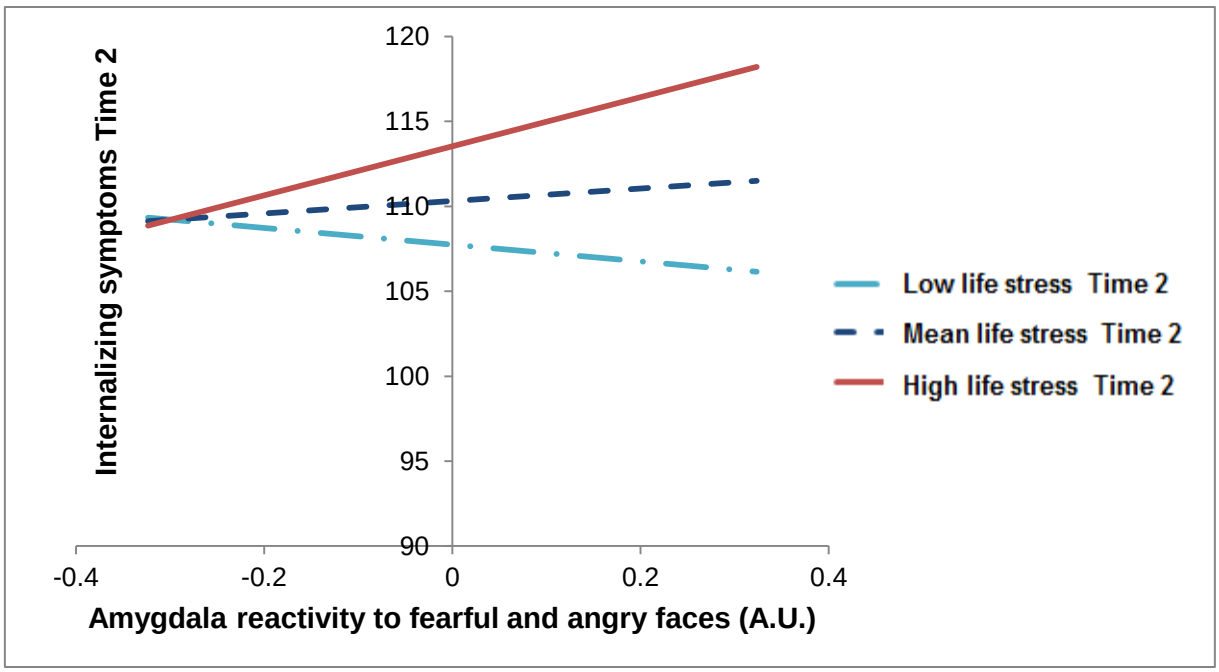
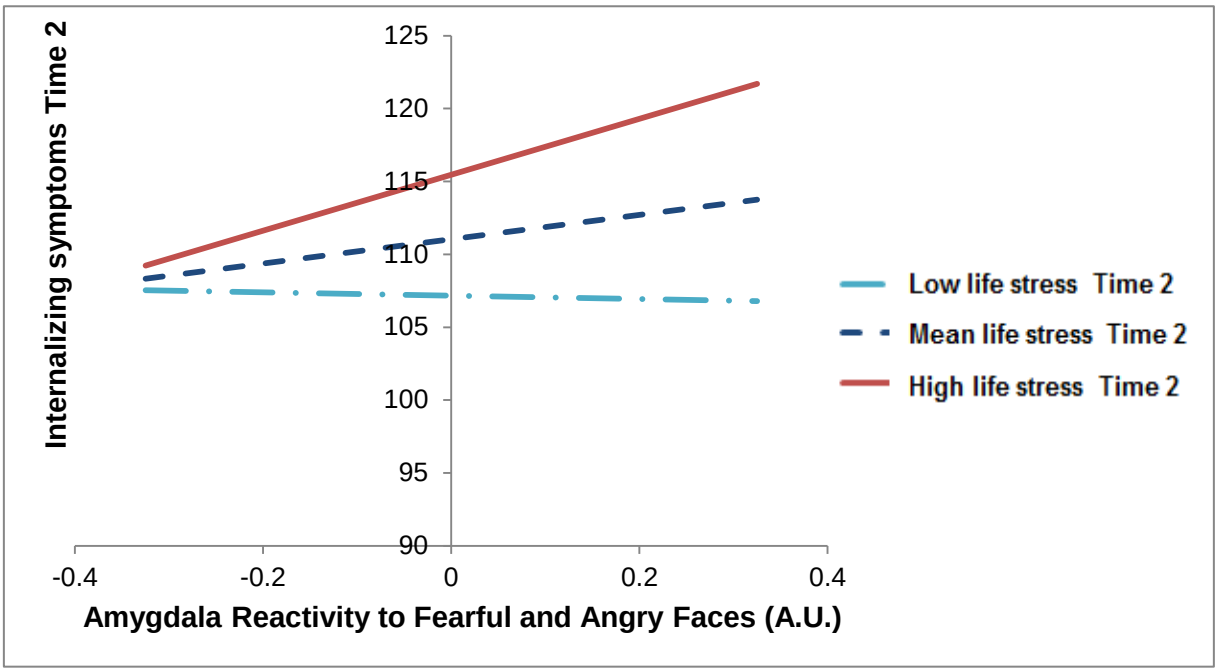
Supplemental Data

Figure S1, Related to Figure 1. Predicted values for Time 2 symptoms based on regression Model A.



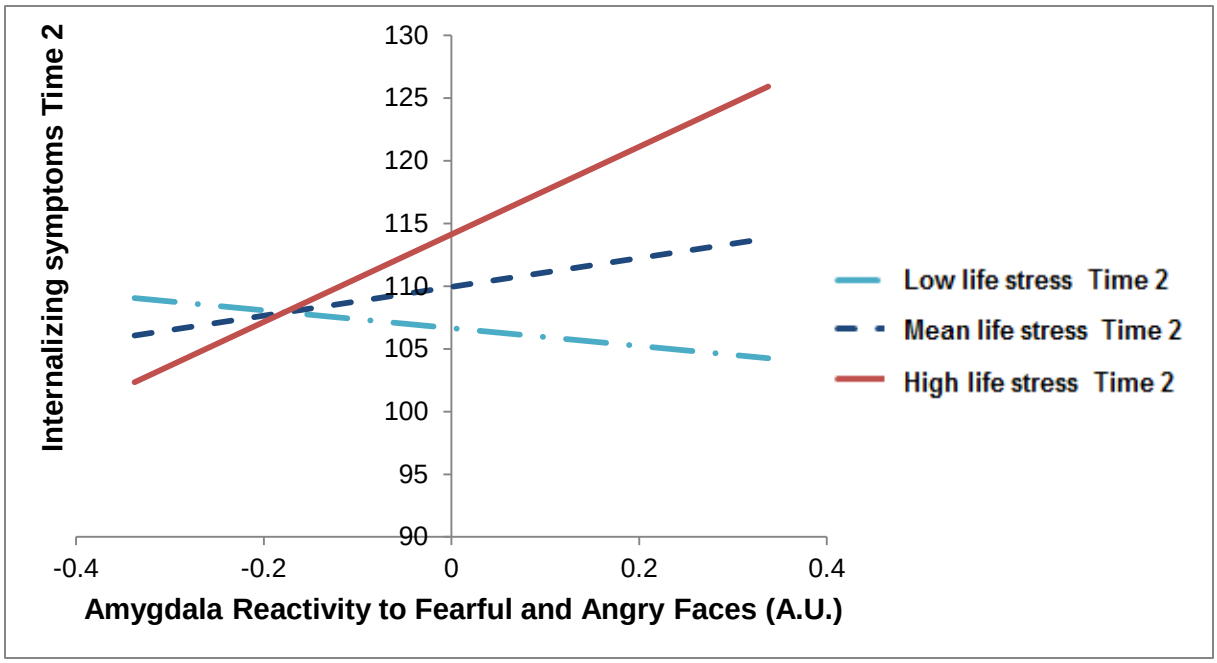
Using the PROCESS macro for SPSS (Hayes, 2013), expected values of depression and anxiety symptoms at Time 2 were estimated from the regression equation for Model A as a function of threat-related amygdala reactivity and life stress experienced post-scanning (mean +/-1 SD), including the covariates of age, gender, childhood trauma, baseline symptoms, presence of non-internalizing psychopathology, stressful life events at baseline, and number of days since imaging.

Figure S2, Related to Figure 2. Predicted values for Time 2 symptoms based on regression Model B.



Expected values of internalizing symptoms estimated as a function of amygdala reactivity and life stress experienced post-scanning (mean +/-1 SD), including covariates.

Figure S3, Related to Figure 3. Predicted values for Time 2 symptoms based on regression Model C.



Expected values of internalizing symptoms estimated as a function of amygdala reactivity and life stress experienced post-scanning (mean +/-1 SD), including covariates.

Table S1, Related to Figure 1. Results of multiple regression for Model A, predicting MASQ Total scores at follow-up for all participants who completed a post-scanning assessment.

Predictor	B Coefficient	Standard Error	t(339)=	p=
Amygdala reactivity	3.67	3.7	1.00	.32
Stressful life events Time 2	**.60**	**.3**	**2.32**	**.02**
Interaction of amygdala reactivity x stressful life events Time 2	**2.01**	**.7**	**3.08**	**.002**
Gender	-3.02	2.3	-1.31	.19
Age	1.21	1.0	1.18	.24
Childhood Trauma	.29	.2	1.72	.09
Internalizing symptoms Time 1	**.55**	**.1**	**8.21**	**<.001**
Psychopathology	2.75	3.5	.79	.43
Stressful life events Time 1	-.05	.2	-.31	.76
Days between MRI and completing questionnaire	-.0004	.004	-.12	.91

Bolded effects are significant. The full model is significant: $\chi^2(10)=175.34$, $p<.001$, $R^2=.40$. Change R^2 due to addition of interaction=.02.

Table S2, Related to Figure 1. Results of whole-brain exploratory analysis for Model A.

Region	Cluster size	$t(329)=$	$p_{uncorrected}=$	Coordinates (x,y,z)
Cluster 1:				
Right calcarine sulcus	2765	5.38	<.001	(24, -74, 10)
Left cerebellum		4.89	<.001	(-18, -46, -16)
Right lingual gyrus		4.82	<.001	(24, -62, -4)
Cluster 2:				
Vermis	197	4.32	<.001	(2, -62, -40)
Cluster 3:				
Cingulate gyrus	157	3.77	<.001	(-22, -32, 38)
Cluster 4:				
Left putamen	183	3.75	<.001	(-20, 14, -8)
Left amygdala		3.07	.001	(-26, -2, -24)
Cluster 5:				
Right thalamus	920	3.50	<.001	(16, -20, 12)
Brain stem		3.40	<.001	(-4, -20, -16)
Right caudate		3.36	<.001	(12, 0, 18)
Cluster 6:				
Left hippocampus	127	3.35	<.001	(-28, -36, -4)
Left inferior temporal cortex		3.15	.001	(-36, -38, -14)
Cluster 7:				
Left middle frontal gyrus	144	3.29	.001	(-24, -2, 38)
Cingulate gyrus		3.19	.001	(-24, -14, 38)

A whole-brain exploratory analysis was run in SPM8 by entering internalizing symptoms at Time 2, stressful life events at Time 2, and their interaction as predictors of activation to angry and fearful faces > shapes in a regression analysis. Internalizing symptoms at Time 2 and stress at Time 2 were centered before calculating the interaction term. The following variables were entered as covariates: stress at Time 1, internalizing symptoms at Time 1, gender, age, childhood trauma, psychopathology, and days since imaging. Correction for multiple comparisons was conducted by using cluster-size thresholding based on Monte Carlo simulation. In brief, this approach creates multiple simulated null datasets from which a distribution of cluster sizes

corresponding to a desired corrected p-value can be determined (using AFNI's 3dClustStim). An initial (uncorrected) statistical threshold of $p<.005$ was chosen. Based on this threshold, the number of comparisons in our imaging volume and the smoothness of our imaging data (as measured by 3dFWHMx), a minimum cluster size of 111 voxels was required to have a corrected $p \leq 0.05$. Coordinates resulting from this analysis were labeled using the Automated Anatomical Labeling (AAL) atlas.

Table S3, Related to Figure 2 and Figure 3. Results of multiple regression for Model B, predicting MASQ Total scores at follow-up for all participants who completed an assessment at least 1 year post-scanning, and for Model C, predicting MASQ Total scores at 1-year post-scanning using the mean life stress score from all assessments completed before then.

Model B				
Predictor	**B Coefficient**	**Standard Error**	**t(191)=**	**p=**
Amygdala reactivity	8.34	5.2	1.61	.11
Stressful life events Time 2	**.71**	**.3**	**2.14**	**.03**
Interaction of amygdala reactivity x stressful life events Time 2	**1.75**	**.8**	**2.33**	**.02**
Gender	-3.03	3.2	-.96	.34
Age	.94	1.4	.65	.52
Childhood Trauma	.46	.3	1.73	.08
Internalizing symptoms Time 1	**.51**	**.1**	**6.19**	**<.001**
Psychopathology	-.80	3.9	-.21	.84
Stressful life events T1	-.13	.2	-.62	.53
Days between MRI and completing questionnaire	.002	.01	.30	.77
Model C				
Predictor	**B Coefficient**	**Standard Error**	**t(98)=**	**p=**
Amygdala reactivity	11.49	6.1	1.87	.06
Stressful life events Time 2	.99	.6	1.54	.12
Interaction of amygdala reactivity x stressful life events Time 2	**5.56**	**1.9**	**2.96**	**.003**
Gender	-5.70	4.9	-1.16	.25
Age	.11	2.3	.05	.96
Childhood Trauma	.20	.5	.42	.67
Internalizing symptoms Time 1	**.61**	**.1**	**4.98**	**<.001**
Psychopathology	-7.82	7.2	-1.08	.28
Stressful life events T1	-.26	.4	-.75	.46
Days between MRI and completing questionnaire	-.003	.1	-.05	.96

Number of stress assessments	1.28	2.2	.58	.56

Bolded effects are significant. For Model B, the full model is significant: $\chi^2(10)=112.03$, $p<.001$, $R^2=.44$. Change R^2 due to addition of interaction=.01. For Model C, the full model is significant: $\chi^2(11)=59.58$, $p<.001$, $R^2=.41$. Change R^2 due to addition of interaction=.05.

Supplemental Experimental Procedures

Functional MRI acquisition

Each participant was scanned using a research-dedicated GE MR750 3T scanner equipped with high-power high-duty-cycle 50-mT/m gradients at 200 T/m/s slew rate, and an eight-channel head coil for parallel imaging at high bandwidth up to 1 MHz at the Duke-UNC Brain Imaging and Analysis Center. A semi-automated high-order shimming program was used to ensure global field homogeneity. A series of 34 interleaved axial functional slices aligned with the anterior commissure-posterior commissure plane were acquired for full-brain coverage using an inverse-spiral pulse sequence to reduce susceptibility artifact (TR/TE/flip angle = 2000 ms/30 ms/60; field of view (FOV) = 240 mm; 3.75 × 3.75 × 4 mm voxels; interslice skip = 0). Four initial radiofrequency excitations were performed (and discarded) to achieve steady-state equilibrium. To allow for spatial registration of each participant's data to a standard coordinate system, high-resolution three-dimensional structural images were acquired in 34 axial slices co-planar with the functional scans (TR/TE/flip angle = 7.7 s/3.0 ms/12; voxel size = 0.9 × 0.9 × 4 mm; FOV = 240 mm, interslice skip = 0).

Functional MRI analysis

Preprocessing of fMRI data was conducted in SPM8, including realigning images to the first volume in the time series to correct for head motion, spatially normalizing images into a standard stereotactic space (Montreal Neurological Institute (MNI) template) using a 12-parameter affine model (final resolution of functional images = 2 mm isotropic voxels), and smoothing to minimize noise and residual difference in gyral anatomy with a Gaussian filter, set at 6-mm full-width at half-maximum. Voxel-wise signal intensities were ratio normalized to the whole-brain

global mean.

Artifact detection software (http://www.nitrc.org/projects/artifact_detect) was used to create a regressor assigning a lower weight to: 1) volumes exhibiting significant mean-volume signal intensity variation (i.e., within volume mean signal greater or less than 4 standard deviations of mean signal of all volumes in time series), and 2) individual volumes where scan-to-scan movement exceeded 2 mm translation or 2° rotation in any direction. Quality control criteria for inclusion of a participant's imaging data were: <5% volumes exceed Artifact detection criteria for motion or signal intensity outliers, ≥ 90% coverage of signal within the anatomically-defined bilateral amygdala region of interest, and accuracy ≥75% on the matching task performed during scanning. A total of 906 participants underwent fMRI scanning; of these, 15 were excluded for coverage, 27 were excluded for exceeding the Art criteria, 35 were excluded for accuracy, and 18 were excluded for other problems associated with data quality including incidental findings, problems with the coil, or artifacts in the images. Thus, a total of 811 participants met all quality control criteria for fMRI data. Mean behavioral accuracy for included participants was 98.4% (SD=.04%).

Covariates

Several additional variables were included as covariates in the regression models, including age and gender (0=male, 1=female). Diagnosis of a current or past DSM-IV Axis I or select Axis II disorder was included as a covariate (0=no psychopathology, 1=psychopathology). Total scores on the Childhood Trauma Questionnaire (Bernstein et al., 2003) were included to control for early life trauma.

Attrition and supplementary analyses

We examined whether participants who completed an online assessment following scanning differed from participants who did not complete a follow-up assessment in terms of age, childhood trauma, stressful life events reported at baseline, internalizing symptoms reported at baseline, or amygdala reactivity recorded at baseline. For Models A and B, there was no significant difference between participants who completed a follow-up questionnaire compared to excluded participants on any of these measures. For Model C, included participants (M=31.4, SD=6.4) had lower scores than excluded participants (M=33.5, SD=8.1) on the Childhood Trauma Questionnaire (used as a covariate), $t(751)=2.41$, $p=.02$. They did not differ on any of the variables of interest assessed at baseline.

Because participants differed in terms of the number of assessments completed and how far apart from scanning assessments were completed, we examined whether these factors were associated with our variables of interest (life stress or depression at any time point). For Model A, the number of days between scanning and the most recent assessment was associated with life stress reported at Time 2, $r=.22$, $p<.001$, indicating that participants completing an online assessment farther apart from scanning reported greater severity of life events. For this reason, the number of days between imaging and the online assessment was entered as a covariate in all analyses. For Models B and C, there was no association between days since imaging and the measures of interest. For Model C, the number of assessments contributing to the mean life stress score (M=2.4, Min-Max=1-4) was negatively associated with depressive symptoms at Time 1, $r=-.25$, $p=.01$, indicating that participants with greater internalizing symptoms at Time 1 completed a smaller number of assessments post-scanning. Thus, we controlled for the number of assessments completed when testing Model C.

To examine whether amygdala reactivity at baseline was associated with the severity of life events reported post-scanning, we tested the association between amygdala reactivity and the life stress measure at Time 2. For Model A (r=.11, p=.04) and for Model B (r=.17, p=.02) the association was significant, suggesting that individuals with greater amygdala reactivity at baseline reported greater life stress at follow-up. However, for Model C, this correlation was not significant, r=-.004, p=.97, indicating this effect was not present when taking the mean of multiple reports of life stress completed post-scanning.

Additional analyses to test specificity of effects

To determine whether reported effects were specific to amygdala reactivity to threat rather than perceptual features common to all face stimuli (e.g., perceptual complexity), we tested results for all models examined using mean amygdala reactivity to neutral faces > shapes rather than amygdala reactivity to angry and fearful faces > shapes. The interaction of amygdala reactivity and life stress was significant for Model A (p=.048), but not for Model B, (p=.22), or Model C (p=.32), indicating that effects were generally specific to amygdala reactivity to threat rather than all facial expressions.

We also performed post hoc analyses to examine the potential laterality of effects, given that we computed a mean amygdala reactivity value to reduce the number of comparisons performed. We tested laterality within the most stringent model, Model C. Here, both left amygdala reactivity to threatening faces (p=.02) and right amygdala reactivity to threatening faces (p=.002) interacted with life stress to predict internalizing symptoms 1 year later. In addition, we tested whether effects were specific to the basolateral or centromedial amygdala by extracting contrast values (for angry and fearful faces>shapes) from functional clusters defined

by ROIs based on probabilistic maps for these subregions (Amunts et al., 2005). For Model A, the interaction between amygdala reactivity and life stress was significant for the centromedial amygdala (p=.002) and the basolateral amygdala (p=.024). For Model B, the interaction was significant for the centromedial amygdala (p=.028) and for the basolateral amygdala (p=.048). For Model C, the interaction was significant for the centromedial amygdala (p=.01) and for the basolateral amygdala (p=.006).

Finally, for Model C, we further examined whether effects were specific to anxiety or depressive symptoms by using subscales of the internalizing measure as the outcome. The interaction remained significant for general distress/anxiety symptoms (p=.013) and anhedonia (p=.003), but was not significant for general distress/depressive symptoms (p=.11) or anxious arousal (p=.19).

Supplemental References

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N.J., Habel, U., Schneider, F., and Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. Anat Embryol *210*, 343-352.

Bernstein, D.P., Stein, J.A., Newcomb, M.D., Walker, E., Pogge, D., Ahluvia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., *et al.* (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. Child Abuse Negl. *27*, 169-190.

Hayes, A.F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (New York: Guilford Press).