SUPPLEMENTAL INFORMATION

Supplementary Materials

Visit 1: Psychophysiology

Participants

A total of 133 individuals underwent fear conditioning. However, 22 anxious youths and 7 healthy youths discontinued participation when they became anxious. Therefore, 23 anxious youths, 18 anxious adults, 42 healthy youths, and 31 healthy adults completed fear acquisition and extinction procedures in the clinic (Table 1).

Factors contributing to the discontinuation rates were determined using logistic regression analyses after removing a non-significant diagnosis×age interaction [$p>0.2$]. Similarly, logistic regression analyses were used to determine other factors in youths that may contribute to these discontinuation rates.

Data with technical problems from 1 anxious youth, 2 healthy youths, 1 healthy adult were excluded from analysis. Furthermore, skin conductance response (SCR) data from 3 anxious adults, 6 healthy youths, and 5 healthy adults in the final sample were not useable due to equipment failure; however, these participants did provide useable subjective and startle response data. Unlike other studies that excluded participants if they are deemed non-responders or fail to condition or extinguish (1), no other subjects were excluded from analysis.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Procedures

Stimuli

Black-white photographs of a light-haired (#01) and dark-haired (#03) woman displaying closed-mouth, neutral facial expressions were selected from the NIMstim facial expression set (2) as the conditioned stimuli (CS+, CS-). The woman selected to represent the CS+ was counterbalanced across participants. The CS+ and CS- were presented for 7-8 seconds followed by a grey screen presented for 8-21 seconds (averaging 15 seconds). The unconditioned stimulus (UCS) consisted of a 1-second image of the CS+ woman displaying an open-mouthed, fearful expression co-terminated with an aversive 95dB scream. The inter-stimulus interval (ISI) ranged from 11-15 seconds.

To measure fear-potentiated startle (FPS), startle probes (i.e., 40ms, 4-10 psi of compressed air delivered to the forehead) were presented during the CS+, CS- and the ISI. During CS+ and CS- trials, the startle probes occurred 5-6 seconds following onset of the face.

Apparatus and Acquisition

PsyLab psychophysiological recording system (PsyLab SAM System Contact Precision Instruments, London, www.psylab.com) presented the visual and auditory stimuli and recorded SCR and startle response. The procedure included four phases: startle habituation, pre-conditioning, fear acquisition, and extinction. Six startle probes in the absence of any stimuli were presented during the startle habituation phase. Each CS+ and CS- was presented four times during pre-conditioning, ten times during fear

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

acquisition, and eight times during extinction. During the fear acquisition phase, the CS+ was followed by a UCS with an 80% reinforcement schedule. No UCS was presented in pre-conditioning or extinction. In all phases, the CS+, CS-, and ISI were presented in blocked counterbalanced order.

Psychophysiological measures were collected continuously during each phase. SCR was measured by two Ag/AgCl electrodes filled with non-saline gel attached to the medial phalanx of the middle and ring fingers of the left hand according to published recommendations (3). FPS was measured by the eye blink startle reflex using two 6mm tin cup electromyography (EMG) electrodes filled with standard electrolyte solution placed under the subject's left eye. A ground electrode was attached to the subject's left forearm. SCR and EMG data were recorded using a sampling rate of 1000Hz, and EMG data was filtered using an amplifier bandwidth of 30-500Hz.

Before startle habituation, after fear acquisition, and after extinction, participants rated their fear level when viewing the CS+ and CS- using a ten-point Likert scale (1=none, 10=extreme). Ratings were collected after each phase to avoid interfering with the fear and safety learning process.

Analyses

To determine baseline measurements, group differences in pre-conditioning measures were examined using ANOVA. In all cases, significant effects were detected at α=0.05 level. Post-hoc analyses were conducted, using Bonferroni correction, when necessary.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Psychophysiology

Using PsyLab software, the SCR to each CS+ and CS- was determined by the difference between peak amplitude (within 1-5 seconds following stimulus onset) and baseline activity. The SCR data were square-root transformed. EMG data were rectified and smoothed using moving averages with 20ms windows. The EMG response to the startle probe during each CS+, CS-, and ISI was calculated as the difference between the peak EMG response (within 150ms following the startle probe) and the baseline activity (50 ms prior to the startle probe). Startle response data for each individual were standardized using a T-score transformation.

Transformed data were averaged across all trials and analyzed. Using diagnosis (anxious, healthy) and age-group (youth, adult) as between-subject factors and phase (pre-conditioning, fear acquisition, and extinction) and stimulus type (CS+, CS-) as within-subject factors, a repeated-measures ANOVA was used to test main effects and interaction effects. For startle response analysis, ISI was also included as a stimulus type. Greenhouse-Geisser and Bonferroni corrections were used where indicated. Significance was determined using α=0.05 and two-tailed tests. In addition, the square-root transformed SCR to the UCS was analyzed for group effects using repeated measures ANOVA.

Behavior

Behavioral analysis was analyzed in a similar manner as the psychophysiological data.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Correlations

Pearson correlations were used to examine relationships among age, psychophysiological responses and subjective ratings across all participants. Significant correlations are reported using α=0.05. Following Z-transformations, group differences in correlations were investigated using ANOVA.

Visit 2: fMRI

Participants

All participants completed the first visit of the study, which involved fear acquisition and extinction, prior to any treatment. Prior to the second visit, involving the MRI scan, some participants may have met with the therapist to begin establishing rapport; however, these sessions did not involve exposure, and none of the participants were on medication at the time of the scan.

Thirteen individuals discontinued participation in the study after completing fear acquisition and extinction. This group included 2 anxious youths, 1 anxious adult, and 1 healthy youth who discontinued due to fear of MRI procedures as well as 2 anxious youths and 3 healthy youths due to MRI contraindications; 1 anxious youth discontinued due to medication status; 2 healthy youths and 1 healthy adult could not be scheduled. Finally, data were excluded from 3 anxious youths, 1 anxious adult, 10 healthy youths, and 1 healthy adult due to excessive head movement (more than 75% data with >3mm) and 1 healthy youth, and 1 healthy adult due to technical problems.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Procedures

Stimuli

The stimuli included the neutral photographs of the two women used in the fear acquisition and extinction procedure. In addition, morphed images of the two women were used. Therefore, the stimuli correspond to 0% (CS-), 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% CS+.

Task

In each of two runs, 4 blocks of each cognitive instruction (threat appraisal, explicit memory, and perceptual discrimination) were presented in random order. In each block, all morphed images (*n*=11) and 2 blank images were presented randomly (Figure 1). Images were presented for 3000 ms with a 500 ms ISI.

The task was presented using E-prime computer software (PST Inc, Pittsburgh, PA) and front-projection. Participants viewed the screen via a mirror mounted on the head coil. A two-button box response device recorded the participants' responses.

Apparatus and Data Acquisition

Using standard sequences, MRI scans were collected with two 3T General Electric Signa scanners (Waukesha, WA) and an 8-channel head coil. To reduce signal dropout around the sinuses due to susceptibility artifact from air/tissue boundaries, an automated shim was used to calibrate the magnetic field. Two functional scans were acquired using a series of 36 contiguous 2.6 mm interleaved axial slices positioned to approximate the AC-PC line. These scans used a 96×96 matrix with echo-planar single shot gradient echo T2* weighting (TR=2300 ms; TE=25 ms; FOV=240 mm; flip

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

angle=90^{o}), yielding 2.5×2.5×2.6 mm voxels. To reach longitudinal magnetization equilibrium, four initial acquisition images were discarded prior to task onset. Finally, a high-resolution T1-weighted volumetric scan of the whole brain was acquired, using a magnetization prepared gradient echo sequence (MPRAGE) [124 1.2 mm axial slices; FOV=220 mm; NEX=1; TR=TE=min; matrix=256×192; TI=725 ms; bandwidth=31.25 kHz for 256 pixels]. This anatomical scan was used for co-registration and normalization procedures.

Analyses

Behavior

Subjective ratings and reaction time data were analyzed using SPSS 17.0 linear mixed model analysis. A full model included subject as a random factor, diagnosis (anxious, healthy), age-group (adult, youth), and cognitive instruction (threat appraisal, explicit memory, and physical characteristic) as fixed factors, linear (Morph-Level) and quadratic (Morph^2-Level) trends across individual morphed images as covariates, and all interactions among these factors. The linear and quadratic trend regressors consisted of the morphed level for each image and its square, respectively. In addition, scanner, days between conditioning and scanning, and IQ were included as additional covariates. Reported statistics derive from a reduced model, which included significant interactions and associated main effects detected in the full model. Significance was determined using α=0.05 and Greenhouse-Geisser correction, when applicable. Post-hoc tests with Bonferroni correction also examined 1) significant linear and quadratic trends 2) significant differences among selected morphed images (CS-, 50% CS+, 100% CS+) and

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

3) the boundary between CS- (on-line response) or 50% CS+ (for reaction time) and the other morphed images.

Although interactions with diagnosis and age-group were non-significant, group differences in response to CS+ have been found in prior work examining only two stimuli (CS+ and CS-). Therefore, post-hoc analyses examined group differences in response to select morphed images during threat appraisal and explicit memory, *i.e.*, the 100% CS+ and the morphed image capturing the threat/safety boundary (70% CS+ for threat appraisal, 60% for explicit memory). Group differences in subjective response to these morphed images were interrogated using separate univariate ANOVAs.

MRI Analysis

fMRI data were analyzed with AFNI software (4). For each subject, standard pre-processing of echo-planar data included slice-time correction, motion correction, spatial normalization to the Talaraich template using manual methods, and spatial smoothing with a 6 mm full-width half-maximum (FWHM) smoothing kernel. BOLD data were temporally scaled at the voxel level so that regression coefficients can be interpreted as percent signal change relative to the baseline mean.

For each subject, a general linear model included condition regressors of interest, six regressors modeling residual motion effects, and regressors modeling linear trends across time for each run. To form each condition regressor, the stimulus onsets for each condition were convolved with a gamma function approximating the hemodynamic response. Thirty-three condition regressors corresponded to each morphed image (11) for each cognitive instruction (3). A series of estimated betas, one for each regressor, was

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

generated to minimize the error term within the model. Beta coefficients from each condition regressor at the individual level were included in the group-level analysis.

Whole-brain group analyses were conducted in AFNI software using in-house adaptations of specially-developed linear mixed modeling (adapted version of 3dLME) (5). All 33 effects of interest for each individual were included in the group-level model. Interaction effects (e.g., diagnosis×age-group×instruction×$Morph^2$-Level) were detected using both a voxel-wise probability threshold and a spatial extent threshold. The spatial extent threshold was determined using AFNI's AlphaSim program (6) using cluster-level probability $\alpha=0.05$, a voxel-wise threshold $p<0.005$, 1000 Monte Carlo simulations, smoothness of 9.11, 9.03, 8.21 mm along x, y, and z axes.

In general, two approaches have been used to correct from multiple comparisons in relatively large regions, such as the ventral prefrontal cortical region targeted in the current study. One approach uses multiple spheres each placed around activations found in prior studies (7). The other approach uses one relatively large sphere, which encompasses all of the various locations implicated based on prior studies. While the first approach minimizes Type II errors and extends directly from prior studies, the second approach is more conservative, as it requires a larger spatial extent to consider the activation "significant". Since this is the first study of its kind, we adopted this second, more-conservative approach. This approach has been used previously in similar situations (8, 9), specifically in papers using AFNI (10). As a result, the approach was used here, defining the single ventral prefrontal cortex regional mask based on the standard template in AFNI for this region. When multiple activations appear in a region

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

with this approach, the procedures for considering each region as distinct from the other are the same as when examining activations arising from a whole-brain approach. The only difference with this region-based approach is that only activations within the region are considered.

With this approach, small volume correction regional masks were used for hypothesized regions, ventral regions of the medial prefrontal cortex (11453 mm^3, effectively a 14 mm radius sphere) and amygdala/hippocampus (1828 mm^3, effectively a 7 mm radius sphere). Brain regions were identified based on known anatomical landmarks identified on a single structural image and then cross-referenced with a Talairach atlas (11). All results are reported in left-posterior-inferior (LPI) coordinates and reflect the peak activation voxel.

Procedures described above were used to determine the locations of between-group differences, falling within the ventral prefrontal cortical regional mask. When significant activations emerged in this initial approach, extracted data within significant voxels in each location were then subjected to the secondary analysis. Specifically, to examine the diagnosis group (anxious, healthy)×age-group (adult, youth)×cognitive instruction (threat appraisal, explicit memory, physical discrimination)×$Morph^2$ interaction, a functionally-defined cluster was limited to active voxels surpassing our $p<.005$ threshold in each of the two regions emerging from our initial *a priori* analysis. These clusters comprised the subgenual anterior cingulate (sgACC, [-9, 26, -9]) and ventromedial prefrontal cortex (vmPFC, [4, 49, -6]). Using linear mixed models in SPSS, the average percent signal change values, relative to baseline, within these clusters were

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

examined in a similar manner as the behavioral data. To decompose the higher-level interactions, each cognitive instruction was examined separately.

Correlations

Significant Pearson correlations among age, conditioning and extinction measures and percent signal change values were tested using α=0.05. Following Z-transformations, group differences in correlations were investigated using ANOVA.

Supplementary Data

Participant Characteristics

Table 1 and S1 summarize characteristics across the four study groups. Groups were well-matched within each age group [all p>0.2]. Nevertheless, as a whole, adolescents had lower IQ than adults [$t(112)$=4.2, p<0.001]. IQ was co-varied in all analyses but did not predict any outcome measures [all p>0.2].

Visit 1: Fear acquisition and extinction

Discontinuation Rates

None of the adults, but 29 youths, discontinued the procedures due to excessive fear. Anxious and younger participants (anxious: n=22, 15 females, 10.7±2.3 years; healthy: n=7, 4 females, 10.2±1.6 years) were more likely to discontinue the procedures. Both findings represented large effects [diagnosis odds ratio: 4.2; odds-ratio for a five-year increment in age: 10.3, both p<0.02]. In youth, females also tended to discontinue more frequently than males [odds-ratio: 5.5, p<0.06]. Finally, anxious youths discontinuing the procedures were rated by clinicians as having more anxiety [odds-ratio

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

for one unit increase on Pediatric Anxiety Rating Scale (PARS): 1.4, $p<0.053$]. However, no differences between youths who discontinued and completed were found for any other measure, including IQ, other symptom scales, or responses to the conditioning procedures [all $p>0.1$].

Psychophysiology

During pre-conditioning, no CS or group differences in conditioning were detected using either physiological measure (i.e., SCR or startle response) [all $p>0.1$]. With respect to startle response, no group differences were detected during startle habituation [all $p>0.7$].

Psychopathology did not impact differential conditioning for either physiological measure [all $p>0.1$]. As shown in Figure S1, following fear acquisition, the response to the CS+ was greater than to the CS-, for both SCR [phase×CS-type: $F(2,192)=7.8$, $p<0.001$] and startle response [phase×CS-type: $F(4,440)=9.2$, p<0.001].

Two age differences were detected, though neither was moderated by anxiety status. First, SCR was greater in youths than adults during fear acquisition and extinction but no difference between healthy youth and adults were detected during pre-acquisition [diagnosis×age-group×phase: $F(2,192)=4.5$, p<0.02; $p>0.5$, all other $F(1,96)>12.7$, p<0.02]. Similar to the response to CS, youths (1.1±0.6 microsiemens) had a greater SCR response to the UCS during fear acquisition than the adults (0.7±0.7 microsiemens) [$F(1,98)=10.3$, $p<0.002$].

In contrast, there was evidence of CS-specific differences in startle response revealed by an age-group×phase×CS-type interaction [$F(4,440)=2.6$, p<0.04]. This three-

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

way interaction reflected age-group differences in EMG after extinction. Namely, following extinction, startle response to CS+ and CS- were more similar in adults than in youths. This represents lower levels of extinction in youths than adults [age-group×CS-type: $F(2,220)=4.2$, $p<0.02$]. However, the startle response was similar in the last pairs of trials [$p>0.6$], with no significant effect of age-group [$p>0.2$].

Subjective Response

Prior to conditioning, subjective reports of anxiety towards CS+ compared to CS- were rated similarly across all groups [all $p>0.2$].

As shown in Figure S1, the anxious groups reported being more afraid of both the CS+ and CS- than the healthy groups [$F(1,109)=19.3$, $p<0.001$]. This finding reflected greater fear in both anxious adults and youths, relative to their respective, age-matched healthy comparison groups. Of note, this finding only emerged for the CS cues; groups did not differ in their anxiety ratings of the UCS (anxious youths: 6.9±2.2, anxious adults: 6.8±2.5, healthy youths: 6.5±2.3, healthy adults 6.0±2.3) or the ISI (anxious youths: 2.4±1.1, anxious adults: 3.2±2.1, healthy youths: 2.8±2.2, healthy adults 2.4±1.6) [all $p>0.2$]. Moreover, this finding did not relate to conditioning. During fear acquisition, all four groups demonstrated differential conditioning [$F(1,110)=53.0$, $p<0.001$], but groups did not differ in their anxiety ratings of the CS+ relative to the CS- [all $p>0.5$].

Two other anxiety-related group differences manifested selectively following extinction (Figure S1). These differences contributed to a four-way, diagnosis×age-

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

group×phase×CS-type interaction [$F(2,218)=4.5$, $p<0.02$]. First, both anxious groups showed deficient extinction; anxiety ratings to the CS+ decreased from acquisition through extinction in the two healthy groups [both $F(2,108)>15.6$, both $p<0.04$], reflecting successful extinction learning. However, no such decreases occurred in either of the two anxious groups [both $p>0.2$]. Second, only anxious youths exhibited similarly high levels of anxiety ratings to the CS+ and CS- after extinction [$p>0.2$]. In the other three groups, anxiety was lower to the CS- than to the CS+ [all $F(1,109)>10.1$, $p<0.002$].

All participants reported that the CS+ (58.6%±22.7) screamed more in the past than the CS- (0%±0). However, three anxious youths, 5 anxious adults, 8 healthy youths, and 5 healthy adults reported the scream was unpredictable. Six anxious youths, 8 anxious adults, 13 healthy youths, and 7 healthy adults reported the scream predictable but the explanation did not pertain to the CS. With this open-ended question, individuals may have thought the contingency was more complicated than reporting who screamed.

Correlational data

Across all participants in the four subject groups, the subjective ratings of the conditioning effects (CS+ > CS-) increased as the subjectively-reported aversiveness of the screaming lady increased [$R(94)=0.34$, $p<0.001$]. During fear acquisition, conditioning effects in SCR were positively correlated with startle response [$R(98)=0.22$, $p<0.03$], albeit weakly. Across all participants, the conditioning effects measured by psychophysiological measures (i.e., SCR in acquisition [$R(98)=-0.26$, $p<0.008$] and startle response during extinction [$R(112)=-0.2$, $p<0.04$]) negatively correlated with age,

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

indicating that younger participants were more physiologically discriminating on these measures. No significant group differences in these correlations were noted [all $p>0.15$].

Visit 2: Recall

Subjective response

Responses to all cognitive instructions had a significant positive quadratic trend [$Morph^2$-Level: $F(1,1061)=35.5$, $p<0.001$; all $t(899)>2.0$, all $p<0.05$]; however, the quadratic pattern differed based on cognitive instruction [instruction×$Morph^2$-Level: $F(2,1931)=4.2$, $p<0.001$]. The responses for the 100% CS+ and 50% CS+ were greater during explicit memory than the threat appraisal task [both $F(2,1810)>63.8$, $p<0.001$], illustrating the more pronounced quadratic pattern in the explicit memory task. This effect was also captured by the different threat/safety boundaries. Compared to the pure CS- image (0% CS+), participants identified the CS+ based on explicit memory when an image contained at least 60% CS+ features [$F(10,2570)=82.7$, $p<0.001$; all $p<0.001$] and reported more fear when an image contained at least 70% CS+ features [$F(10,2570)=8.9$, $p<0.001$; all $t(81)>4.2$, all $p<0.05$].

Interactions with diagnosis and age-group were non-significant. Nevertheless, to extend results from prior studies, post-hoc analyses did examine group differences in response to selected morphed images. At these identified threat/safety boundaries, anxious groups, relative to the two healthy groups, did report significantly more fear during threat appraisal [$F(1,78)=4.7$, $p<0.03$] and tended to remember the CS+ better [$F(1,78)=3.2$, $p<0.08$]. In addition, adults remembered the 100% CS+ to a greater extent

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

than youths [$F(1,78)=4.3$, $p<0.04$], and anxious participants tended to rate more fear [$F(3, 78)=3.7$, $p<0.06$] and show better memory for the 100% CS+ than healthy groups [$F(1,78)=3.6$, $p<0.06$]. In separate analyses based on CS+ assignment (i.e., light-hair or dark-hair), no group differences or interaction effects were noted [all $p>0.1$], suggesting that all groups discriminated the physical characteristics of the stimuli similarly.

The quadratic pattern of the explicit memory task was also more pronounced than the physical discrimination task [$t(1769)=2.4$, $p<0.02$]. Due to counterbalancing of the CS+, the physical discrimination responses across participants were approximately 0.5 at every morphed image; however, the physical discrimination response pattern differed based on the CS+ assignment as expected given that the cognitive instruction pertained to whether the woman had jet black hair. When the CS+ was the dark-haired woman, a quadratic and positive linear response were detected [regression: both $F(2,448)>743.2$, $p<0.001$; both $\beta>0.005$, $t(450)>4.6$, $p<0.001$]. When the CS+ was the light-haired woman, a negative linear response was detected [regression: $F(2,447)=649.2$, $p<0.001$; $t(449)=-9.7$, $p<0.001$]. Responses to the physical discrimination task were less than the explicit memory task but greater than the threat appraisal task for the CS-, 50% CS+ and 100% CS+ [all $F(2,1810)>63.8$; $p<0.001$; all $p<0.003$].

Reaction Time

As shown in Figure S2, no effects of diagnosis or age-group were detected [all $p>0.2$]; however, the reaction time pattern across morphed images differed based on cognitive instruction [instruction×Morph^2-Level interaction: $F(2,1960)=19.3$, $p<0.001$;

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

instruction×Morph-Level interaction: $F(2,1968)=16.3$, $p<0.001$; instruction: $F(3,1057)=245.1$, $p<0.001$].

All cognitive processes exhibited a positive linear term in reaction time [Morph^2-Level: $F(1,636)=34.6$, $p<0.001$; all $t(899)>2.1$, all $p<0.04$], but only explicit memory and physical discrimination exhibited a negative quadratic pattern as well [instruction×Morph^2-Level: $F(2,1960)=19.3$, $p<0.001$; both $t(899)>-6.5$, both $p<0.001$]. Reaction times to the explicit memory and the physical discrimination task showed a greater inverted "U" pattern (*i.e*., a negative quadratic term and positive linear term) than the threat appraisal task [all $t(1907)>4.6$, all $p<0.001$]. In reference to the CS-, participants required more time when judging the 40%, 50%, and 60% CS+ in the physical discrimination task [$F(10,1496)=5.8$, $p<0.001$, all $p<0.03$] and 50%, 60%, and 70% in the explicit memory task [all $p<0.02$]; there were no differences in the threat appraisal task [$p>0.9$].

Neuroimaging data

No region exhibited a four-way interaction corrected for the whole-brain level. Beyond the two regions in the ventral regions of the medial prefrontal cortex, only two other regions across the entire brain, cerebellum [(11, -41, -39), 66 voxels, $F(2,2592)=12.1$] and superior frontal gyrus [(1, 56, 26), 22 voxels, $F(2,2592)=6.1$], surpassed the $p<0.005$ and 15 voxel threshold, with voxel size of 15.625 mm^3.

Interaction effects among diagnostic group, age-group, instruction, and trend among morphed images were not detected in the amygdala or hippocampus. However, additional analyses were conducted to confirm the sensitivity of the imaging methods to

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

amygdala and hippocampal engagement. Specifically, across all participants, morphed images with a larger CS+ contribution (above 50%) were compared to all morphed images with a larger CS- contribution (below 50%) using AFNI 3dttest++ program, including mean-centered covariates of non-interest to mimic the main analysis. Significant activation was identified using $p<0.05$ corrected levels for the amygdala and hippocamus. Of note, in the explicit memory task, amygdala regions [(21, -1, -6), extending from putamen, 11 voxels, Z=-3.15] and hippocampal regions [(19, 21, -9), 14 voxels, Z=-3.30; (29, 34, -9), 14 voxels, Z=-2.83] deactivated (i.e., responded more to the all CS- relative to the all CS+ morphed images), providing evidence that these regions are involved in recall; however, no interactions between diagnostic or age groups were detected.

Correlational data

Across all fMRI participants, self-report ratings of conditioning effects (CS+>CS-) following fear acquisition and extinction correlated positively with the post-scan ratings following recall several weeks later [acquisition: $R(79)=0.22$, $p<0.05$, extinction: $R(79)=0.46$, $p<0.001$], indicating moderate stability of levels of subjective fear during recall. As with post-scan ratings, self-report ratings of fear during acquisition and fear extinction correlated positively with online ratings during recall several weeks later [acquisition: $R(80)=0.33$, $p<0.002$, extinction: $R(80)=0.22$, $p<0.04$]. No group differences were noted [all p>0.2].

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Across all participants, neither on-line ratings of fear during threat appraisal nor post-scan ratings correlated with neural activation in either sgACC or vmPFC regions for any cognitive process (e.g., threat appraisal and explicit memory) during recall [all $p>0.4$].

Correlations of conditioning effects (CS+>CS-) also emerged for self-reported ratings during fear conditioning and brain-imaging measures assessed several weeks later. Subjective ratings following fear acquisition negatively correlated with data acquired several weeks later, both for the sgACC [$R(80)=-0.25$, $p<0.02$] and vmPFC [$R(80)=-0.24$, $p<0.03$] activations during threat appraisal. For the sgACC, this behavior-brain association varied by group, with stronger negative correlations in anxious youth [$R(12)=-0.71$, $p<0.005$] than the other groups [diagnosis×age-group×covariate: $F(1,75)=5.9$, $p<0.02$, all other groups $R<-0.07$, $p>0.8$]. Across all participants, a similar negative correlation between subjective ratings during extinction and brain activation manifest only in sgACC [$R(80)=-0.23$, $p<0.03$]. Compared to adults who showed no correlation [$p>0.7$], youths have stronger negative correlation between subjective rating during extinction and sgACC during threat appraisal during recall [$R(37)=-0.39$, $p<0.02$, age-group×covariate: $F(1,75)=4.75$, $p<0.03$]. No group differences in vmPFC were detected [p>0.1].

In addition, startle response of conditioning effects (CS+>CS-) during fear acquisition correlated positively with conditioning effects in the sgACC [$R=0.22$, $p<0.05$] and vmPFC when appraising threat during recall [$R(80)=0.30$, $p<0.006$] . Group differences in this correlation were detected, but only in the sgACC. In the sgACC,

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

youths exhibited a positive correlation between these measures [$R(37)=0.35$, $p<0.03$]; whereas, adults had a non-significant correlation [$p>0.5$; age-group×covariate: $F(1, 75)=4.88$, $p<0.03$]; however, there were no group differences in the correlations in the vmPFC [$p>0.5$].

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Supplementary References

1. Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: role of the amygdala and vmPFC. Neuron. 2004 Sep 16;43(6):897-905.

2. Tottenham N. NimStim Face Stimulus Set. [cited 2009 January]; Available from: http://www.macbrain.org/resources.htm.

3. Prokasy WF, Ebel HC. Three components of the classically conditioned GSR in human subjects. Journal of Experimental Psychology. 1967;73:247-56.

4. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996 Jun;29(3):162-73.

5. Chen G, Saad ZS, Britton JC, Pine DS, Cox RW. Linear Mixed-Effects Modeling Approach to FMRI Group Analysis. Neuroimage. in press.

6. Ward BD. AlphaSim. 2000 [cited 2010 May]; Available from: http://afni.nimh.nih.gov/afni/doc/manual/AlphaSim.

7. Siegle GJ, Thompson WK, Collier A, Berman SR, Feldmiller J, Thase ME, Friedman ES. Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. Arch Gen Psychiatry. 2012 Sep;69(9):913-24.

8. Beesdo K, Lau JY, Guyer AE, McClure-Tone EB, Monk CS, Nelson EE, Fromm SJ, Goldwin MA, Wittchen HU, Leibenluft E, Ernst M, Pine DS. Common and distinct amygdala-function perturbations in depressed vs anxious adolescents. Arch Gen Psychiatry. 2009 Mar;66(3):275-85.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

9. Linke J, King AV, Rietschel M, Strohmaier J, Hennerici M, Gass A, Meyer-Lindenberg A, Wessa M. Increased medial orbitofrontal and amygdala activation: evidence for a systems-level endophenotype of bipolar I disorder. Am J Psychiatry. [Research Support, Non-U.S. Gov't]. 2012 Mar;169(3):316-25.

10. Monk CS, Nelson EE, McClure EB, Mogg K, Bradley BP, Leibenluft E, Blair RJ, Chen G, Charney DS, Ernst M, Pine DS. Ventrolateral prefrontal cortex activation and attentional bias in response to angry faces in adolescents with generalized anxiety disorder. American Journal of Psychiatry. 2006 Jun;163(6):1091-7.

11. Talairach J, and Tournoux, P. Co-Planar Stereotactic Atlas of the Human Brain. Thieme, Stuttgart.1988.

12. RUPP Anxiety Study Group. The Pediatric Anxiety Rating Scale (PARS): development and psychometric properties. J Am Acad Child Adolesc Psychiatry. 2002;41:1061-9.

13. Spielberger CD, Edwards CD, et al. STAI. [cited 2009 January]; Available from: http://www.mindgarden.com/products/staisad.htm.

14. Spielberger CD, Edwards CD, Montuori J, Lushene R. STAI-C. [cited 2009 January]; Available from: http://www.mindgarden.com/products/staisch.htm#about.

15. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, Neer SM. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997 Apr;36(4):545-53.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

16. Beck AT, Steer, R.A., & Brown, G.K. . "Manual for the Beck Depression Inventory-II". San Antonio, TX: Psychological Corporation 1996

17. Bartko J. The intraclass correlation coefficient as a measure of reliability. Psychological Reports. 1966;19:3-11.

18. Kovacs M. Children's Depression Inventory [cited 2009]; Available from: http://www.pearsonassessments.com/depressioninvent.aspx.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)

Supplementary Table S1: Anxiety Symptomatology [a]

			PARS		STAIc/STAI state[b]		SCARED[b]		CDI/BDI[b]	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Fear Acquisition and Extinction	Anxious	Youths	13.7	3.8	35.8	4.9	37.4	18.1	14.0	8.0
		Adults	-	-	45.3	11.1	-	-	6.6	3.8
	Healthy	Youths	-	-	28.6	3.5	8.9	6.0	2.3	2.7
		Adults	-	-	25.6	6.1	-	-	1.3	1.9
Recall	Anxious	Youths	13.7	3.6	34.6	5.6	36.2	18.7	14.0	8.0
		Adults	-	-	42.3	6.1	-	-	6.9	4.0
	Healthy	Youths	-	-	28.5	4.0	8.5	5.0	2.3	2.9
		Adults	-	-	25.4	5.4	-	-	1.0	1.3

[a] PARS=Pediatric Anxiety Rating Scale (12). STAI = Spielberger State-Trait Inventory (12-14); SCARED = Screen for Child Anxiety Related Emotional Disorders (15); BDI = Beck Depression Inventory (16); CDI = Child Depression Inventory (17, 18).
[b] Significant diagnosis group difference, $p < .05$.

Data Supplement for Britton et al. (doi: 10.1176/appi.ajp.2013.12050651)