

Empirical Paper

**Cite this article:** Rhoads SA, Cardinale EM, O'Connell K, Palmer AL, VanMeter JW, and Marsh AA. (2020) Mapping neural activity patterns to contextualized fearful facial expressions onto callous-unemotional (CU) traits: intersubject representational similarity analysis reveals less variation among high-CU adolescents. *Personality Neuroscience*. Vol 3: e12, 1–18. doi: [10.1017/pen.2020.13](https://doi.org/10.1017/pen.2020.13)

Received: 28 January 2020

Revised: 20 July 2020

Accepted: 10 August 2020





**Keywords:**

callous-unemotional traits; fearful faces; representational similarity analysis

**Author for correspondence:**

Shawn A. Rhoads,  
Email: [sr1209@georgetown.edu](mailto:sr1209@georgetown.edu)

# Mapping neural activity patterns to contextualized fearful facial expressions onto callous-unemotional (CU) traits: intersubject representational similarity analysis reveals less variation among high-CU adolescents

Shawn A. Rhoads<sup>1</sup> , Elise M. Cardinale<sup>1</sup>, Katherine O'Connell<sup>2</sup> , Amy L. Palmer<sup>3</sup>, John W. VanMeter<sup>4</sup>  and Abigail A. Marsh<sup>1</sup> 

<sup>1</sup>Department of Psychology, Georgetown University, Washington DC 20057, USA; <sup>2</sup>Interdisciplinary Program in Neuroscience, Georgetown University, Washington DC 20057, USA; <sup>3</sup>Independent Scholar and <sup>4</sup>Department of Neurology, Georgetown University Medical Center, Washington DC 20057, USA

## Abstract

Callous-unemotional (CU) traits are early-emerging personality features characterized by deficits in empathy, concern for others, and remorse following social transgressions. One of the interpersonal deficits most consistently associated with CU traits is impaired behavioral and neurophysiological responsiveness to fearful facial expressions. However, the facial expression paradigms traditionally employed in neuroimaging are often ambiguous with respect to the nature of threat (i.e., is the perceiver the threat, or is something else in the environment?). In the present study, 30 adolescents with varying CU traits viewed fearful facial expressions cued to three different contexts (“afraid for you,” “afraid of you,” “afraid for self”) while undergoing functional magnetic resonance imaging (fMRI). Univariate analyses found that mean right amygdala activity during the “afraid for self” context was negatively associated with CU traits. With the goal of disentangling idiosyncratic stimulus-driven neural responses, we employed intersubject representational similarity analysis to link intersubject similarities in multivoxel neural response patterns to contextualized fearful expressions with differential intersubject models of CU traits. Among low-CU adolescents, neural response patterns while viewing fearful faces were most consistently similar early in the visual processing stream and among regions implicated in affective responding, but were more idiosyncratic as emotional face information moved up the cortical processing hierarchy. By contrast, high-CU adolescents’ neural response patterns consistently aligned along the entire cortical hierarchy (but diverged among low-CU youths). Observed patterns varied across contexts, suggesting that interpretations of fearful expressions depend to an extent on neural response patterns and are further shaped by levels of CU traits.

Callous-unemotional (CU) traits are early-emerging personality features characterized by deficits in empathy, concern for others, and remorse following social transgressions (Frick & White, 2008). These traits, which encompass the core affective and interpersonal traits of psychopathy, are associated with severe and persistent disruptive behavioral problems in childhood and adolescence, including aggression and delinquency, and pose a high risk for persistent aggressive and criminal behavior in adulthood (Barry et al., 2000; Burke, Loeber, & Lahey, 2007; Pardini, 2006; Pardini & Frick, 2013; Salekin, Brannen, Zalot, Leistico, & Neumann, 2006; Vasey, Kotov, Frick, & Loney, 2005). Among the interpersonal deficits most consistently associated with CU traits are reduced behavioral and neurophysiological responsiveness to signs of others’ distress (e.g., fearful facial expressions) (Fanti et al., 2017; Jusyte, Mayer, Künzel, Hautzinger, & Schönenberg, 2015; Lozier et al., 2014; Marsh et al., 2008; Sebastian et al., 2014) as well as difficulty accurately interpreting these expressions (Dawel, O’Kearney, McKone, & Palermo, 2012; Marsh & Blair, 2008; Wilson, Juodis, & Porter, 2011). However, because most neuroimaging studies of facial expression responding in CU youths feature passive viewing paradigms, there is little direct evidence linking how these expressions are interpreted by high-CU youths to their anomalous neural responses. The present study is the first to investigate neural responses to fearful facial expressions in different interpretative contexts in high-CU adolescents. We employed a novel data-driven approach to explore how variations in neural responding across interpretive contexts correspond to variations in CU traits.

High levels of CU traits are observed in approximately one-third of children with clinically significant conduct problems (Frick & Viding, 2009; Mills-Koonce et al., 2015) and are consistently linked to atypical patterns of neural responding to a variety of fear-associated stimuli

© The Author(s) 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Carré, Hyde, Neumann, Viding, & Hariri, 2013; Jones, Laurens, Herba, Barker, & Viding, 2009; Marsh *et al.*, 2008). Such stimuli include fearful facial expressions (Marsh, 2016), which serve important social functions that include signaling internal states of distress to others (Horstmann, 2003), signaling vulnerability and appeasement to inhibit aggression or elicit care (Hammer & Marsh, 2015; Marsh, Ambady, & Kleck, 2005), and promoting social learning about environmental threats (Hooker, Germine, Knight, & D'Esposito, 2006; Olsson & Phelps, 2007). High-CU youths and adults exhibit atypical responses to fear or other negative emotions expressed by the face, voice, or body, including reduced autonomic responding, reduced fear-potentiated startle, and impaired aversive conditioning (Blair, 1999; Blair, Jones, Clark, & Smith, 1997; Fanti *et al.*, 2017; Fanti, Panayiotou, Kyranides, & Avraamides, 2016; Kimonis, Fanti, Goulter, & Hall, 2017; Marsh *et al.*, 2011; Muñoz, 2009; Rothmund *et al.*, 2012). High-CU individuals also consistently show atypical patterns of neural activation while viewing fearful expressions, including reduced amygdala activation and striatum (Carré *et al.*, 2013; Jones *et al.*, 2009; Lozier *et al.*, 2014; Marsh *et al.*, 2008). However, less is known about where the stream of neural information processing of fearful faces in high-CU youths diverges from that of low-CU youths.

Information conveyed by faces is processed in distributed but overlapping systems of cortical and subcortical brain regions (Freiwald, Duchaine, & Yovel, 2016; Haxby *et al.*, 2001; Ishai, 2008; Pessoa & Adolphs, 2010). This information is systematically transformed from early representations in visual cortex (e.g., distinct facial features) to more complex, identity-specific representations in fusiform gyrus (e.g., combinations of features, then further reduced to aggregated, feature-invariant information) hierarchically along a posterior–anterior axis. Human neuroimaging finds evidence of activation to emotional faces along this hierarchy, including the visual cortex, fusiform gyrus (FFG), amygdala, superior temporal cortex, and more frontal interpretative regions that include the medial prefrontal cortex (mPFC), and lateral prefrontal cortex (LPFC; including the inferior, middle, and superior frontal gyri) (Sabatinelli *et al.*, 2011). The amygdala plays a key role in coordinating adaptive responses particularly to fearful expressions and other fear-linked stimuli through its reciprocal connections with other subcortical (e.g., hypothalamus, periaqueductal gray, midbrain) and cortical regions (e.g., insula, subgenual cingulate, and medial and lateral prefrontal cortex) (Blair, 2008; Davis & Whalen, 2001; Garvert, Friston, Dolan, & Garrido, 2014; Marsh, 2016; Robinson, Laird, Glahn, Lovullo, & Fox, 2010; Rosen & Donley, 2006). Furthermore, functional coupling among face-selective regions shows evidence for hierarchical clustering roughly into three subnetworks likely corresponding to processing individual identity (e.g., inferior occipital gyri, FFG), retrieval of semantic knowledge (e.g., LPFC, inferior parietal sulci, supramarginal gyri), and representation of emotional information (e.g., mPFC, orbital frontal cortex, insula, superior temporal sulci, temporal pole) (Zhen, Fang, & Liu, 2013). These regions receive input from and send modulatory feedback to lower-level sensory areas, which may enable the derivation of socioaffective meaning from emotional face stimuli.

But despite this relatively comprehensive understanding of the brain networks underlying emotional face processing, little is known about how patterns of responses to fearful expression vary as function of CU traits along the cortical hierarchy. This is in part because interpretations of fearful expression are unconstrained in

most neuroimaging studies. Because of the variable functions that fearful expressions serve, when these expressions are viewed in decontextualized settings, the signal of these expressions could be interpreted in multiple ways – for example, as a readout of an internal state (i.e., the expresser is afraid for themselves), as an effort to inhibit aggression (i.e., signaling that the expresser is afraid of the perceiver), or as a cooperative social cue (i.e., the expresser is signaling that they are afraid for the perceiver). This fact has several implications. For one, different subsets of participants (e.g., those with high- versus low-CU traits) may tend to interpret these expressions differently, which could contribute to inconsistencies in patterns of neural, physiological, and behavioral responding observed in response to them. In addition, CU traits may be more closely associated with atypical responses to fearful expressions in specific interpretative contexts. For example, CU traits may be more closely associated with atypical responses to fear when it is interpreted as an aggression-inhibition cue, given links between CU traits and failures to inhibit aggression despite the distress is caused (Cardinale & Marsh, 2015). This suggests the importance of understanding variation in neurophysiological responses to fearful expressions as a function of how these expressions are interpreted in high-CU youths.

The present study investigated whether adolescents with high-CU traits differentially respond to fearful facial expressions in each of three distinct contexts. Our primary aim was to explore whether variations in patterns of neural activation associated with viewing fearful facial expressions in these different contexts map onto variation in CU traits. To pursue this question, we employed a paradigm that presented participants with contextually-ambiguous fearful expressions as well as expressions for which the interpretative context is constrained (i.e., participants are instructed to interpret expressions as readouts of internal states) and measured blood-oxygenated level-dependent (BOLD) signal in a sample of adolescents with varying levels of CU during fMRI.

To analyze the resulting data, we conducted both univariate analyses and multivoxel pattern analyses that examined how idiosyncratic neural response patterns mapped onto differential intersubject models of CU traits. This allowed us to both assess the consistency of our results with past results obtained from univariate analyses and test novel questions such as whether high-CU participants show neural response patterns that are more similar to one another than low-CU participants. We employed a novel technique called Intersubject Representational Similarity Analysis (IS-RSA; Chen, Jolly, Cheong, & Chang, 2020; Finn *et al.*, 2020; Nguyen, Vanderwal, & Hasson, 2019; van Baar, Chang, & Sanfey, 2019) that assesses how interindividual variability in multivoxel brain activity patterns relates to individual differences in behaviors or traits using second-order statistics (Kriegeskorte, Mur, & Bandettini, 2008). This approach leverages between-participant differences in CU traits and enables us to treat participant-level phenotypic differences as signal instead of noise (Foulkes & Blakemore, 2018; van Baar *et al.*, 2019). Specifically, we tested where multivoxel neural activity patterns associated with viewing fearful facial expressions in each of three contexts would be: 1) more similar among high scoring CU adolescents, but more dissimilar among all others; 2) more similar among low scoring CU adolescents, but more dissimilar among all others; and 3) more similar among adolescents who are very similar in their CU scores regardless of their absolute position on the scale, but more dissimilar for participants with dissimilar scores.

**1. Method**

**1.1 Participants**

The MRI sample consisted of 37 adolescents who varied in conduct problem severity and CU traits. Of the sample, 20 were assessed as exhibiting clinically significant CU traits; the remaining 17 adolescents did not exhibit significant levels of either CU traits or disruptive behavior. The sample included males and females aged 10 to 17 years old who were recruited from Washington, D. C. and surrounding areas through referrals, advertisements, fliers seeking both healthy children and children with conduct problems. All youths and a parent or guardian completed an initial screening visit to determine eligibility for the scanning portion of the study. During this visit, demographic and clinical assessments were conducted, including a test of cognitive intelligence (K-BIT, Kaufman & Kaufman, 2004), and assessments of clinical symptomology using the Strengths and Difficulties Questionnaire (Goodman & Scott, 1999), Child Behavior Checklist (CBCL; Achenbach, 1991), and the Inventory of Callous-Unemotional Traits (ICU; Frick & Ray, 2015; Kimonis et al., 2008) completed separately by parent and child. Participants were excluded based on the following criteria: full-scale IQ scores <80, history of head trauma, neurological disorder, parent report of pervasive developmental disorder, or magnetic resonance imaging (MRI) contraindications. In addition, no siblings were permitted to participate. Additional exclusion criteria for healthy controls included any history of mood, anxiety, or disruptive behavior disorders. Written informed assent and consent were obtained from children and parents before testing. Approval for all procedures was obtained from the Georgetown University Institutional Review Board.

Of this sample, four participants were excluded from analyses due to excessive motion during the MRI scanning and three participants were excluded from analyses for failing the attention checks during the task (>50% errors). The resulting sample consisted of 30 adolescents ( $M_{age} = 13.33, SD = 2.29$ ; 14 females; see Tables 1a and 1b). All participants were native English speakers.

**1.2 Inventory on callous-unemotional traits**

In accordance with standard procedures, ICU scores were calculated using the highest item response from either the child or parent for each item, which reduces susceptibility to social desirability biases and optimizes accuracy across multiple contexts (Frick et al., 2003). Two subjects did not report an answer for a question item but their parents did, and one parent did not provide an answer for a question item but the subject did. For these instances, we used the only available item response. We also assigned “somewhat true” for two question items where one subject’s parent selected both “not at all true” and “somewhat true.” We then calculated a summary score of all item responses (ICU total) and each subscale (callous, uncaring, and unemotional) for each participant (Kimonis et al., 2008). Internal consistency was high for the ICU total scale ( $\alpha = .92$ ), callousness ( $\alpha = .82$ ), and uncaring ( $\alpha = .90$ ) subscales, but low-to-moderate for the unemotional subscale ( $\alpha = .49$ ), which is commonly observed among studies (Cardinale & Marsh, 2017).

**1.3 fMRI task**

We adapted an experimental task from Palmer and colleagues (2013), which comprised six runs. In the first two runs, participants viewed four 18-s blocks of rapidly presented fearful faces interleaved with four 18-s blocks of fixation for 3 min. During the

**Table 1a.** Demographic and behavioral characteristics

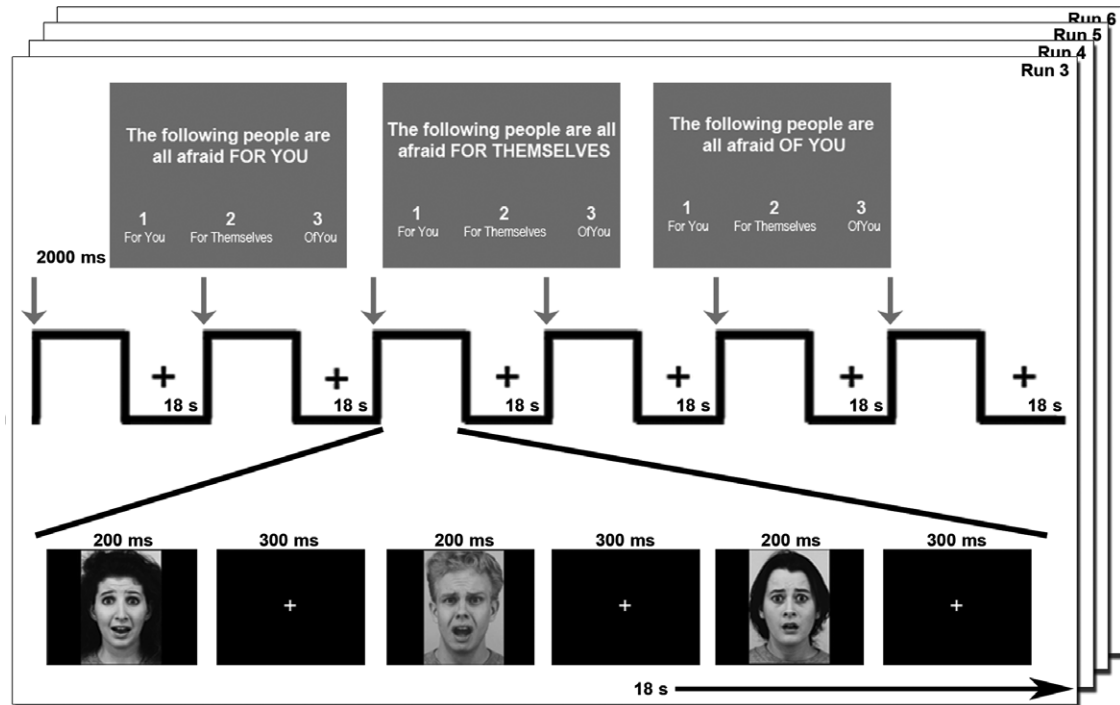
Characteristics	(N = 30)
Male:Female ratio	16:14
Age, mean (SD)	13.33 years (2.29)
IQ, mean (SD)	106.83 (13.92)
Race (n)	
White	15
Black or African American	9
Hispanic or Latino/a	3
Other	3
ICU Total, mean (SD)	35.82 (13.63)
Callous, mean (SD)	12.41 (6.44)
Unemotional, mean (SD)	8.73 (2.49)
Uncaring, mean (SD)	14.68 (5.96)
CBCL Externalizing, mean (SD)	18.93 (16.52)
Percent Accuracy (out of 24) during fMRI task, mean (SD)	85.83% (2.72)

**Table 1b.** Spearman  $\rho$  correlations among demographic and CU variables (N = 30)

	1	2	3	4	5	6	7
1. Age	–						
2. ICU Total	.10	–					
3. Callous	.03	.95**	–				
4. Uncaring	.16	.94**	.86**	–			
5. Unemotional	–.01	.65**	.54**	.54*	–		
6. CBCL <sup>^</sup> (Externalizing)	–.02	.83**	.81**	.80**	.62**	–	
7. IQ (KBIT)	–.12	–.38*	–.39*	–.27	–.36	–.30	–

\*\* $p < 0.001$ , \* $p < 0.05$ ; <sup>^</sup>imputed one question item as average of remaining responses for one subject due to blank data

fearful face blocks, faces were presented for 200 ms followed by a 300-ms fixation cross. The purpose of these baseline runs was to orient the participants to the fearful faces before presenting them in three experimental contexts. During the next four runs, participants again viewed 18-s blocks of rapidly presented fearful faces interleaved with 18-s blocks of fixation. During these runs, a 2000-ms cue appeared before each block that placed the faces in one of three contexts: “afraid of you,” “afraid for you,” or “afraid for themselves.” During this prompt, participants pressed one of three buttons corresponding to the condition to indicate that they had read and understood the context for the upcoming block. The interval between the offset of the cue and the onset of a face block pseudorandomly varied between 2500 and 3600 ms. Each contextualized block was presented twice per run for a total of six blocks of contextualized fearful faces interleaved with six blocks of fixation in each run. These final four runs of the task were 4 min 13 s each. The blocked design was selected to reduce cognitive load resulting from changing contexts (Figure 1).



**Figure 1. Visualization of the fMRI Task.** During the first two runs of fMRI, participants viewed 18-s blocks of noncontextualized fearful facial expressions (200 ms) and fixation (300 ms) interleaved with 18-s blocks of fixation (not depicted). During the final four runs, participants again viewed 18-s blocks of fearful facial expressions (200 ms) and fixation (300 ms) followed by 18-s blocks of fixation. Prior to each face block, a sentence appeared for 2000 s indicating that the “following people are all afraid ... ‘FOR YOU’, ‘FOR THEMSELVES’, or ‘OF YOU’.” Participants were asked to press one of three buttons that corresponded to the instruction as an attentional check.

Face stimuli included fearful facial expressions from 24 different exemplar individuals from the Karolinska face database (Lundqvist, Flykt, & Ohman, 1998), selected because the actors in this database were between the ages of 20 and 30, which is younger than most facial expression sets. Faces presented in grayscale were normalized for size and luminance. Each individual face was presented in only one of the three context condition blocks (i.e., six identities per context). The other six identities were presented within the first two (baseline) runs.

#### 1.4 fMRI data acquisition

MRI data were acquired on a 3.0 T Siemens TIM Trio MRI scanner (Erlangen, Germany), located at the Georgetown University Center for Functional and Molecular Imaging. The first two functional scans consisted of 72 contiguous T2\*-weighted echo planar imaging (EPI) whole-brain functional volumes while the final four functional scans consisted of 102 T2\*-weighted EPI volumes. All functional scan contained the following parameters: repetition time (TR) = 2500 ms; echo time (TE) = 35 ms; flip angle = 90°, 43 slices, matrix = 64 × 64; field of view (FOV) = 240 × 240 × 129 mm<sup>3</sup>; acquisition voxel size = 3.75 × 3.75 × 3.00 mm<sup>3</sup>. A T1-weighted high-resolution anatomical image was acquired for coregistration and normalization of functional images with the following parameters: TR = 1900 ms; TE = 2.52 ms; flip angle = 9°; 176 slices; FOV = 176 × 250 × 250 mm<sup>3</sup>; acquisition voxel size = 1.00 × 0.98 × 0.98 mm<sup>3</sup>.

#### 1.5 fMRI preprocessing and first-level analysis

Preprocessing was performed using fMRIPrep 1.3.2 (Esteban et al., 2017, 2018), which is based on Nipype 1.1.9 (Gorgolewski et al.,

2019; Gorgolewski et al., 2011) and included anatomical T1-weighted brain extraction, head-motion estimation and correction, slice-timing correction, intrasubject registration, and spatial normalization to the Montreal Neurological Institute 152 Nonlinear Asymmetric 2009c Template. The data were smoothed using a 6 mm<sup>3</sup> full-width half-maximum Gaussian kernel prior to first-level beta parameter estimation. Using Statistical Parametric Mapping (SPM12; <https://www.fil.ion.ucl.ac.uk/spm/>), a general linear model (GLM) was constructed for each participant across each of the six runs using boxcar regressors for each of the three instruction prompts, three fearful faces conditions with contexts, six motion parameters, one fearful faces condition without context, and a constant regressor for each run. The resulting GLM was convolved using a canonical hemodynamic response function and corrected for temporal autocorrelations using a first-order autoregressive model. Lastly, a standard high-pass filter (cutoff at 128 s) was used to exclude low-frequency drifts. Subject-level beta maps are available in Supplementary Data.

#### 1.6 fMRI univariate analysis

We first conducted a series of whole-brain univariate multiple regression analyses in SPM12 to examine how neural activation to fearful facial expressions varied as a function of CU traits. For each model, we controlled for age, gender, and IQ. A family-wise error (FWE) corrected threshold of  $p < 0.05$  with >10 voxels per cluster at the whole-brain level was applied to all analyses. To address *a priori* hypotheses regarding the role of the amygdala in fearful face processing, we also applied small volume correction using an anatomically-defined bilateral amygdala parcel from the Harvard Oxford Subcortical Atlas (Desikan et al., 2006; Goldstein et al., 2007; Makris et al., 2006) and conducted

region-of-interest (ROI) analyses that averaged across voxels in the left and right amygdala.

Our first regression model examined the relationship between CU traits and activation during the first two baseline runs. The second regression analysis modeled how neural activation during the final four runs to fearful facial expressions varied across contexts (“afraid of you,” “afraid of you,” and “afraid for self”) and as a function of CU traits. We conducted each test for ICU total scores and subscale scores.

### 1.7 Searchlight intersubject representational similarity analysis (IS-RSA)

To test where multivoxel neural activity patterns associated with viewing contextualized fearful facial expressions corresponded to each of three models (high-CU alike, low-CU alike, and nearest neighbors), we employed IS-RSA (Chen et al., 2020; Finn et al., 2020; Nguyen et al., 2019; van Baar et al., 2019) using the NITools package version 0.3.11 (Chang et al., 2019) in Python 3.5.2 and the CoSMoMMPA Toolbox version 1.1.0 in MATLAB 2017a. We first constructed three intersubject models (30 subjects  $\times$  30 subjects symmetrical matrices) based on ICU total scores, with each cell corresponding to the models’ prediction of a subject pairs’ neural pattern dissimilarity based on their CU scores. The first two models were generated using the Anna Karenina principle that “all low (or high) CU scorers are alike; each high (or low) CU scorer is different in his or her own way” (Finn et al., 2020) according to the following formulas: the low-CU alike model was constructed using the maximum score for each subject pair; the high-CU alike model was constructed using 1 minus the minimum score for each subject pair. The third nearest neighbors model was constructed using the absolute value of the difference between subject pairs’ scores. Prior to constructing each model, all subjects’ summary scores were normalized relative to the maximum CU score in the sample. To investigate finer-grained patterns of CU traits, we also constructed a nearest neighbors model by calculating the Euclidean distance between subject pairs’ item-wise responses on the ICU. Each matrix consisted of 435 unique combinations of subject pairs ( $30!/2! \times (30 - 2)!$ ).

Next, we searched for brain regions that responded to fearful facial expressions similarly 1) among high-CU subjects, 2) among low-CU subjects, and 3) for subjects who were similar in CU traits in a relative rather than an absolute sense. We conducted this analysis using a spherical searchlight comprised of 100 voxels. For each task condition (baseline, “afraid for you,” “afraid of you,” and “afraid for self”), we created a  $30 \times 30$  dissimilarity matrix in each searchlight in terms of the multivoxel activity pattern correlation distances (1 minus Spearman  $\rho$ ) between all pairs of participants while they viewed the expressions in each context (again, 435 unique combinations). We assessed correspondence between the lower triangles of these matrices using a Spearman  $\rho$  correlation and assigned  $\rho$  values to the center voxel of each searchlight. Statistical significance was determined using a Mantel permutation test (Mantel, 1967; Nummenmaa et al., 2012), in which both the rows and columns of one subject  $\times$  subject dissimilarity matrix (e.g., subject labels) were shuffled and the Spearman correlation between both correlation matrices was recomputed 1000 times to generate an empirical null distribution of rank correlations (Figure 2). For each searchlight, we calculated the  $p$ -value as the proportion of instances in which the permuted Spearman  $\rho$  statistic exceeded the true Spearman  $\rho$  statistic, and

thresholded the resulting maps using a cluster-forming threshold at  $p < .005$  and cluster-extent threshold at  $k = 10$ , and using a false-discovery rate (FDR) at  $q = .05$ .

For each condition, we assessed the voxel-wise similarity between the nearest neighbors model constructed using subjects’ item-wise responses and the nearest neighbors model based on the absolute difference of subject pairs’ summary scores by conducting a Spearman rank correlation, list-wise excluding nonoverlapping voxels. Because these tests revealed significant similarity across voxels between models for all conditions ( $\rho_{baseline} = .78$ ,  $p < .0001$ ;  $\rho_{ForYou} = .76$ ,  $p < .0001$ ;  $\rho_{OfYou} = .68$ ,  $p < .0001$ ;  $\rho_{ForSelf} = .70$ ,  $p < .0001$ ), we opted to interpret results from the absolute difference nearest neighbors model across all conditions. We report findings using the Euclidean distance model in Table 2.

## 2. Results

### 2.1 Behavioral data

Age was neither related to ICU total nor any of the subscales ( $r_s < .16$ ,  $p_s > .4$ ). Gender was related to CU traits with males ( $M = 40.31$ ,  $SE = 3.00$ ) scoring higher than females ( $M = 30.79$ ,  $SE = 3.75$ ),  $t(28) = 2.00$ ,  $p = .03$ . Males also had higher scores than females on the unemotional ( $t(28) = 1.70$ ,  $p = .049$ ) and uncaring ( $t(28) = 2.05$ ,  $p = .025$ ) subscales, and their scores trended higher on the callous ( $t(28) = 1.64$ ,  $p = .056$ ).

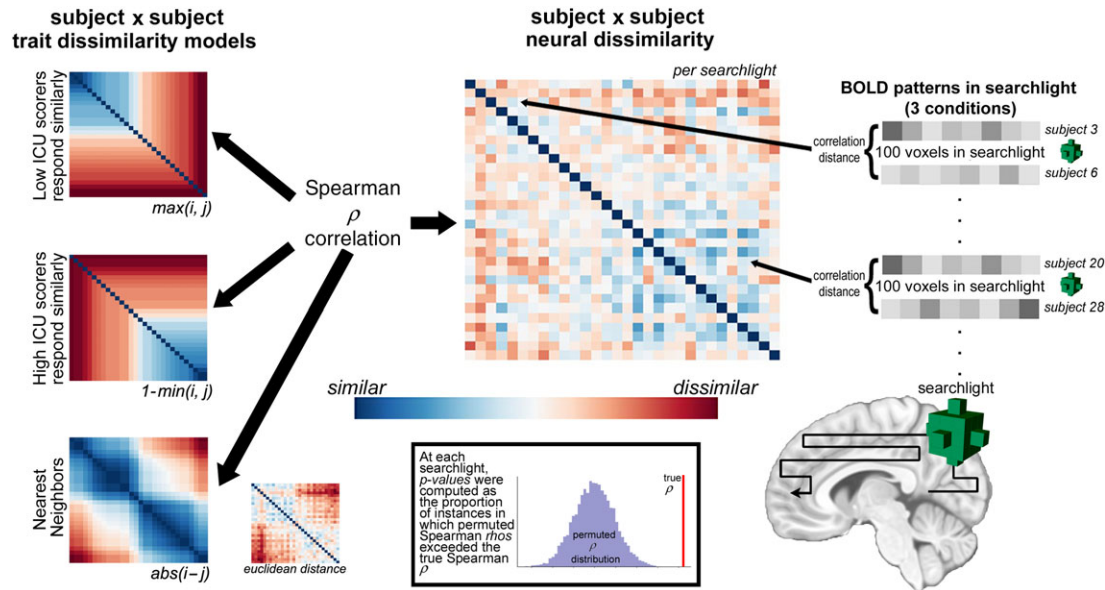
Behavioral data during the fMRI task consisted of adolescents’ responses to our attention-check questions in the final four runs of the fMRI task, where they pressed one of three buttons to confirm that they read and understood each prompt. The average accuracy score was 85.83% ( $SD = 2.72$ ). A Poisson regression analysis controlling for age, gender, and IQ indicated no statistically significant association between number of errors and CU traits (ICU total) ( $\beta = .009$ ,  $SE = .008$ ,  $p = .29$ ) or any covariates ( $.18 < p < .51$ ).

### 2.2 Neuroimaging

#### 2.2.1 Univariate

Our first analysis assessed how activation in response to fearful facial expressions varied as a function of CU traits using a traditional univariate approach across the whole-brain and within anatomically defined *a priori* bilateral amygdala masks. Multiple regression analyses revealed no statistically significant association between CU traits (total, callousness, uncaring, or unemotional) and neural activation to fearful faces at baseline when examining the whole brain. We found a significant main effect of activation to fearful faces versus fixation among all subjects in a small cluster in left visual cortex (13 voxels). Simultaneously, another multiple regression model predicting neural activation to fearful facial expressions across the different interpretative contexts found neither a significant effect of context nor an effect of CU traits. Applying small-volume correction in bilateral amygdala did not yield any findings across these two models.

Only when averaging across voxels in right amygdala did we find negative associations between mean activation and CU traits, controlling for age, gender, and IQ. During the “Afraid for Self” condition, we found associations between mean right amygdala activity and callousness ( $partial \rho = -.409$ ,  $p = .02$ ) and ICU total scores ( $partial \rho = -.349$ ,  $p = .06$ ). We present plots of mean activation as a function of CU traits for each context in bilateral, right, and left amygdala in the Supplementary Materials.



**Figure 2. Visualization of Searchlight Intersubject Representational Similarity Analysis.** Searchlight Intersubject Representational Similarity Analysis (IS-RSA) consisted of the following steps: (1) We computed three subject  $\times$  subject dissimilarity matrices based on CU summary scores across subjects. The first matrix tested a model in which low scoring CU adolescents' neural response patterns were more alike while all others' were different from each other; the second matrix tested a model in which high scoring CU adolescents' neural response patterns were more alike while all others' were different from each other; and the third matrix tested a model where adolescents' neural response patterns were similar to each other in a relative rather than an absolute sense. Depicted trait dissimilarity models are sorted by ICU total scores in ascending order. (2) For each condition ("Afraid for you", "Afraid for self", "Afraid of you"), we then computed a subject  $\times$  subject neural dissimilarity matrix within 100-voxel searchlights across gray matter. (3) Again for each condition, we vectorized the lower triangle of each matrix and performed a Spearman  $\rho$  correlation at each searchlight between intersubject behavioral dissimilarity and intersubject neural dissimilarity matrices, and assigned the  $\rho$  statistic to the center voxel in the searchlight. (4) Statistical significance was determined using a Mantel permutation test, in which both the rows and columns of the subject  $\times$  subject model dissimilarity matrix were shuffled and the Spearman correlation between both correlation matrices was recomputed 1000 times to generate an empirical null distribution of rank correlations. (5) At each searchlight, we calculated the  $p$ -value as the proportion of instances in which the permuted Spearman  $\rho$  statistic exceeded the true Spearman  $\rho$  statistic.

### 2.2.2 Intersubject RSA

We next investigated whether multivoxel neural activation patterns associated with viewing fearful facial expressions in the different contexts would correspond to each of our three models (Table 2). Our first analyses focused on the baseline runs, during which the context was unspecified. Here, results from the IS-RSA revealed that a cluster in the precuneus ( $\rho = .257, p < .0001$ ) exhibited an intersubject representational geometry where low-CU adolescents' neural response patterns were similar while others' were dissimilar during baseline fearful face viewing. A much larger set of clusters exhibited neural response patterns whereby high-CU adolescents' neural response patterns were similar while others' were dissimilar, including right posterior superior temporal gyrus extending to posterior superior temporal sulcus (pSTS) ( $\rho = .277, p < .0001$ ), right middle temporal gyrus (MTG) ( $\rho = .168, p = .001$ ), bilateral precentral gyrus ( $\rho = .266, p < .0001$ ), left superior frontal gyrus ( $\rho = .247, p = .001$ ), pars opercularis portion of the left inferior frontal gyrus (IFG) ( $\rho = .241, p = .002$ ), left anterior STG ( $\rho = .227, p < .0001$ ), and right postcentral gyrus extending to supramarginal gyrus ( $\rho = .177, p = .001$ ). Finally, adolescents who are similar in CU traits, regardless of whether they were low or high, exhibited similar neural responses in regions that included right posterior MTG (pMTG) extending to pSTS ( $\rho = .409, p < .0001$ ), right posterior inferior temporal gyrus ( $\rho = .200, p < .0001$ ), and right MTG ( $\rho = .194, p = .003$ ).

Our next analyses focused on response patterns during the cued contexts. When expressers were described as afraid for the viewer ("Afraid for you" condition), results from the IS-RSA revealed that

specific regions exhibited an intersubject representational geometry where low-CU adolescents' neural response patterns are similar while others' are dissimilar. Identified brain regions included right intracalcarine cortex ( $\rho = .436, p < .0001$ ), right frontal pole ( $\rho = .334, p < .0001$ ), right frontal operculum ( $\rho = .292, p = .001$ ), bilateral medial prefrontal cortex ( $\rho = .289, p = .004$ ), right pMTG ( $\rho = .269, p < .0001$ ), and bilateral paracingulate gyrus, extending to anterior cingulate cortex (ACC) ( $\rho = .263, p = .002$ ). Results in this condition also showed geometry whereby high-CU adolescents' neural response patterns were similar while others' were dissimilar in regions including several early-visual stream areas: bilateral superior occipital cortex ( $\rho_{left} = .230, p = .003; \rho_{right} = .342, p = .001$ ) and two clusters in right FFG ( $\rho = .551, p < .0001; \rho = .407, p = .005$ ); and right cerebellum lobule VI ( $\rho = .242, p < .0001$ ), right precuneus ( $\rho = .498, p < .0001$ ), right posterior supramarginal gyrus ( $\rho = .399, p < .0001$ ), right anterior supramarginal gyrus ( $\rho = .310, p < .0001$ ), right superior parietal lobule ( $\rho = .273, p < .0001$ ), pars opercularis portion of the right IFG ( $\rho = .270, p < .0001$ ), and right middle frontal gyrus ( $\rho = .261, p = .001$ ). Finally, among adolescents who were similar in CU traits, regardless of whether they were high or low on the scale, similar neural response patterns were observed in two clusters in the pars opercularis portion of right IFG ( $\rho = .283, p = .001; \rho = .188, p = .001$ ), right temporoparietal junction ( $\rho = .262, p < .0001$ ), two clusters in right FFG ( $\rho = .234, p = .001; \rho = .225, p < .0001$ ), right aSTG ( $\rho = .222, p = .002$ ), right ACC ( $\rho = .220, p = .001$ ), right intracalcarine cortex ( $\rho = .214, p < .0001$ ), and left posterior supramarginal gyrus ( $\rho = .206, p = .001$ ).

**Table 2.** Table of searchlight IS-RSA results meeting significance threshold.

Condition & Model	Brain Region(s)	Cluster Size	Spearman $\rho$	P-value	MNI Peak Coordinate		
					x	y	z
<b>Baseline</b>							
Low ICU Alike	Precuneus (R)	25	.257	<.0001 <sup>+</sup>	9	-53	39
	Inferior Frontal Gyrus, pars opercularis (R)	55	-.166	<.0001 <sup>+</sup>	50	14	9
High ICU Alike	Superior Temporal Gyrus, posterior division, extending to Superior Temporal Sulcus (R)	130	.277	<.0001 <sup>+</sup>	69	-23	3
	Middle Temporal Gyrus, temporooccipital part (R)	130	.168	.001	69	-46	0
	Precentral Gyrus extending to Middle Cingulate Cortex (R)	62	.266	<.0001 <sup>+</sup>	9	-20	51
	Precentral Gyrus to Middle Cingulate Cortex (L)	62	.163	.004	-10	-27	48
	Superior Frontal Gyrus (L)	15	.247	.001	-6	48	36
	Inferior Frontal Gyrus, pars opercularis (L)	31	.241	.002	-55	11	0
	Cerebral White Matter, extending to Superior Temporal Gyrus (L)	21	.227	<.0001 <sup>+</sup>	-36	-12	-12
	Postcentral Gyrus, extending to Supramarginal Gyrus (R)	26	.177	.001	58	-20	21
	Middle Temporal Gyrus, temporooccipital part (R)	11	.171	.004	62	-57	-6
	Precuneus (L)	16	-.188	.001	-6	-65	21
ICU Nearest Neighbors (Absolute Difference)	Middle Temporal Gyrus, posterior division, extending to Superior Temporal Sulcus (R)	147	.409	<.0001 <sup>+</sup>	69	-31	-12
	Planum Polare (L)	34	.229	<.0001 <sup>+</sup>	-40	-8	-15
	Inferior Temporal Gyrus, posterior division (R)	47	.200	<.0001 <sup>+</sup>	43	-23	-21
	Middle Temporal Gyrus, temporooccipital part (R)	11	.194	.003	54	-57	-3
	Cerebellum Lobule VI (R)	12	.188	.003	32	-57	-33
	Lateral Occipital Cortex, inferior division (R)	39	-.149	<.0001 <sup>+</sup>	43	-68	-9
	Parahippocampal Gyrus, posterior division (L)	16	-.135	<.0001 <sup>+</sup>	-14	-38	-12
	Lateral Occipital Cortex, superior division (R)	23	-.109	<.0001 <sup>+</sup>	43	-83	30
ICU Nearest Neighbors (Euclidean Distance)	Middle Temporal Gyrus, posterior division, extending to Superior Temporal Sulcus (R)	117	.347	<.0001 <sup>+</sup>	69	-31	-12
	Lingual Gyrus (L)	22	.229	.001	-10	-80	-6
	Superior Temporal Gyrus, anterior division (L)	30	.218	<.0001 <sup>+</sup>	-51	-16	-15
	Lateral Occipital Cortex, superior division (R)	17	.213	<.0001 <sup>+</sup>	20	-80	42
	Middle Occipital Gyrus (L)	12	.195	.002	-29	-76	18
	Middle Temporal Gyrus, temporooccipital part (R)	21	.179	.001	54	-57	-3
	Parahippocampal Gyrus, anterior division (R)	11	.170	.001	32	-23	-24

(Continued)

Table 2. (Continued)

Condition & Model	Brain Region(s)	Cluster Size	Spearman $\rho$	P-value	MNI Peak Coordinate		
<b>Afraid for You</b>							
Low ICU Alike	Intracalcarine Cortex (R)	43	.436	<.0001 <sup>+</sup>	5	-80	6
	Dorsolateral Prefrontal Cortex (R)	41	.334	<.0001 <sup>+</sup>	32	44	39
	Frontal Operculum Cortex (R)	11	.292	.001	43	22	3
	Medial Prefrontal Cortex (L, R)	16	.289	.004	-6	29	-18
	Middle Temporal Gyrus, posterior division (R)	23	.269	<.0001 <sup>+</sup>	65	-12	-21
	Paracingulate Gyrus, extending to Anterior Cingulate Cortex (L, R)	30	.263	.002	2	48	3
	Precuneus (R)	204	-.325	<.0001 <sup>+</sup>	17	-72	36
	Lateral Occipital Cortex, superior division (R)	204	-.220	.002	50	-61	48
	Superior Parietal Lobule (R)	204	-.213	.001	28	-57	54
	Precentral Gyrus, lateral part (R)	97	-.265	.001	43	-5	63
	Middle Frontal Gyrus (R)	97	-.184	.002	50	11	39
	Hippocampus (R)	23	-.238	.003	39	-31	-9
Lateral Occipital Cortex, superior division	19	-.174	.001	-25	-68	63	
High ICU Alike	Occipital Fusiform Gyrus (R)	39	.551	<.0001 <sup>+</sup>	24	-80	-9
	Cerebellum Lobule VI (R)	39	.242	<.0001 <sup>+</sup>	17	-65	-21
	Precuneus (R)	109	.498	<.0001 <sup>+</sup>	20	-72	33
	Lateral Occipital Cortex, superior division (L)	109	.230	.003	-6	-83	48
	Occipital Fusiform Gyrus (R)	14	.407	.005	35	-38	-18
	Supramarginal Gyrus, posterior division (R)	128	.399	<.0001 <sup>+</sup>	62	-38	36
	Supramarginal Gyrus, anterior division (R)	128	.310	<.0001 <sup>+</sup>	69	-20	39
	Lateral Occipital Cortex, superior division (L)	18	.342	.001	-17	-80	36
	Superior Parietal Lobule (R)	15	.273	<.0001 <sup>+</sup>	17	-53	57
	Inferior Frontal Gyrus, pars opercularis (R)	42	.270	<.0001 <sup>+</sup>	54	22	24
	Middle Frontal Gyrus (R)	29	.261	.001	47	7	54
	Postcentral Gyrus, lateral part (L)	15	-.247	.003	-55	-12	33
	Middle Frontal Gyrus (R)	21	-.226	<.0001 <sup>+</sup>	35	26	36
ICU Nearest Neighbors (Absolute Difference)	Inferior Frontal Gyrus, pars opercularis (R)	54	.283	.001	62	18	6
	Supramarginal Gyrus, Temporoparietal Junction (R)	114	.262	<.0001 <sup>+</sup>	69	-27	42
	Occipital Fusiform Gyrus (R)	32	.234	.001	17	-76	-18
	Temporal Occipital Fusiform Cortex (R)	30	.225	<.0001 <sup>+</sup>	32	-46	-18
	Superior Temporal Gyrus, anterior division (R)	13	.222	.002	58	-1	-6
	Anterior Cingulate Cortex (R)	31	.220	.001	2	41	3
	Intracalcarine Cortex (R)	17	.214	<.0001 <sup>+</sup>	2	-80	6
	Supramarginal Gyrus, posterior division (L)	33	.206	.001	-55	-42	27
	Inferior Frontal Gyrus, pars opercularis (R)	13	.188	.001	54	22	24
	Central Opercular Cortex (L)	11	-.135	<.0001 <sup>+</sup>	-47	-16	12
Superior Parietal Lobule (L)	11	-.122	<.0001 <sup>+</sup>	-36	-53	51	

(Continued)



**Table 2.** (Continued)

Condition & Model	Brain Region(s)	Cluster Size	Spearman $\rho$	P-value	MNI Peak Coordinate		
ICU Nearest Neighbors (Euclidean Distance)	Temporal Occipital Fusiform Cortex (R)	124	.310	<.0001 <sup>+</sup>	32	-46	-18
	Cerebellum Lobule VI (R)	124	.219	<.0001 <sup>+</sup>	13	-61	-24
	Cerebellum Lobule IV and V (L)	124	.180	.003	-2	-53	-12
	Intracalcarine Cortex (L)	77	.291	.001	-10	-83	0
	Occipital Fusiform Gyrus (L)	56	.275	<.0001 <sup>+</sup>	-32	-80	-21
	Occipital Fusiform Gyrus (R)	14	.236	.003	20	-87	-18
	Middle Occipital Gyrus (L)	11	.231	<.0001 <sup>+</sup>	-21	-95	3
	Frontal Operculum Cortex (R)	30	.225	.001	47	18	0
	Supramarginal Gyrus, anterior division (R)	43	.221	.001	69	-23	42
	Dorsolateral Prefrontal Cortex (R)	11	.214	<.0001 <sup>+</sup>	20	56	27
	Dorsolateral Prefrontal Cortex (R)	11	.175	.002	28	48	39
	Anterior Cingulate Cortex (R)	11	.167	.003	2	44	3
	Postcentral Gyrus, dorsomedial part (R)	14	-.155	.001	2	-38	69
	Posterior Cingulate Cortex (R)	20	-.144	<.0001 <sup>+</sup>	5	-31	42
<b>Afraid of You</b>							
Low ICU Alike	Intracalcarine Cortex (R)	18	.368	.002	5	-83	6
	Posterior Cingulate Cortex (R)	40	-.328	<.0001 <sup>+</sup>	5	-31	30
	Hippocampus (L)	41	-.316	<.0001 <sup>+</sup>	-32	-31	-6
	Lingual Gyrus (L)	41	-.105	.002	-25	-53	-3
	Cuneal Cortex (R)	19	-.285	.001	17	-72	30
	Superior Parietal Lobule	16	-.279	<.0001 <sup>+</sup>	-17	-53	63
	Middle Temporal Gyrus, posterior division (R)	45	-.278	<.0001 <sup>+</sup>	47	-20	-9
	Lateral Occipital Cortex, superior division (R)	16	-.252	.001	35	-80	42
	Lateral Occipital Cortex, superior division (R)	17	-.226	.003	39	-80	12
	Postcentral Gyrus, dorsomedial part (R)	19	-.217	<.0001 <sup>+</sup>	13	-42	66
	Middle Temporal Gyrus, temporooccipital part (R)	11	-.208	.002	62	-53	3
	Lateral Occipital Cortex, superior division (L)	31	-.184	.003	-55	-72	30
	Dorsolateral Prefrontal Cortex (L)	12	-.144	.002	-17	59	33
High ICU Alike	Posterior Cingulate Cortex (L)	29	.411	.002	5	-27	33
	Intracalcarine Cortex (L)	23	-.271	.002	-6	-76	6
ICU Nearest Neighbors (Absolute Difference)	Lateral Occipital Cortex, superior division (R)	28	-.152	<.0001 <sup>+</sup>	32	-83	39
	Lateral Occipital Cortex, superior division (L)	32	-.151	<.0001 <sup>+</sup>	-21	-61	54
	Middle Frontal Gyrus (R)	20	-.138	.001	43	11	30
	Paracingulate Gyrus (R)	20	-.132	.002	9	22	45
	Brain-Stem (L)	34	-.120	<.0001 <sup>+</sup>	-10	-46	-45
	Dorsolateral Prefrontal Cortex (R)	18	-.117	.001	35	56	27
	Lateral Occipital Cortex, inferior division (R)	17	-.111	.003	43	-72	15
	Lateral Occipital Cortex, superior division (R)	28	-.152	<.0001 <sup>+</sup>	32	-83	39

(Continued)

Table 2. (Continued)

Condition & Model	Brain Region(s)	Cluster Size	Spearman $\rho$	P-value	MNI Peak Coordinate		
ICU Nearest Neighbors (Euclidean Distance)	Intracalcarine Cortex (L)	52	.264	.001	-6	-80	3
	Frontal Operculum Cortex (L)	12	.245	<.0001 <sup>+</sup>	-44	26	0
	Middle Occipital Gyrus (L)	14	.244	<.0001 <sup>+</sup>	-21	-95	3
	Cerebellum Lobule VI (R)	37	.237	.001	17	-65	-27
	Superior Temporal Gyrus, anterior division (L)	13	.210	.001	-51	-12	-9
	Lateral Occipital Cortex, superior division (L)	18	-.170	<.0001 <sup>+</sup>	-10	-61	69
	Middle Temporal Gyrus, temporooccipital part (R)	43	-.170	.002	65	-57	6
	Lateral Occipital Cortex, inferior division (R)	43	-.162	<.0001 <sup>+</sup>	54	-76	12
	Middle Temporal Gyrus, temporooccipital part (R)	43	-.141	.001	43	-57	12
<b>Afraid for Self</b>							
Low ICU Alike	Caudate (R)	12	.174	.001	17	14	18
	Insular Cortex (L)	53	-.220	<.0001 <sup>+</sup>	-36	11	-6
	Putamen (L)	53	-.172	<.0001 <sup>+</sup>	-29	-12	6
	Cerebellum Lobule VI (L)	15	-.206	.001	-17	-68	-24
	Superior Parietal Lobule (R)	20	-.190	.003	28	-53	48
	Precuneus (R)	13	-.185	.003	13	-38	57
	Precentral Gyrus, lateral part (L)	17	-.180	.003	-29	-12	57
	Postcentral Gyrus, dorsomedial part (L)	13	-.166	.004	-17	-38	45
High ICU Alike	Central Opercular Cortex (L)	20	-.284	<.0001 <sup>+</sup>	-51	-5	9
	Superior Frontal Gyrus (R)	26	-.217	<.0001 <sup>+</sup>	2	14	63
	Lateral Occipital Cortex, superior division (R)	11	-.212	<.0001 <sup>+</sup>	28	-83	39
	Cerebellum Crus II (R)	13	-.173	.003	13	-80	-33
	Entorhinal cortex, anterior division (L)	18	-.153	<.0001 <sup>+</sup>	-25	3	-15
ICU Nearest Neighbors (Absolute Difference)	Inferior Frontal Gyrus, pars triangularis (R)	16	.240	<.0001 <sup>+</sup>	50	18	-3
	Angular Gyrus (R)	44	.233	<.0001 <sup>+</sup>	54	-50	18
	Frontal Pole extending to Dorsolateral Prefrontal Cortex (R)	11	.185	.001	17	63	15
	Precentral Gyrus, lateral part (R)	113	-.172	<.0001 <sup>+</sup>	43	-12	36
	Postcentral Gyrus, lateral part (R)	113	-.137	<.0001 <sup>+</sup>	62	-16	30
	Postcentral Gyrus, lateral part (L)	10	-.142	.001	-44	-16	33
	Precentral Gyrus, lateral part (L)	29	-.142	<.0001 <sup>+</sup>	-32	-8	45
	Brain-Stem (R)	27	-.141	.002	13	-38	-21
	Supramarginal Gyrus, posterior division (L)	11	-.137	<.0001 <sup>+</sup>	-51	-42	33
	Frontal Pole extending to Ventrolateral Prefrontal Cortex (R)	12	-.125	.001	47	44	-15
ICU Nearest Neighbors (Euclidean Distance)	Inferior Frontal Gyrus, pars triangularis (R)	16	.242	.001	50	18	-3
	Occipital Pole (L)	12	.222	<.0001 <sup>+</sup>	-17	-102	6
	Frontal Pole extending to Dorsolateral Prefrontal Cortex (R)	16	.217	.001	24	63	15
	Precuneus (L)	11	.194	.001	-2	-68	48

**Note.** Results are reported using a cluster-forming threshold at  $p < .005$  and cluster-extent threshold at  $k = 10$  with voxel size =  $3\text{mm}^3$ . <sup>+</sup> indicates a cluster surviving false discovery rate (FDR) correction at  $q = .05$ .  $\rho$  values represent the degree to which the specific cluster corresponds to the inter-subject model with positive values indicating higher correspondence and negative values indicating anticorrespondence. Regions were labeled using the Harvard Oxford Atlas, and organized by condition, intersubject model, positive-to-negative  $\rho$  values, and then posterior-to-anterior.

When expressers were described as afraid of the viewer (“Afraid of you” condition), IS-RSA results revealed that low-CU adolescents’ neural response patterns were similar while others’ were dissimilar in right intracalcarine cortex ( $\rho = .368$ ,  $p = .002$ ). High-CU adolescents’ exhibited similar neural response patterns while others’ were dissimilar in left posterior cingulate cortex (PCC) ( $\rho = .411$ ,  $p = .002$ ). We did not identify any regions that corresponded to the nearest neighbor model in this condition.

In the “Afraid for self” condition (i.e., when expressers were described as afraid for themselves), right caudate was the only cluster that corresponded to the low-CU alike model ( $\rho = .174$ ,  $p = .001$ ). We did not identify any clusters meeting significance threshold that corresponded to the high-CU alike model. The only clusters that corresponded to the nearest neighbor model were the pars triangularis portion of the right IFG ( $\rho = .240$ ,  $p < .0001$ ), right angular gyrus ( $\rho = .233$ ,  $p < .0001$ ), and right frontal pole ( $\rho = .185$ ,  $p = .001$ ).

Regions showing a negative relationship with any of the models (clusters showing structure that is anticorrelated with the model) are listed in Table 2, but (for clarity) only clusters showing a positive relationship across models are displayed in Figures 3–5. Thresholded and unthresholded group-level statistical maps for IS-RSA are available in Supplementary Materials.

### 3. Discussion

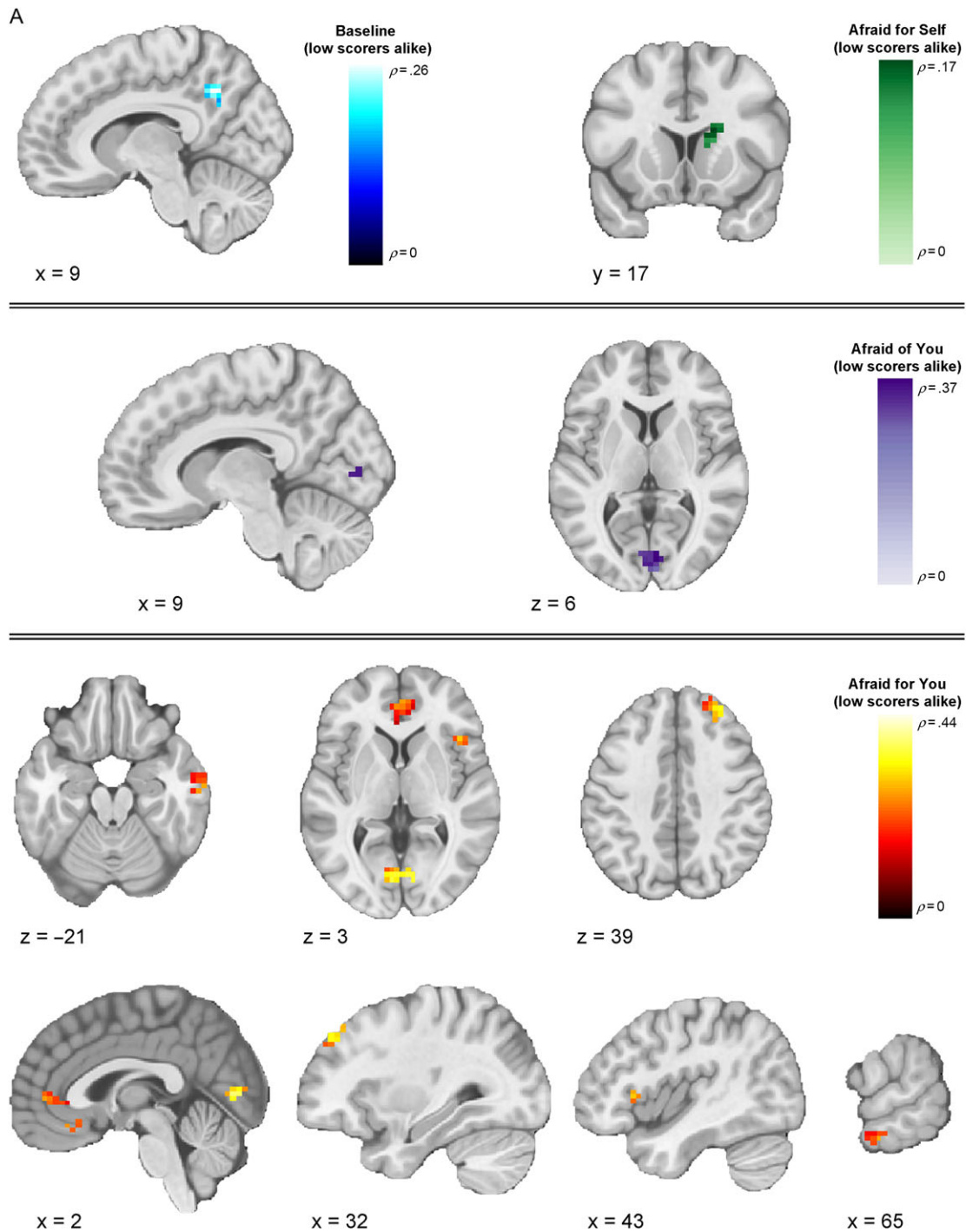
The present findings provide the first evidence that context-specific patterns of brain activity in response to fearful facial expressions align when people occupy similar positions in a phenotypic feature space of CU traits. IS-RSA identified how variations in adolescent- and parent-reported CU traits directly map onto variations in neural activation patterns associated with the context-specific interpretation of fearful facial expressions. Of note, we generally found activation patterns in regions implicated in low-level visual processing (e.g., occipital cortex), and the detection of personal threat from social cues (e.g., ACC) exhibited intersubject structure whereby low-CU adolescents were alike and others were dissimilar. By contrast, activation patterns in regions implicated in emotional face processing and perception (e.g., FFG, STG, pSTS) and higher-level social cognition (e.g., LPFC, mPFC, PCC, precuneus) more often showed intersubject structure whereby high-CU adolescents were alike while others were more dissimilar. These findings were highly right-lateralized, consistent with evidence for the right lateralization of emotional face processing and related socioaffective processes (Gläscher & Adolphs, 2003; Hung et al., 2010; Noesselt, Driver, Heinze, & Dolan, 2005). CU traits also shaped patterns of responses to fearful expressions as a function of interpretative contexts, providing insight beyond prior findings of reduced amygdala responsiveness to these expressions (Byrd, Kahn, & Pardini, 2013; Cardinale & Marsh, 2017; Carré et al., 2013; Ciucci, Baroncelli, Franchi, Golmaryami, & Frick, 2014; Essau, Sasagawa, & Frick, 2006; Houghton, Hunter, & Crow, 2013; Marsh et al., 2008) – a finding that was partially replicated here, with CU traits linked to reduced univariate amygdala responsiveness to fearful expressions only when expressers were described as afraid for themselves.

In line with evidence from our univariate findings regarding the importance of contextual framing, our IS-RSA analyses for the first-time link intersubject similarities in neural response patterns to contextualized fearful expressions to differential intersubject models of CU traits. Two superordinate novel patterns emerged

when youths varying in CU traits viewed fearful facial expressions across varying contexts. First, neural response patterns among low-CU adolescents became more idiosyncratic as emotional face information moved along the cortical processing hierarchy (as evidenced by the high-CU alike model). By contrast, neural response patterns aligned more among high-CU adolescents in regions along the entire cortical hierarchy, especially those that are implicated in social information processing, such as the FFG, STG, pSTS, PCC, and LPFC. These patterns may reflect these youths’ impairments in interpreting and responding to the social messages that fearful facial expressions convey. Second, these observations vary somewhat according to the context in which the fearful facial expressions are interpreted. Specifically, higher-order association regions exhibit intersubject neural activity pattern alignment even among low-CU adolescents when the context is specified. These findings align with recent work demonstrating the importance of contextual information for understanding emotional faces (Davis, Neta, Kim, Moran, & Whalen, 2016; Petro, Tong, Henley, & Neta, 2018).

Using IS-RSA to investigate neural responding to fearful expressions, we found that intersubject neural response patterns reflected intersubject differences in CU traits, especially in contexts where the expressers were described as afraid for the participant (“Afraid for you” condition). For example, we found higher alignment for high-CU participants in right FFA (while other participants show more idiosyncratic patterns). Known for its role in face identity processing (Kanwisher, 2017; Kanwisher, McDermott, & Chun, 1997), the right FFA also plays a role in emotion category discrimination, with selectivity for fearful facial expressions possibly resulting from feedback from the amygdala and mPFC, as suggested by findings from human fMRI and intracranial recordings (Harry, Williams, Davis, & Kim, 2013; Ishai, Pessoa, Bickle, & Ungerleider, 2004; Kawasaki et al., 2012). The observed pattern alignment among multivoxel patterns in this region may underpin impairments in the category-based perceptions of fearful facial expressions in CU adolescents, potentially from disrupted modulation from the amygdala.

We also observed differential alignment across high- and low-CU adolescents in regions implicated in emotional appraisal and introspection, as well as defensive responses to threats. Also within the “Afraid for you” context, low-CU adolescents were more similar (and other adolescents were most dissimilar to each other) in regions that included mPFC and ACC. Previous studies have found that perspective-taking, emotional appraisal, and cognitive empathy tasks recruit mPFC in both adults (Bzdok et al., 2012; Rameson, Morelli, & Lieberman, 2012; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Tusche, Bockler, Kanske, Trautwein, & Singer, 2016; Van Overwalle, 2009) and adolescents (Sebastian et al., 2012). For example, mPFC activation increases when adolescents perform tasks that require making inferences about social interactions (Kilford, Garrett, & Blakemore, 2016; Sebastian, Burnett, & Blakemore, 2008; Vollm et al., 2006). mPFC is also functionally correlated with the amygdala during fearful face processing, which is compromised in CU individuals (Blair, 2008; Breden, Cardinale, Lozier, VanMeter, & Marsh, 2015; Marsh et al., 2008). Evidence also demonstrates that ACC activates in response to threat-related stimuli (Bishop, Duncan, Brett, & Lawrence, 2004), exhibits strong structural and functional connectivity with the amygdalae (Carlson, Cha, & Mujica-Parodi, 2013; Kim et al., 2011; Williams et al., 2006), and displays blunted responses to others’ distress in CU youths (Lockwood et al., 2013). Together, the present findings suggest a deficit within these regions among

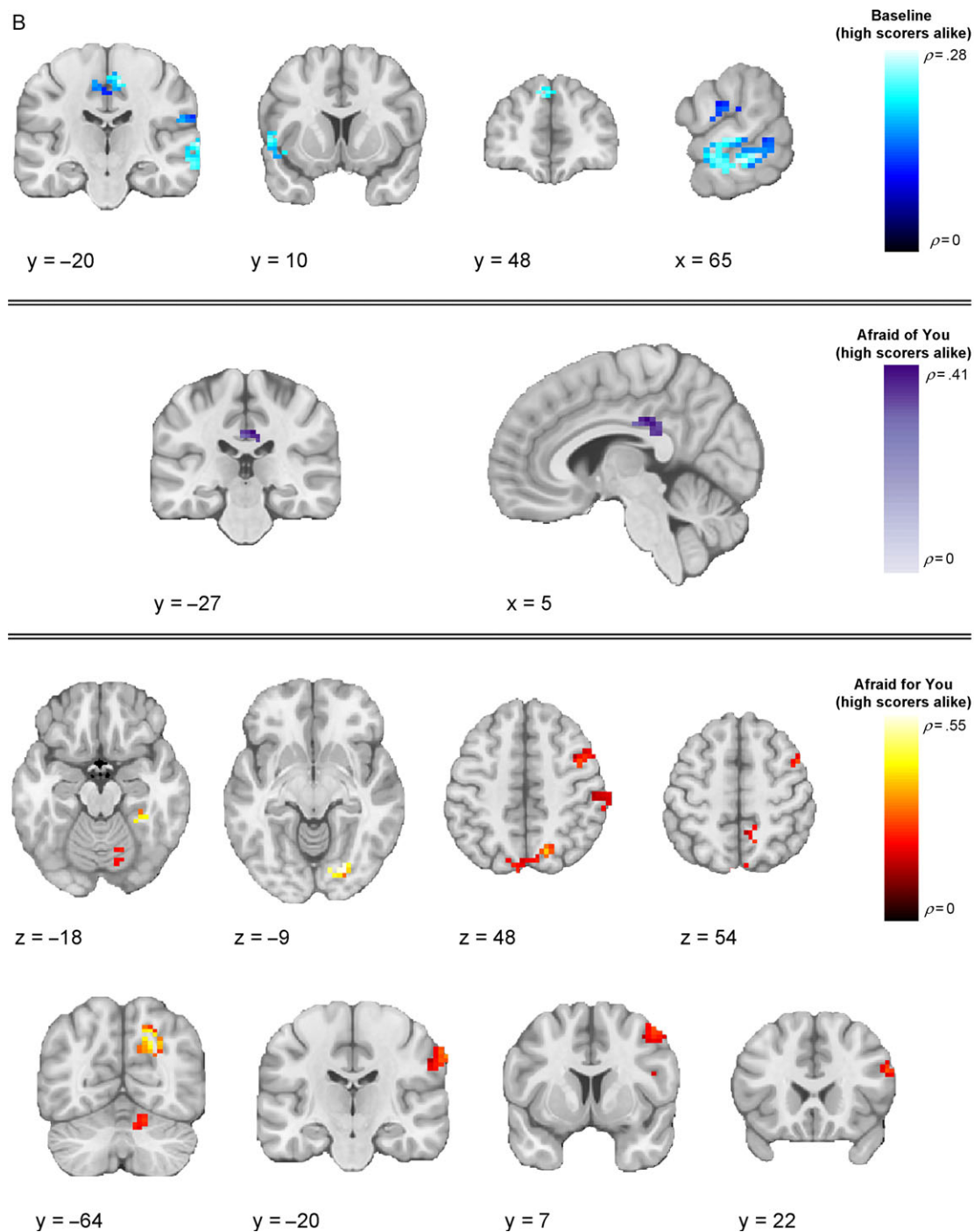


**Figure 3. Thresholded IS-RSA Results (Low CU Scorers Alike Model).** Visualization of clusters across conditions showing significant intersubject pattern response structure whereby low-CU adolescents were similar while all others were dissimilar. Clusters are thresholded at  $p < .005$  and  $k = 10$  (FDR-corrected results are reported in Table 2).

high-CU adolescents, particularly when viewing social cues that are contextualized as cooperative social signals.

Intersubject differences among subjects viewing fearful faces in the “Afraid of you” and “Afraid for self” conditions were less sensitive to the models testing relative and absolute differences among subjects. This could suggest that univariate patterns of activation better explain variation among CU adolescents within these contexts (particularly as we observed the stereotypical association between CU traits and reduced univariate amygdala response in

the “afraid for self” condition). Notably, when expressers were described as afraid for themselves, we observed that low-CU adolescents were more similar (and all others dissimilar) in right caudate. Individual differences in activation to fearful facial expressions within this region has been linked to CU traits (Lozier *et al.*, 2014), in which increased activation was related to lower CU traits. This finding in particular may reflect this region’s role in motivating prosocial approach towards relieving another’s personal distress (Schlund, Magee, & Hudgins, 2011).



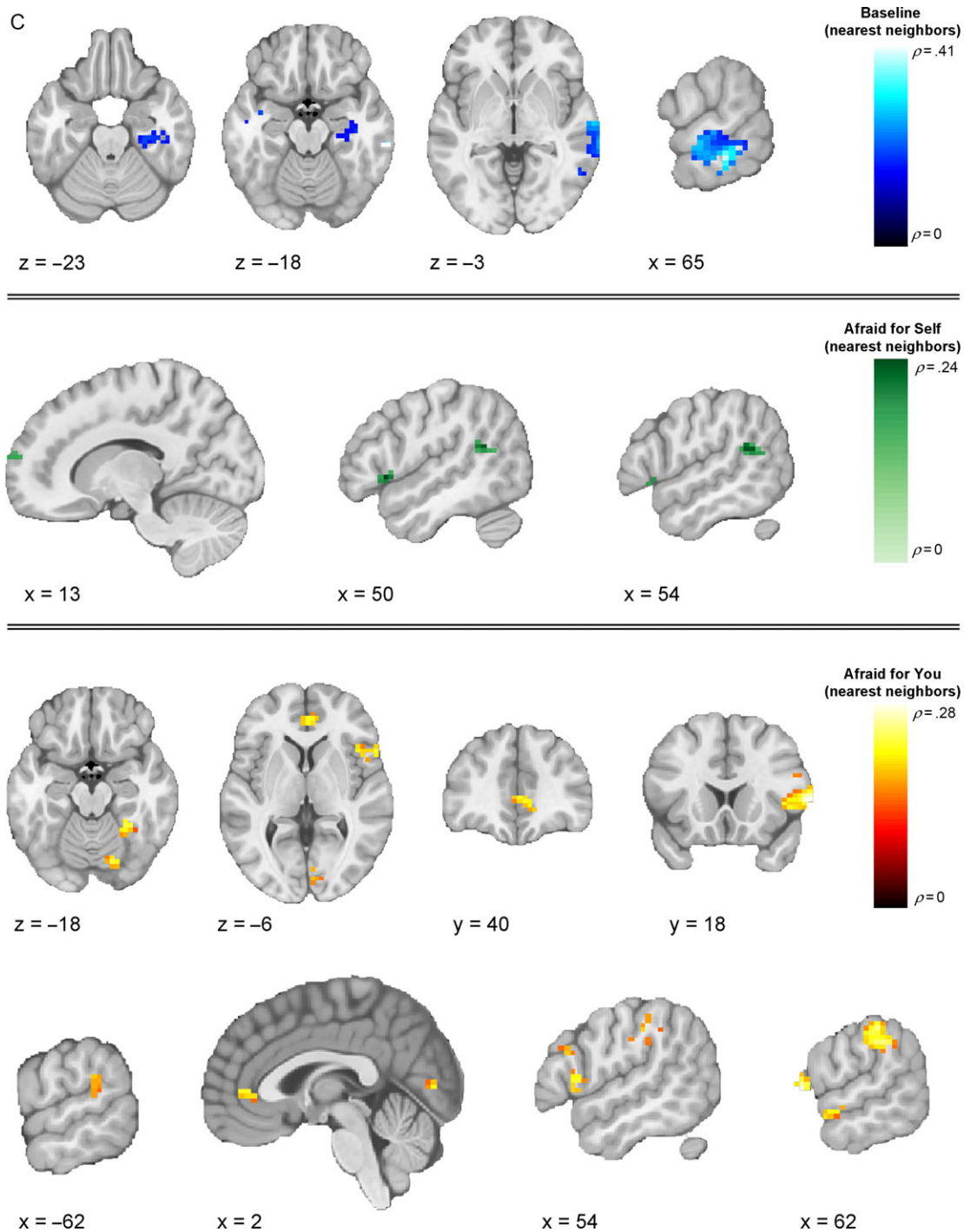
**Figure 4. Thresholded IS-RSA Results (High CU Scorers Alike).** Visualization of clusters across conditions showing significant intersubject pattern response structure whereby high-CU adolescents were similar while all others were dissimilar. Clusters are thresholded at  $p < .005$  and  $k = 10$  (FDR-corrected results are reported in Table 2).

### 3.1 Limitations

The reported findings should be considered in light of several limitations. First, our task was not selected for its ecological validity, but rather because the incorporated prompts allowed us to specify the context in which each expression should be interpreted. In the real world, of course, perceivers typically interpret others' facial expressions using a larger variety of rich contextual cues and semantic information that require shifts in attention and reference

to prior knowledge (Becker, 2009). Future studies should explore how more naturalistic cues related to the context of emotional facial expressions shape responses to and interpretations of these expressions.

Our MRI session included six runs of passively viewed fearful facial expressions, such that attentional requirements of the task should be considered. While the initial two baseline runs oriented and habituated subjects to the fearful expressions, and attentional



**Figure 5. Thresholded IS-RSA Results (Nearest Neighbors Model).** Visualization of clusters across conditions showing significant intersubject pattern response structure whereby adolescents were similar in CU traits regardless of being low or high. Clusters are thresholded at  $p < .005$  and  $k = 10$  (FDR-corrected results are reported in Table 2).

checks were included in the final four runs, it is possible that neural responses related to attention could vary as a function of CU traits. Mitigating this concern, however, is that we did not find a relationship between CU traits and accuracy in the attention checks. Our sample size also was limited by practical considerations and our goal of oversampling adolescents with moderate- to high-CU traits.

It is also important to note that greater similarity in neural activity patterns does not reflect increased or reduced activity.

High similarity could occur if two participants exhibit different mean levels of activation in a given brain region if the participants covary in their multivoxel activity patterns similarly. Additionally, our study utilizes anatomically-aligned data across pairs of subjects, but evidence suggests that functional alignment (e.g., hyperalignment) can improve detection of individual differences (Feilong, Nastase, Guntupalli, & Haxby, 2018). Our findings warrant further investigations to examine interindividual neural responses to the contextualization of fearful facial expressions in

association with CU traits and other psychological constructs – particularly given the novelty of the approach.

It should lastly be noted that although CU traits are often understood to be a unitary construct, recent studies found that ICU subcomponents (which include uncaring, callousness, and unemotionality) exhibit divergent associations with external psychological and neural variables (Cardinale et al., 2018; Cardinale & Marsh, 2017). In light of this, we calculated intersubject models for the ICU subscales separately and present these intersubject models in Supplementary Materials for comparison to the models constructed using ICU total scores.

#### 4. Conclusions

Prior studies have consistently indicated disrupted neural responses to fearful facial expressions in CU youths, and found evidence that these disruptions are closely linked to the disruptive behavior characteristic of CU traits. Nonetheless, many questions remain regarding the nature of the observed disruptions. While some of these questions reflect the standard approaches to assessing responses to fearful expressions in CU youths, the present study observed more specific information about how CU traits disrupt responding to fearful expressions both by constraining the context in which the expressions were interpreted and applying a novel analytic approach that yields greater specificity in mapping CU traits onto patterns of neural responses. This work underscores the utility of using techniques that explore interindividual variation in behavioral and personality characteristics and multivoxel patterns of neural responding.

**Data deposit.** Data and code are available on the Open Science Framework at <https://osf.io/bqdm5>.

**Supplementary material.** To view supplementary material for this article, please visit <https://doi.org/10.1017/pen.2020.13>.

**Acknowledgements.** We would like to thank Luke J. Chang, Emily S. Finn, and the instructors and participants of the Methods in Neuroscience at Dartmouth (MIND) computational summer school 2018, for helpful suggestions, discussions, and feedback regarding intersubject representational similarity analysis. We would also like to thank the Georgetown Center for Functional and Molecular Imaging staff for all of their assistance with data collection.

**Funding.** This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant R03 HD064906-01) to A.A.M., the National Science Foundation Graduate Research Fellowship Program Award to S.A.R., and the National Center for Advancing Translational Sciences Award (1KL2RR031974-01) to J.W.V. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (Grant ACI-1548562).

**Conflicts of interest.** Authors have nothing to disclose.

#### References

- Achenbach, T. M. (1991). Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington, VT: University of Vermont, Department of Psychiatry.
- Barry, C. T., Frick, P. J., Dshazo, T. M., McCoy, M. G., Ellis, M., & Loney, B. R. (2000). The importance of callous-unemotional traits for extending the concept of psychopathy to children. *Journal of Abnormal Psychology, 109*, 335–340. <https://doi.org/10.1037/0021-843X.109.2.335>
- Becker, M. W. (2009). Panic search: Fear produces efficient visual search for nonthreatening objects: Short Report. *Psychological Science, 20*, 435–437. <https://doi.org/10.1111/j.1467-9280.2009.02303.x>
- Bishop, S., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: Controlling attention to threat-related stimuli. *Nature Neuroscience, 7*, 184–188. <https://doi.org/10.1038/nn1173>
- Blair, R. J. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences, 27*, 135–145. [https://doi.org/10.1016/S0191-8869\(98\)00231-1](https://doi.org/10.1016/S0191-8869(98)00231-1)
- Blair, R. J. (2008). The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 2557–2565. <https://doi.org/10.1098/rstb.2008.0027>
- Blair, R. J., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology, 34*, 192–198. <https://doi.org/10.1111/j.1469-8986.1997.tb02131.x>
- Breeden, A. L., Cardinale, E. M., Lozier, L. M., VanMeter, J. W., & Marsh, A. A. (2015). Callous-unemotional traits drive reduced white-matter integrity in youths with conduct problems. *Psychological Medicine, 45*, 3033–3046. <https://doi.org/10.1017/S0033291715000987>
- Burke, J. D., Loeber, R., & Lahey, B. B. (2007). Adolescent conduct disorder and interpersonal callousness as predictors of psychopathy in young adults. *Journal of Clinical Child and Adolescent Psychology, 36*, 334–346. <https://doi.org/10.1080/15374410701444223>
- Byrd, A. L., Kahn, R. E., & Pardini, D. A. (2013). A validation of the inventory of callous-unemotional traits in a community sample of young adult males. *Journal of Psychopathology and Behavioral Assessment, 35*, 20–34. <https://doi.org/10.1007/s10862-012-9315-4>
- Bzdok, D., Schilbach, L., Voegeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function, 217*, 783–796. <https://doi.org/10.1007/s00429-012-0380-y>
- Cardinale, E. M., Connell, K. O., Robertson, E. L., Meena, L. B., Breeden, A. L., Lozier, L. M., ... Marsh, A. A. (2018). Callous and uncaring traits are associated with reductions in amygdala volume among youths with varying levels of conduct problems. *Psychological Medicine, 49*, 1449–1485. <https://doi.org/10.1017/S0033291718001927>
- Cardinale, E. M., & Marsh, A. A. (2015). Impact of psychopathy on moral judgments about causing fear and physical harm. *PLoS ONE, 10*, 1–21. <https://doi.org/10.1371/journal.pone.0125708>
- Cardinale, E. M., & Marsh, A. A. (2017). The reliability and validity of the Inventory of Callous Unemotional Traits: A meta-analytic review. *Assessment, 27*, 57–71. <https://doi.org/10.1177/1073191117747392>
- Carlson, J. M., Cha, J., & Mujica-Parodi, L. R. (2013). Functional and structural amygdala - Anterior cingulate connectivity correlates with attentional bias to masked fearful faces. *Cortex, 49*, 2595–2600. <https://doi.org/10.1016/j.cortex.2013.07.008>
- Carré, J. M., Hyde, L. W., Neumann, C. S., Viding, E., & Hariri, A. R. (2013). The neural signatures of distinct psychopathic traits. *Social Neuroscience, 8*, 122–135. <https://doi.org/10.1080/17470919.2012.703623>
- Chang, L., Sam, Jolly, E., Cheong, J. H., Burnashev, A., Chen, A., & Frey, S. (2019). cosanlab/NLTools. <https://doi.org/10.5281/ZENODO.3251172>
- Chen, P. A., Jolly, E., Cheong, J. H., & Chang, L. J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage, 216*, 1–10. <https://doi.org/10.1016/j.neuroimage.2020.116851>
- Ciucci, E., Baroncelli, A., Franchi, M., Golmaryami, F. N., & Frick, P. J. (2014). The association between callous-unemotional traits and behavioral and academic adjustment in children: Further validation of the inventory of callous-unemotional traits. *Journal of Psychopathology and Behavioral Assessment, 36*, 189–200. <https://doi.org/10.1007/s10862-013-9384-z>
- Davis, F. C., Neta, M., Kim, M. J., Moran, J. M., & Whalen, P. J. (2016). Interpreting ambiguous social cues in unpredictable contexts. *Social Cognitive and Affective Neuroscience, 11*, 775–782. <https://doi.org/10.1093/scan/nsw003>
- Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*, 13–34. <https://doi.org/10.1038/sj.mp.4000812>
- Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for

- facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, 36, 2288–2304. <https://doi.org/10.1016/j.neubiorev.2012.08.006>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31, 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, 13, 454–469. <https://doi.org/10.1177/1073191106287354>
- Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., . . . Gorgolewski, K. J. (2017). poldracklab/FMRIPrep. <https://doi.org/10.5281/ZENODO.1095198>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Aye, I., Erramuzpe, A., . . . Gorgolewski, K. J. (2018). FMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16, 111–116. <https://doi.org/10.1101/306951>
- Fanti, K. A., Kyranides, M. N., Georgiou, G., Petridou, M., Colins, O. F., Tuvblad, C., & Andershed, H. (2017). Callous-unemotional, impulsive-irresponsible, and grandiose-manipulative traits: Distinct associations with heart rate, skin conductance, and startle responses to violent and erotic scenes. *Psychophysiology*, 54, 663–672. <https://doi.org/10.1111/psyp.12837>
- Fanti, K. A., Panayiotou, G., Kyranides, M. N., & Avraamides, M. N. (2016). Startle modulation during violent films: Association with callous-unemotional traits and aggressive behavior. *Motivation and Emotion*, 40, 321–333. <https://doi.org/10.1007/s11031-015-9517-7>
- Feilong, M., Nastase, S. A., Guntupalli, J. S., & Haxby, J. V. (2018). Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage*, 183, 375–386. <https://doi.org/10.1016/j.neuroimage.2018.08.029>
- Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosyncrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, 215, 1–12. <https://doi.org/10.1016/j.neuroimage.2020.116828>
- Foulkes, L., & Blakemore, S. J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience*, 21, 315–323. <https://doi.org/10.1038/s41593-018-0078-4>
- Freiwald, W., Duchaine, B., & Yovel, G. (2016). Face processing systems: From neurons to real-world social perception. *Annual Review of Neuroscience*, 39, 325–346. <https://doi.org/10.1146/annurev-neuro-070815-013934>
- Frick, P. J., Cornell, A. H., Bodin, S. D., Dane, H. E., Barry, C. T., & Loney, B. R. (2003). Callous-unemotional traits and developmental pathways to severe conduct problems. *Developmental Psychology*, 39, 246–260. <https://doi.org/10.1037/0012-1649.39.2.246>
- Frick, P. J., & Ray, J. V. (2015). Evaluating callous-unemotional traits as a personality construct. *Journal of Personality*, 83, 710–722. <https://doi.org/10.1111/jopy.12114>
- Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology*, 21, 1111–1131. <https://doi.org/10.1017/S0954579409990071>
- Frick, P. J., & White, S. F. (2008). Research Review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 49, 359–375. <https://doi.org/10.1111/j.1469-7610.2007.01862.x>
- Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *NeuroImage*, 102, 309–316. <https://doi.org/10.1016/j.neuroimage.2014.07.047>
- Gläscher, J., & Adolphs, R. (2003). Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. *Journal of Neuroscience*, 23, 10274–10282. <https://doi.org/10.1523/jneurosci.23-32-10274.2003>
- Goldstein, J. M., Seidman, L. J., Makris, N., Ahern, T., O'Brien, L. M., Caviness, V. S., . . . Tsuang, M. T. (2007). Hypothalamic abnormalities in schizophrenia: Sex effects and genetic vulnerability. *Biological Psychiatry*, 61, 935–945. <https://doi.org/10.1016/j.biopsych.2006.06.027>
- Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27, 17–24. <https://doi.org/10.1023/A:1022658222914>
- Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Jarecka, D., . . . Ghosh, S. (2019). nipy/nipype. <https://doi.org/10.5281/ZENODO.2685428>
- Gorgolewski, K. J., Madison, C., Burns, C. D., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, 5, 1–15. <https://doi.org/10.3389/fninf.2011.00013>
- Hammer, J. L., & Marsh, A. A. (2015). Why do fearful facial expressions elicit behavioral approach? Evidence from a combined approach-avoidance implicit association test. *Emotion*, 15, 223–231.
- Harry, B., Williams, M. A., Davis, C., & Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Frontiers in Human Neuroscience*, 7, 1–6. <https://doi.org/10.3389/fnhum.2013.00692>
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430. <https://doi.org/10.1126/science.1063736>
- Hooker, C. I., Germine, L. T., Knight, R. T., & D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *Journal of Neuroscience*, 26, 8915–8922. <https://doi.org/10.1523/JNEUROSCI.3048-05.2006>
- Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or action requests? *Emotion*, 3, 150–166. <https://doi.org/10.1037/1528-3542.3.2.150>
- Houghton, S., Hunter, S. C., & Crow, J. (2013). Assessing callous unemotional traits in children aged 7- to 12-years: A confirmatory factor analysis of the inventory of callous unemotional traits. *Journal of Psychopathology and Behavioral Assessment*, 35, 215–222. <https://doi.org/10.1007/s10862-012-9324-3>
- Hung, Y., Smith, M. L., Bayle, D. J., Mills, T., Cheyne, D., & Taylor, M. J. (2010). Unattended emotional faces elicit early lateralized amygdala-frontal and fusiform activations. *NeuroImage*, 50, 727–733. <https://doi.org/10.1016/j.neuroimage.2009.12.093>
- Ishai, A. (2008). Let's face it: It's a cortical network. *NeuroImage*, 40, 415–419. <https://doi.org/10.1016/j.neuroimage.2007.10.040>
- Ishai, A., Pessoa, L., Bickle, P. C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences*, 101, 9827–9832. <https://doi.org/10.1073/pnas.0403559101>
- Jones, A. P., Laurens, K. R., Herba, C. M., Barker, G. J., & Viding, E. (2009). Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. *American Journal of Psychiatry*, 166, 95–102. <https://doi.org/10.1176/appi.ajp.2008.07071050>
- Jusyte, A., Mayer, S. V., Künzel, E., Hautzinger, M., & Schönberg, M. (2015). Unemotional traits predict early processing deficit for fearful expressions in young violent offenders: An investigation using continuous flash suppression. *Psychological Medicine*, 45, 285–297. <https://doi.org/10.1017/S0033291714001287>
- Kanwisher, N. (2017). The quest for the FFA and where it led. *The Journal of Neuroscience*, 37, 1056–1061. <https://doi.org/10.1523/JNEUROSCI.1706-16.2016>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311. <https://doi.org/10.1523/jneurosci.17-11-04302.1997>
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test*, Second Edition. Circle Pines, MN: AGS Publishing. <https://doi.org/10.1002/9781118660584.ese1325>
- Kawasaki, H., Tsuchiya, N., Kovach, C. K., Nourski, K. V., Oya, H., Howard, M. A., & Adolphs, R. (2012). Processing of facial emotion in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, 24, 1358–1370. [https://doi.org/10.1162/jocn\\_a\\_00175](https://doi.org/10.1162/jocn_a_00175)
- Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience and Biobehavioral Reviews*, 70, 106–120. <https://doi.org/10.1016/j.neubiorev.2016.08.016>



- Kim, M. J., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N., & Whalen, P. J. (2011). The structural and functional connectivity of the amygdala: From normal emotion to pathological anxiety. *Behavioural Brain Research*, 223, 403–410. <https://doi.org/10.1016/j.bbr.2011.04.025>
- Kimonis, E. R., Fanti, K. A., Goulter, N., & Hall, J. (2017). Affective startle potentiation differentiates primary and secondary variants of juvenile psychopathy. *Development and Psychopathology*, 29, 1149–1160. <https://doi.org/10.1017/S0954579416001206>
- Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., . . . Morris, A. S. (2008). Assessing callous-unemotional traits in adolescent offenders: Validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law and Psychiatry*, 31, 241–252. <https://doi.org/10.1016/j.ijlp.2008.04.002>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 1–28. <https://doi.org/10.3389/neuro.06.004.2008>
- Lockwood, P. L., Sebastian, C. L., McCrory, E. J., Hyde, Z. H., Gu, X., De Brito, S. A., & Viding, E. (2013). Association of callous traits with reduced neural response to others' pain in children with conduct problems. *Current Biology*, 23, 901–905. <https://doi.org/10.1016/j.cub.2013.04.018>
- Lozier, L. M., Cardinale, E. M., VanMeter, J. W., Marsh, A. A., Van Meter, J. W., & Marsh, A. A. (2014). Mediation of the relationship between callous-unemotional traits and proactive aggression by amygdala response to fear among children with conduct problems. *JAMA Psychiatry*, 71, 627–636. <https://doi.org/10.1001/jamapsychiatry.2013.4540>
- Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet. <https://doi.org/10.1017/S0048577299971664>
- Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., . . . Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, 83, 155–171. <https://doi.org/10.1016/j.schres.2005.11.020>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Marsh, A. A. (2016). Understanding amygdala responsiveness to fearful expressions through the lens of psychopathy and altruism. *Journal of Neuroscience Research*, 94, 513–525. <https://doi.org/10.1002/jnr.23668>
- Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion*, 5, 119–124. <https://doi.org/10.1037/1528-3542.5.1.119>
- Marsh, A. A., & Blair, R. J. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 32, 454–465. <https://doi.org/10.1016/j.neubiorev.2007.08.003>
- Marsh, A. A., Finger, E. C., Mitchell, D. G. V., Reid, M. E., Sims, C., Kosson, D. S., . . . Blair, R. J. (2008). Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *American Journal of Psychiatry*, 165, 712–720. <https://doi.org/10.1176/appi.ajp.2007.07071145>
- Marsh, A. A., Finger, E. C., Schechter, J. C., Jurkowitz, I. T. N., Reid, M. E., & Blair, R. J. (2011). Adolescents with psychopathic traits report reductions in physiological responses to fear. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 52, 834–841. <https://doi.org/10.1111/j.1469-7610.2010.02353.x>
- Mills-Koonce, W. R., Wagner, N. J., Willoughby, M. T., Stifter, C., Blair, C., & Granger, D. A. (2015). Greater fear reactivity and psychophysiological hyperactivity among infants with later conduct problems and callous-unemotional traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56, 148–154. <https://doi.org/10.1111/jcpp.12289>
- Muñoz, L. C. (2009). Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 554–562. <https://doi.org/10.1097/CHI.0b013e31819c2419>
- Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, 184, 161–170. <https://doi.org/10.1016/j.neuroimage.2018.09.010>
- Noesselt, T., Driver, J., Heinze, H. J., & Dolan, R. (2005). Asymmetrical activation in the human brain during processing of fearful faces. *Current Biology*, 15, 424–429. <https://doi.org/10.1016/j.cub.2004.12.075>
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 9599–9604. <https://doi.org/10.1073/pnas.1206095109>
- Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, 10, 1095–1102. <https://doi.org/10.1038/nn1968>
- Palmer, A. L. (2013). *Psychopathological contributions to the interpretation of ambiguous facial expressions*. Dartmouth College. Retrieved from <https://search.proquest.com/docview/1614153683>
- Pardini, D. A. (2006). The callousness pathway to severe violent delinquency. *Aggressive Behavior*, 32, 590–598. <https://doi.org/10.1002/ab.20158>
- Pardini, D. A., & Frick, P. J. (2013). Multiple developmental pathways to conduct disorder: Current conceptualizations and clinical implications. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 22, 20–25.
- Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a “low road” to “many roads” of evaluating biological significance. *Nature Reviews Neuroscience*, 11, 773–782. <https://doi.org/10.1038/nrn2920>
- Petro, N. M., Tong, T. T., Henley, D. J., & Neta, M. (2018). Individual differences in valence bias: fMRI evidence of the initial negativity hypothesis. *Social Cognitive and Affective Neuroscience*, 13, 687–698. <https://doi.org/10.1093/scan/nsy049>
- Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2012). The neural correlates of empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, 24, 235–245. [https://doi.org/10.1162/jocn\\_a\\_00130](https://doi.org/10.1162/jocn_a_00130)
- Robinson, J. L., Laird, A. R., Glahn, D. C., Lovallo, W. R., & Fox, P. T. (2010). Metaanalytic connectivity modeling: Delineating the functional connectivity of the human amygdala. *Human Brain Mapping*, 31, 173–184. <https://doi.org/10.1002/hbm.20854>
- Rosen, J. B., & Donley, M. P. (2006). Animal studies of amygdala function in fear and uncertainty: Relevance to human research. *Biological Psychology*, 73, 49–60. <https://doi.org/10.1016/j.biopsycho.2006.01.007>
- Rothmund, Y., Ziegler, S., Hermann, C., Gruesser, S. M., Foell, J., Patrick, C. J., & Flor, H. (2012). Fear conditioning in psychopaths: Event-related potentials and peripheral measures. *Biological Psychology*, 90, 50–59. <https://doi.org/10.1016/j.biopsycho.2012.02.011>
- Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., . . . Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage*, 54, 2524–2533. <https://doi.org/10.1016/j.neuroimage.2010.10.011>
- Salekin, R. T., Brannen, D. N., Zalot, A. A., Leistico, A. M., & Neumann, C. S. (2006). Factor structure of psychopathy in youth: Testing the applicability of the new four-factor model. *Criminal Justice and Behavior*, 33, 135–157. <https://doi.org/10.1177/0093854805284416>
- Schlund, M. W., Magee, S., & Hudgins, C. D. (2011). Human avoidance and approach learning: Evidence for overlapping neural systems and experiential avoidance modulation of avoidance neurocircuitry. *Behavioural Brain Research*, 225, 437–448. <https://doi.org/10.1016/j.bbr.2011.07.054>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Sebastian, C. L., Burnett, S., & Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences*, 12, 441–446. <https://doi.org/10.1016/j.tics.2008.07.008>
- Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, 7, 53–63. <https://doi.org/10.1093/scan/nsr023>
- Sebastian, C. L., McCrory, E. J., Dadds, M. R., Cecil, C. A. M., Lockwood, P. L., Hyde, Z. H., . . . Viding, E. (2014). Neural responses to fearful eyes in children with conduct problems and varying levels of callous-unemotional traits.

- Psychological Medicine*, 44, 99–109. <https://doi.org/10.1017/S0033291713000482>
- Tusche, A., Bockler, A., Kanske, P., Trautwein, F.-M., & Singer, T.** (2016). Decoding the charitable brain: Empathy, perspective taking, and attention shifts differentially predict altruistic giving. *Journal of Neuroscience*, 36, 4719–4732. <https://doi.org/10.1523/JNEUROSCI.3392-15.2016>
- van Baar, J. M., Chang, L. J., & Sanfey, A. G.** (2019). The computational and neural substrates of moral strategies in social decision-making. *Nature Communications*, 10, 1–14. <https://doi.org/10.1038/s41467-019-09161-6>
- Van Overwalle, F.** (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30, 829–858. <https://doi.org/10.1002/hbm.20547>
- Vasey, M. W., Kotov, R., Frick, P. J., & Loney, B. R.** (2005). The latent structure of psychopathy in youth: A taxometric investigation. *Journal of Abnormal Child Psychology*, 33, 411–429. <https://doi.org/10.1007/s10802-005-5723-1>
- Vollm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., ... Elliott, R.** (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, 29, 90–98. <https://doi.org/10.1016/j.neuroimage.2005.07.022>
- Williams, L. M., Das, P., Liddell, B. J., Kemp, A. H., Rennie, C. J., & Gordon, E.** (2006). Mode of functional connectivity in amygdala pathways dissociates level of awareness for signals of fear. *Journal of Neuroscience*, 26, 9264–9271. <https://doi.org/10.1523/JNEUROSCI.1016-06.2006>
- Wilson, K., Juodis, M., & Porter, S.** (2011). Fear and loathing in psychopaths: A meta-analytic investigation of the facial affect recognition deficit. *Criminal Justice and Behavior*, 38, 659–668. <https://doi.org/10.1177/0093854811404120>
- Zhen, Z., Fang, H., & Liu, J.** (2013). The hierarchical brain network for face recognition. *PLoS ONE*, 8, 1–9. <https://doi.org/10.1371/journal.pone.0059886>