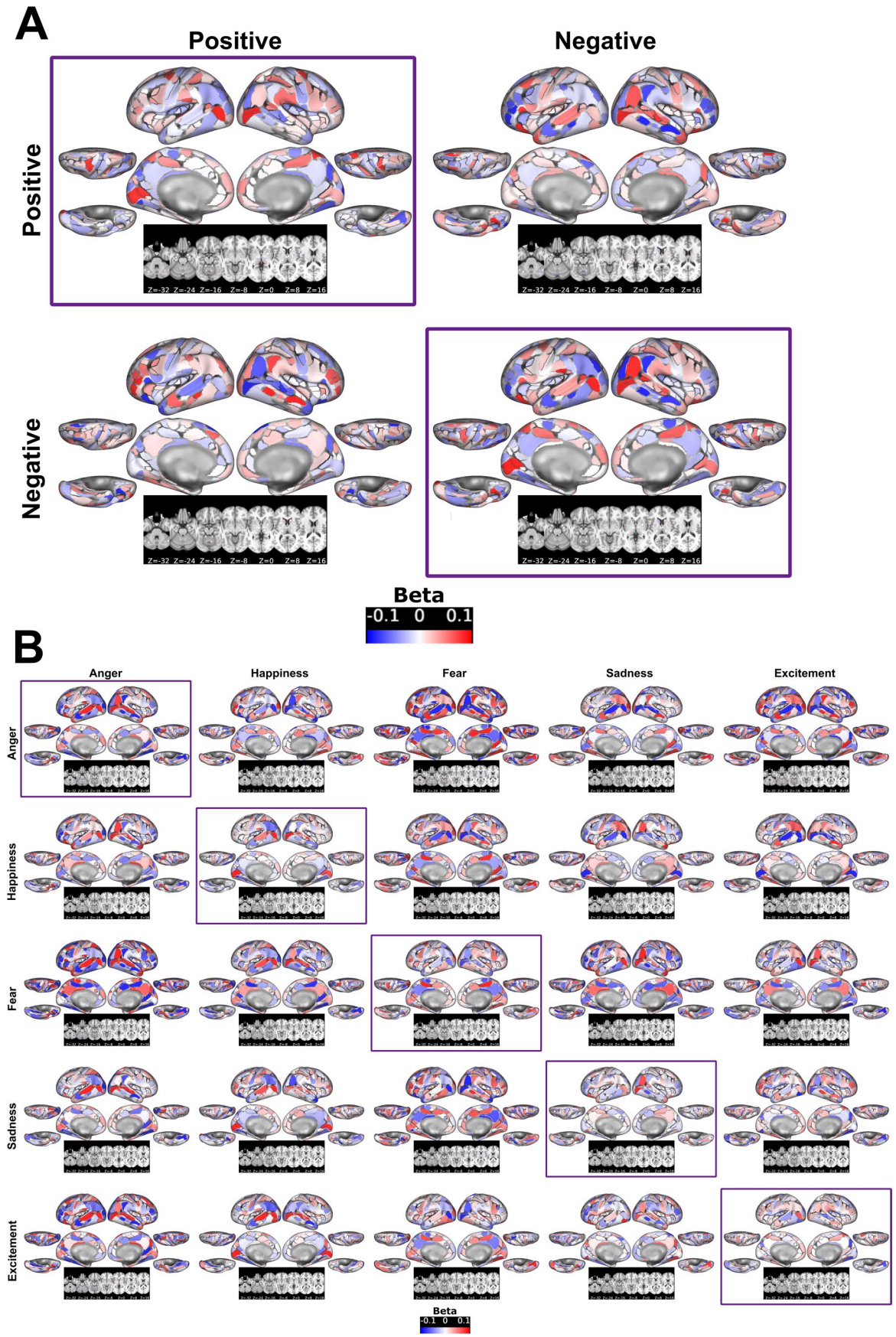
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41593-023-01358-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-023-01358-9.

Correspondence and requests for materials should be addressed to M. Catalina Camacho.

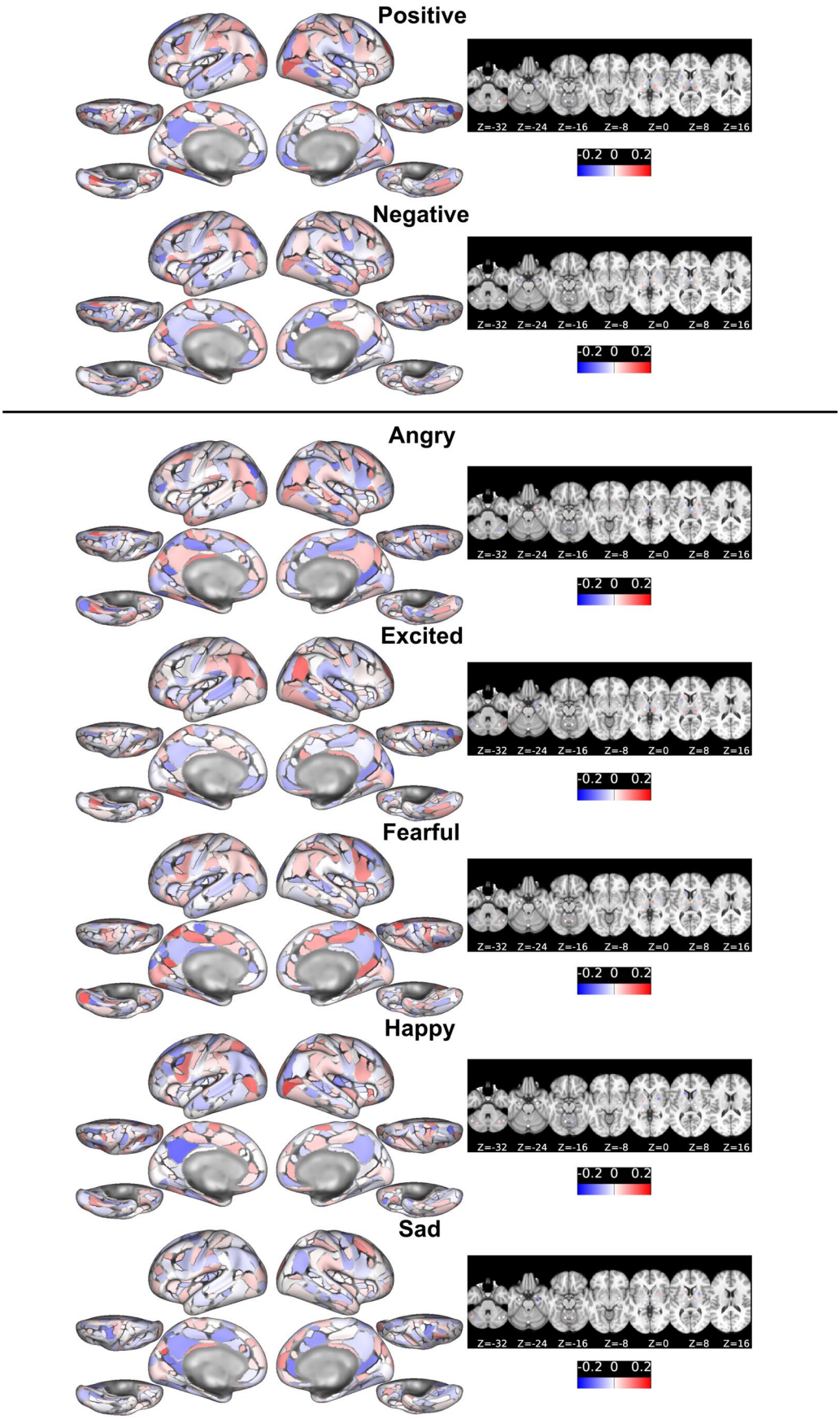
Peer review information *Nature Neuroscience* thanks Erik Nook, Heini Saarimäki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41593-023-01358-9



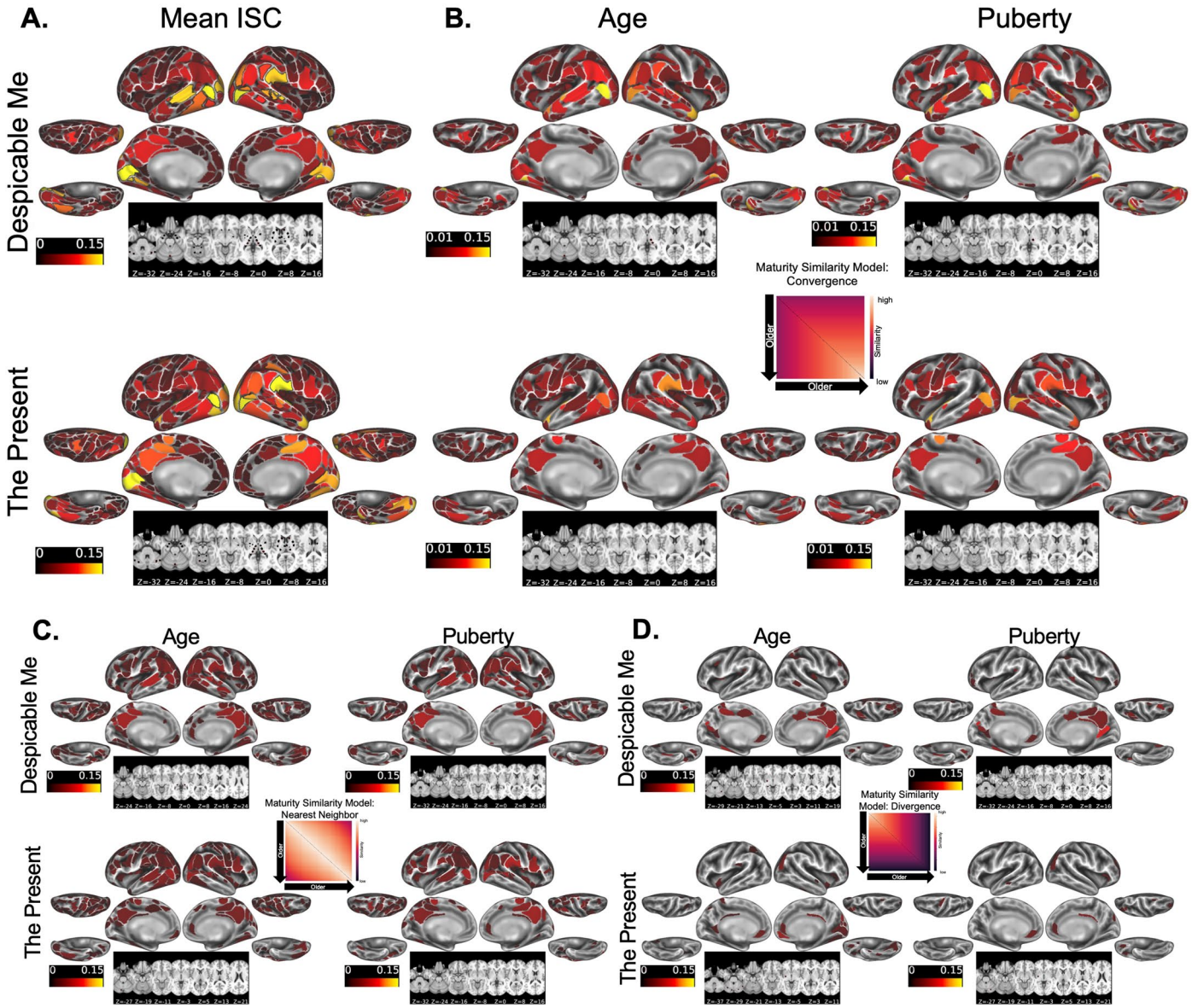
Extended Data Fig. 1 | Emotion-specific activation differences maps. (a) Difference maps for general emotions (column minus row). Mean activation maps are shown on the diagonal. **(b)** Difference maps for specific emotions (column minus row). Mean activation maps are shown on the diagonal.

https://doi.org/10.1038/s41593-023-01358-9



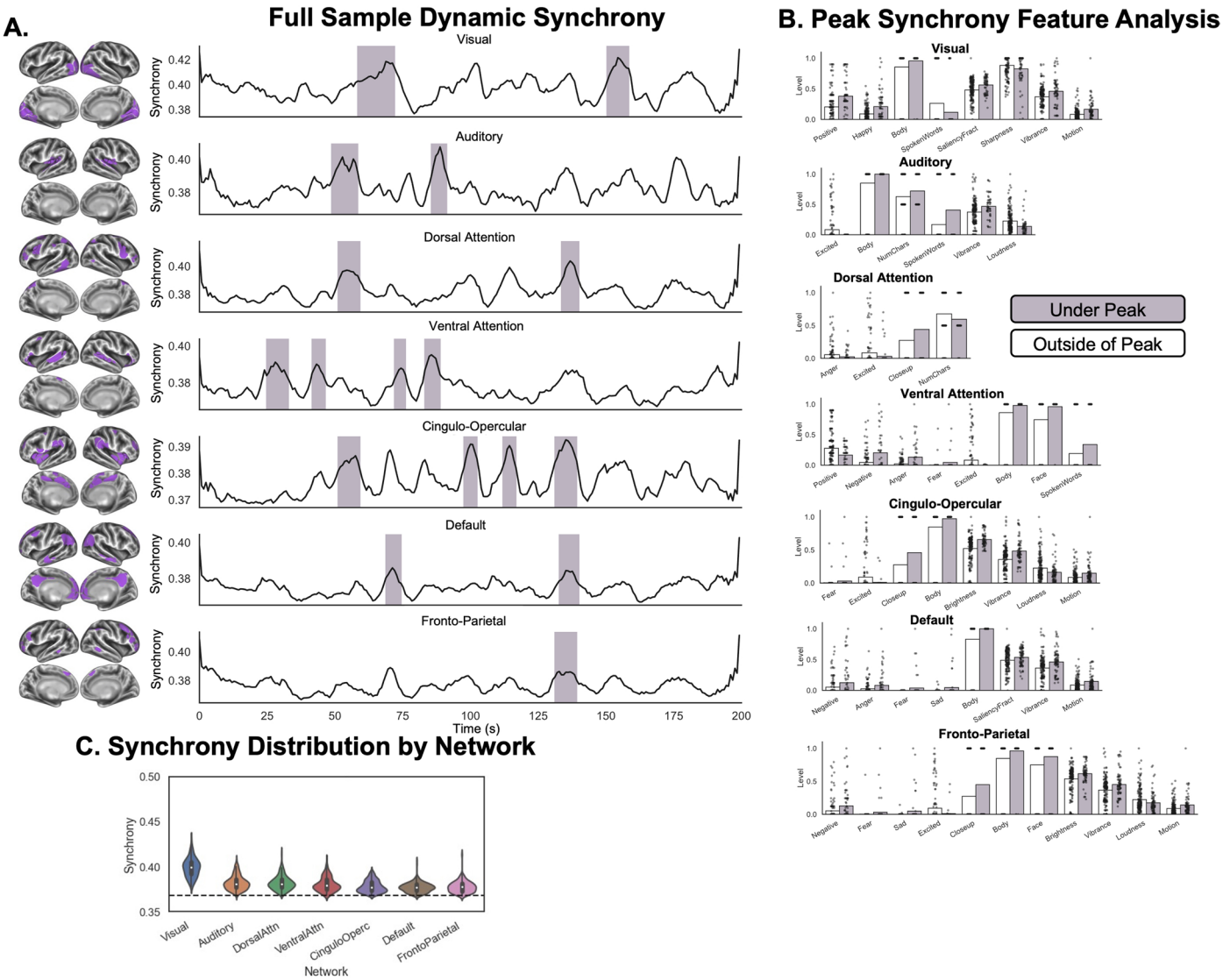
Extended Data Fig. 2 | Parcel-wise Pearson Correlations between chronological age and activation to each emotion. Age explained at most 3.8% of the variance in parcel-level activation (Age-Activation r2 range: Negative 0-0.017, Positive 0-0.030, Anger 0-0.020, Happy 0-0.026, Sad 0-0.023, Excited 0-0.028, Fearful 0-0.023; Puberty-Activation r2 range: Negative 0-0.029, Positive 0-0.030, Anger 0-0.024, Happy 0-0.038, Sad 0-0.014, Excited 0-0.030, Fearful 0-0.021).

https://doi.org/10.1038/s41593-023-01358-9



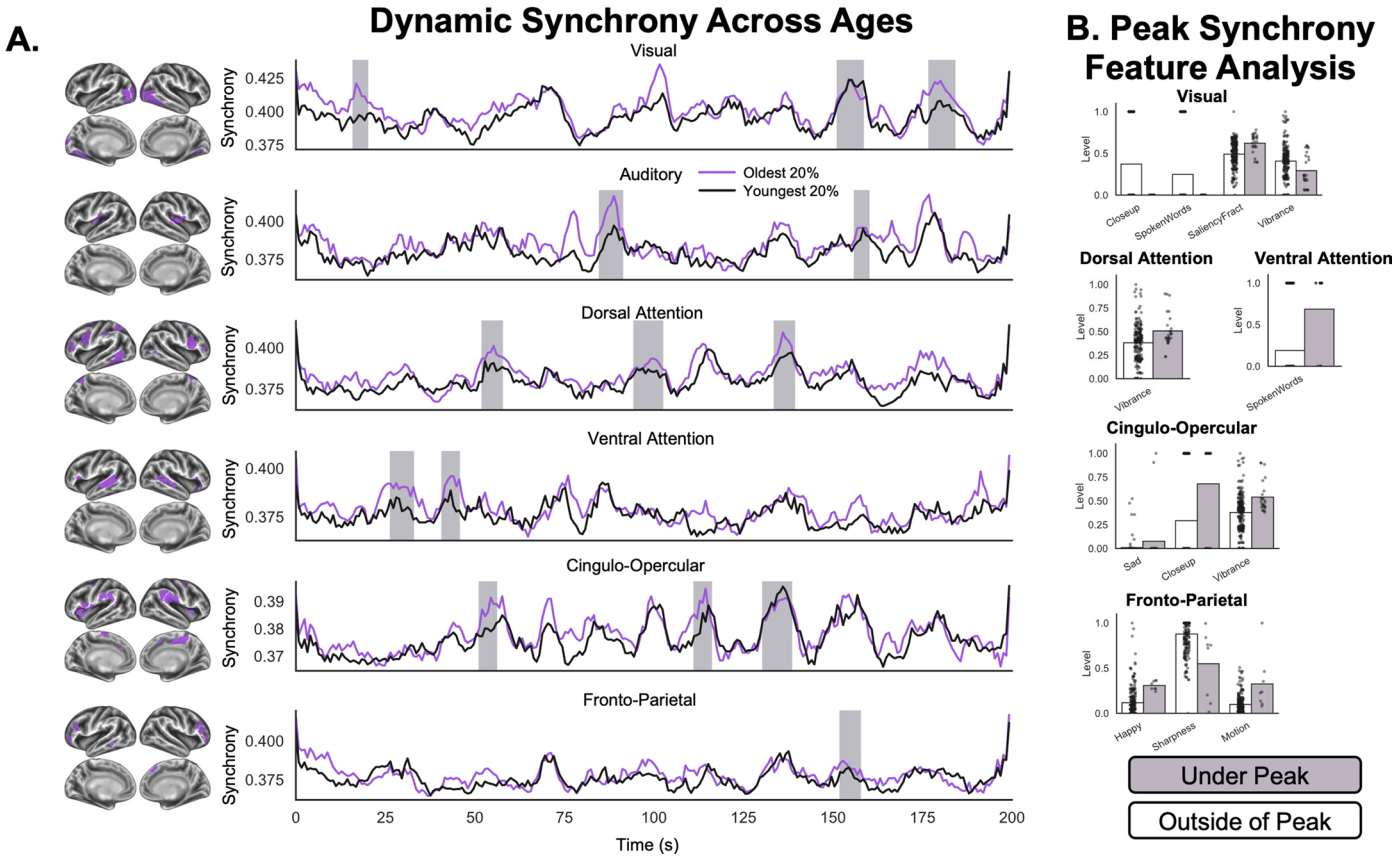
Extended Data Fig. 3 | Significant parcels and similarity coefficients for each model of maturation after FDR correction. (a) Average inter-subject correlations for each parcel for each movie. Significant parcels (one-sided Pearson's r; permutation based and FDR-corrected $p < 0.05$) are outlined in black. **(b)** Significant parcels and their coefficient magnitudes for the Convergence similarity model of maturation. **(c)** Significant parcels for the Nearest Neighbor and **(d)** Divergence models of maturation across chronological age and puberty.

https://doi.org/10.1038/s41593-023-01358-9



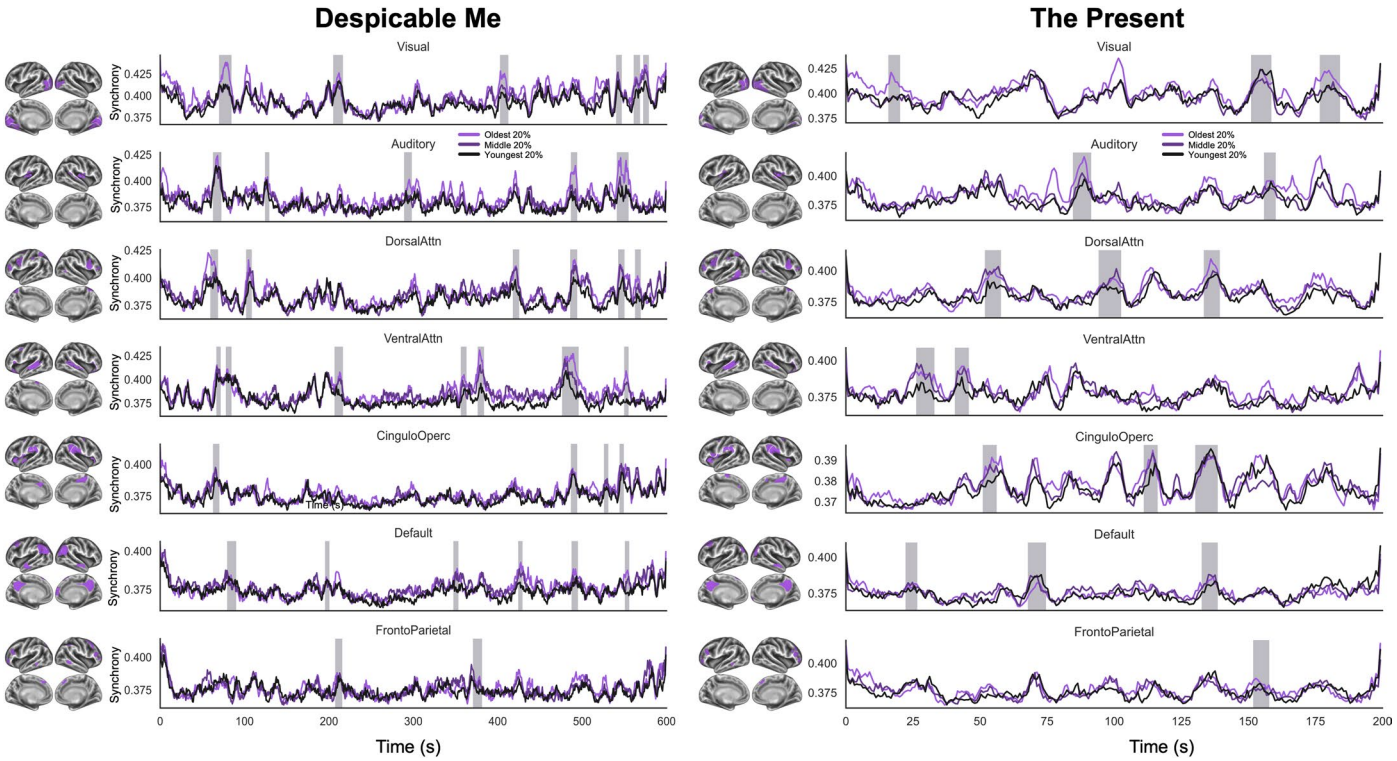
Extended Data Fig. 4 | Full sample dynamic analysis results for The Present. (**a**) Dynamic synchrony across the full sample with replicating peaks in synchrony shaded purple. Peaks at least 5 seconds wide and with a prominence higher than the 95th percentile value for that network (permutation-based $p < 0.001$). Parcels were limited to those that were significantly correlated across the sample at the group level after FDR correction. (**b**) Video feature means within the peaks were compared to features outside of the peaks using a two-sided t-test to test if specific video features elicited increased synchronization. Plotted features were those that were significantly different (FDR-corrected $p < 0.05$). Bars are plotted at the mean with dots indicating each measure used in computing the mean. (**C**) Violin plots of mean synchrony values by network. White dots indicate median value, the box the 50% interquartile range, and the whiskers each the upper and lower 25%. The dashed horizontal line indicates the value at permuted $p < 0.05$ after FDR correction.

https://doi.org/10.1038/s41593-023-01358-9



Extended Data Fig. 5 | Dynamic Synchrony analysis results in the oldest children for *The Present*. (**a**) Network dynamic activation similarity (synchrony) for each the oldest and youngest children in the sample. Included parcels are shown to the left of each trace. Shaded areas denote significant increases in synchrony (1.5 standard deviations above the mean) in the oldest children. (**b**) Bar plots of the mean value show the results from the video feature analysis comparing portions of the video within peaks of inter-subject synchrony to outside the peaks. Dots overlaid on the bar plots indicate each measure used in computing the mean Only features that significantly differed after FDR-correction are plotted.

https://doi.org/10.1038/s41593-023-01358-9



Extended Data Fig. 6 | Network-wise dynamic similarity plots for 3 age groups: oldest, middle, and youngest, for each movie. Included parcels are shown to the left of each trace. Shaded areas denote significant increases in synchrony (1.5 standard deviations above the mean) in the oldest children.

nature portfolio

Corresponding author(s): M. Catalina Camacho

Last updated by author(s): Jun 17, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Software files used to collect these data are included on the study website: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/MRI%20Protocol.html
Data analysis	Preprocessing was carried out using the Human Connectome Project minimal preprocessing pipeline (available at https://github.com/Washington-University/HCPpipelines). Additional processing and analyses were carried out using custom scripts written in python 3.7.2 (available at: https://github.com/catcamacho/hbn_analysis) using the numpy v1.21.6, scipy v1.7.3, nibabel v3.2.1, and pandas v1.1.2 libraries. Analyses were carried out using the pliers v0.4.1, statsmodels v0.13.2, and scikit-learn v0.24.2 libraries. Plotting was carried out using the seaborn v0.11.1 and matplotlib v3.4.2 libraries.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

fMRI and behavioral data are available as the first 9 releases of the Healthy Brain Network Biobank dataset. Video codes obtained using the EmoCodes system are available at https://emocodes.org/datasets/.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We report the sex distribution for the sample in Table SC1. We did not select for a specific distribution of sex or gender in this study. We also did not test for sex-specific effects since there is no previous work indicating consistent sex differences in emotion processing. Rather, the previous literature points to differences on the basis of experience or age rather than sex or gender.
Population characteristics	Participants included children ages 5-to-15 years of age. Approximately 60% of the sample was male. No genotyping or treatment-related data were used in this study.
Recruitment	Because this is an open dataset, we refer the reader to the study website for more information: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Recruitement.html
Ethics oversight	Study ethical oversight was provided by the Child Mind Institute and the site institutional review boards.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The Healthy Brain Network Biobank is a large multi-site quantitative study directed by the Child Mind Institute in New York with the goal of creating a richly phenotyped and multimodal sample from 5-21 year old youth (target N=10,000) to advance pediatric psychiatry and neuroscience research. Briefly, children and their parents were recruit to take part in 3 in-person visits during which they underwent neuroimaging, cognitive and behavioral assessments, clinical interviews, and completed questionnaires.
Research sample	The sample used in this paper is N=823 children ages 5-15 years who are a subsample of the data from the first nine releases. The first 9 releases include MRI data from 2,232 individuals. We removed 179 individuals who were ages 16 or older then further removed individuals with 1) missing fMRI sequences, 2) intracranial anomalies, or 3) too much motion to analyze, leaving 823 for our final analysis. This is not a representative sample of the United States as a whole. We chose this sample because of the large sample size and use of emotional videos.
Sampling strategy	The Healthy Brain Network uses a community-referred model, oversampling for children ages 5-12 years. Advertisements appeal to parents who may be concerned about a specific behavior in their child, though meeting criteria for a psychiatric symptom was not inclusion criteria for this study (see http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Recruitement.html for more information). The current sample is therefore enriched for a full spectrum of psychopathology symptoms relative to the public, (though the psychopathology data are beyond the scope of this analysis. More details can be found here: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/About.html
Data collection	Details on how MRI data were collected can be found here: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/MRI%20Protocol.html Details on maturity (age and puberty) data collection can be found here: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Pheno%20Protocol.html Procedures relevant to this manuscript are as follows: Briefly participants underwent fMRI while watching emotionally evocative

	videos and reported on child age. At a separate session, parents reported on child pubertal stage. We do not have information about whether or not researchers were blinded to all information about the child, however since the MRI visit was conducted before the other behavioral visits, we can reasonable assume that these data were collected blinded to other information. There were no experimental conditions for researchers to be blinded to.
Timing	Data collection is ongoing. We did not determine our sample based on data collection efforts.
Data exclusions	We removed 179 individuals who were ages 16 or older due to previous work indicating that emotion reasoning is well-developed by mid to late adolescence and due the the poorer representation of the 16-21 ages in the available data (less than 10 subjects per year per site). We further removed individuals with 1) missing fMRI sequences (522), 2) intracranial anomalies (N=2), or 3) too much motion to analyze (N=679), leaving 823 for our final analysis. These criteria we established prior to data processing.
Non-participation	We do not have information on who declined to participate available in the released data. We only have information on participants who chose to participate in the study.
Randomization	N/A. There is no experiment, condition designation, or treatment involved in this study. This study utilized a quantitative, observational design.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Magnetic resonance imaging

Experimental design

Design type	Children watch 2 video clips during fMRI. Children were instructed to watch the videos.
Design specifications	N/A
Behavioral performance measures	N/A

Acquisition

Imaging type(s)	Structural MRI for preprocessing, functional MRI for analysis
Field strength	3T
Sequence & imaging parameters	Imaging parameters for each site are detailed in the study protocols: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/MRI%20Protocol.html
Area of acquisition	Brain (whole head)

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Data were processed using the Human Connectome Project minimal preprocessing pipeline. Subsequent denoising was carried out using custom scripts written in python
Normalization	Data were normalized to MNI surface and volume space using the HCP pipeline.
Normalization template	MNI305

Noise and artifact removal	motion parameters, global signal, framewise displacement, and temporal censoring of high motion volumes were included in nuisance regression for each fMRI timeseries
Volume censoring	Volumes with a framewise displacement greater than 0.9mm were censored.

Statistical modeling & inference

Model type and settings	We used three kinds of analyses: 1) support vector machine learning, 2) inter-subject representational similarity and 3) dynamic similarity analysis. For the support vector machine learning analysis, we used a linear kernel and re-ran the same models using a nonlinear kernel (RBF). All other model settings were left as the default for the scikit-learn library.
Effect(s) tested	We tested 1) if we could accurately label emotions from their activation maps across the sample using support vector classification, 2) if we could predict maturity (age or puberty) from activation maps using linear or curvilinear support vector regression, 3) which nonlinear developmental model best fit the activation data, and 4) which scenes are processed differently in older versus younger children. Together, these tests inform how the brain encodes schematic emotional information that is presented fully in context and how activation to emotional cues may shift across development.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	No clustering was conducted. Analyses were performed at the parcel level.
Correction	We used Bonferroni FDR correction on p-values obtained via permutation-based approaches and took a discover/replication approach to ensure generalization of results. P-values obtained from standard t-tests were corrected using Benjamini-Hochsberg FDR correction (this only applies to the video feature peak analysis as all other analyses used a permuted p-value and Bonferroni-style FDR correction).

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis	Data analyses were performed on parcel-level maps and no further data reduction was completed prior to data analysis. All analyses were carried out using a discovery and replication approach to ensure generalization of findings. Support vector models: independent variables/features were the whole brain parcel-level activation maps and the dependent variable was either the activation map label (classification models) or the participant age/puberty stage (regression models). Significant features were identified by systematically permuting parcel and network-level activation and noting the change in model accuracy. Follow-up analyses were conducted on regions/network subsets of the features to test if that region or network was sufficient to train an accurate model. Inter-subject representational similarity analysis (IS-RSA): Here, the similarity between participants in parcel level activation time series (neural similarity) was compared to the similarity in maturity (maturational similarity). This approach is model-free, thus there are no dependent or independent variables.