Amygdala Reward Reactivity Mediates the Association Between Preschool Stress Response and Depression Severity

Supplemental Information

Diagnostic Assessment

Diagnostic assessments were conducted using the Kiddie Schedule for Affective Disorders-Early Childhood version (K-SADS-EC; (1)), a developmentally modified version of the Kiddie Schedule for Affective Disorders and Schizophrenia for School age Children-Present and Lifetime Version (K-SADS-PL) (2) adapted for use in preschool age children. Revisions include adjusting and/or adding questions and threshold anchors reflecting developmentally appropriate symptom and diagnostic criteria for 3-6 year old children, removing sections less applicable to young children, editing probes to use parent-directed wording, and providing directions for clarifying symptom severity and genesis where necessary. Previous research has demonstrated diagnostic consistency between the K-SADS-PL and the Preschool Age Psychiatric Assessment (3), a valid and commonly used semi-structured psychiatric interview for parents of young children. Following completion of the K-SADS-EC by a trained research assistant, relevant symptom and duration criteria gathered during the interview were individually reviewed for each child by a licensed clinical psychologist with expertise in early childhood development and psychopathology (author MSG) and subsequently used to generate DSM-5 (4) diagnoses when appropriate.

Depression Severity

The Preschool Feelings Checklist – Scale Version (PFC-S; (5) is a 23 item measure that uses a Likert rating scale (0 = never, 4 = most of time; range of possible scores 0-92) designed to assess depression severity in preschool children and has established validity at this age (6). To further

support the validity of the Preschool Feelings Checklist – Scale version (PFC-S) as a measure of depression severity and to demonstrate its convergence with a widely available clinical questionnaire commonly used to examine emotion difficulties in preschoolers, we examined the relationship between scores from the PFC-S and the Child Behavior Checklist (CBCL) (7) subscale scores for Depression/Anxiety and Affective Problems (reflecting many of the symptoms of depression including sadness, irritability, diminished positive affect, sleep disturbances, altered eating, and fatigue in the current sample). CBCL scores were available for 50 of the 52 preschoolers included in the current study. As anticipated, the PFC-S and both CBCL scores were highly positively correlated (PFC-S and CBCL Depression/Anxiety raw score: $r = .55$, $p < .001$; PFC-S and CBCL depression/anxiety t-score: $r = .53$, $p < .001$; PFC-S and CBCL Affective Problems raw score: $r = .73$, $p < .001$; PFC-S and CBCL Affective Problems t-score: $r = .72$, $p < .001$), providing further support of the PFC-S as a valid measure of depression severity and its use as such in the current study."

Cortisol Assay and Quality Control Methods

Saliva samples (n=395) were collected using Salivette collection devices (Sarstedt, Rommelsforf, Germany) then stored at -80°C immediately following collection. Salivary cortisol was measured using enzyme-linked immunosorbent assay (ELISA; DRG International kit SLV-2930 lot #s 64K035 and 64K105; Springfield, New Jersey USA). All samples were run in duplicate and the average value was used for analyses. Intra- and inter-assay coefficients of variation (CV) were 3.99% and 6.85% respectively.

Prior to use, kits, reagents, and samples were brought to room temperature. Samples were centrifuged at 3,000 x g for 10 minutes. Next, 120 µL of sample, standard (0.0, 2.0, 5.0, 10.0, 20.0,

40.0, 80.0 ng/mL), or high and low cortisol control samples (to allow for interplate comparison) were aliquoted to a clean 96 well plate. Then, 100 µL from each well was transferred to the 96 well ELISA plate pre-coated with mouse anti-cortisol antiserum. Horseradish peroxidase-conjugated cortisol (200 µL) was added to each well on the ELISA plate and incubated on a mixer for 60 minutes. After emptying well contents, plates were washed 3 times with wash solution (400 µL/well) using an ELx50 plate washer (BioTek; Winooski, Vermont, USA). Residual wash solution was removed before 200 µL of tetramethylbenzidine (TMB) substrate solution was added to each well. The plate was then incubated on a mixer for 30 minutes. The reaction was stopped by adding 400 µL of 0.5M H_2SO_4 stop solution and then read at 450 nm using an Epoch microplate spectrophotometer (BioTek; Winooski, Vermont, USA) and calculated using Gen5 software (BioTek; Winooski, Vermont, USA). Cortisol concentrations (ng/mL) were calculated from the optical densities by the Gen5 software using 4-parameter logistic regression.

Fifty-two of the 60 children with usable fMRI data also provided usable cortisol data from 5 or more samples during the stress reactivity task. Of the 8 children without usable cortisol data, 5 were unable to produce saliva sufficient for measurement and 3 took 30 minutes or longer to produce their baseline cortisol sample. Of the 52 children with cortisol data, 1 child provided 5 samples, 6 children provided 6 samples, and 45 children provided 7 samples. Children with missing samples were required to have one of the first two samples (prior to frustration task and immediately after the frustration task [2 were missing the first sample and 4 were missing the second sample]) and the last sample. Number of saliva samples was not related to AUC_g ($r = .01$, $p = .9$) and the mediation models reported in the main manuscript remained significant when number of cortisol samples was included as a covariate (PROCESS Indirect Effect [10,000 bootstrap samples]: .19 (.11), bias corrected 95% CI: .04/.5) and when children with fewer than 7

cortisol samples are excluded (PROCESS Indirect Effect [10,000 bootstrap samples]: .17 (.12), bias corrected 95% CI: .03/.5).

Stress Evoked Cortisol Response in Young Children

While cortisol scores increased at the group level between cortisol measures 1 (baseline mean = .5[.07]) and 6 (mean peak at 40 minutes post stressor = .52 [.1]) following our stressor paradigm, this increase did not reach statistical significance ($t = -.88$, $p = .39$). Previous research examining stress evoked cortisol response in preschool age children has reported mean level cortisol increases, decreases, as well as significant variability at the individual level (e.g., some children exhibiting increases in cortisol and others exhibiting decreases in cortisol despite mean level increases or decreases at the group level). As an illustrative example, Tolep & Dougherty (8) recently reported an overall mean decrease in preschooler cortisol reactivity following exposure to the same in-lab stressor paradigm used in the current study. However, when cortisol increases were examined at the individual child level (i.e., peak value that was used to calculate increase was child specific), the authors found a stress-invoked cortisol response in 51.3% of their preschoolers. Similarly, despite reporting a mean level increase at the group level, De Weerth et al. (9) found that a significant minority of their preschoolers (39%) did not exhibit an increase in cortisol following an age appropriate, in-lab stressor paradigm. In line with the current study's focus on individual differences, this research has also suggested that variability stress-evoked cortisol response following a stressor may be related to inter-individual differences in emotion regulation at this early age (e.g., high negative affect, low positive affect, etc) (8, 10, 11). Importantly, similar research in adults has suggested that individual variability in stress evoked cortisol response may be meaningfully related to brain activity. More specifically, attenuated

amygdala reactivity has been associated with a heightened stress evoked cortisol response following a psychosocial stressor in adults (12). Given the previous research reviewed above, it's likely that variability in stress evoked cortisol response in preschoolers also reflects meaningful individual differences in stress reactivity and response that can be uniquely informative when interpreted within a broader context (i.e., not group mean(s) alone), including measures of brain function and emotion like those used in the current study.

Specificity of Findings to AUC_g

The current study included a measure of the overall stress evoked cortisol response (i.e., the amount of cortisol produced following the in-lab stressor) referred to as Area Under the Curve with respect to ground (AUC_g). This choice was based on previous research indicating that total cortisol output, rather than reactivity, following a stressor is associated with depression and depression risk (13, 14). Nevertheless, to provide additional support for the specificity of the reported findings to AUC_g, we re-ran our primary mediation analysis using AUC_i, a measure of cortisol change from baseline following a stressor (15). When using AUCi, the mediation model including AUC_i as the independent variable, amygdala reactivity as the mediator, and depression severity as the outcome was not significant (PROCESS Indirect Effect [10,000 bootstrap samples]: -.06 (.21), bias corrected 95% CI: -.51/.31).

Specificity of Findings to Depression Severity

To further bolster the specificity of our mediation findings to depression severity, we replaced PFC-S scores with raw scores from the CBCL (7) DSM-Oriented Anxiety and ODD subscales in our mediation model. CBCL scores were available for 50 of the 52 preschoolers included in the

current study. Mediation models including CBCL Anxiety and ODD raw scores as the outcome measure (rather than PFC-S scores) did not reach significance. In addition, providing additional support for the specificity of our findings to depression severity, amygdala reactivity continued to mediate the relationship between cortisol response (AUC_g) and PFC-S scores when CBCL Anxiety and ODD subscale raw scores were included as covariates in the model (PROCESS Indirect Effect [10,000 bootstrap samples]: .18 (.11), bias corrected 95% CI: .04/.49). The CBCL Affective Problems subscale primarily includes items measuring symptoms of depression, such as sadness, irritability, diminished positive affect, sleep disturbances, altered eating, and fatigue. Given the strong positive correlation ($r = .727$, $p < .001$) between PFC-S scores and CBCL Affective Problems subscale raw scores, we replaced PFC-S scores with CBCL Affective Problems subscale raw scores in our mediation model to investigate whether our findings would replicate with this closely aligned measure. The mediation model including CBCL Affective Problems in place of PFC-S scores was significant (PROCESS Indirect Effect [10,000 bootstrap samples]: .05 (.03), bias corrected 95% CI: .0023/.1327), though with a reduced effect size. Collectively, these findings further support the specificity of our findings to depressive symptoms.

Results of Mediation Model When Excluding Two Children Without Reaction Time Data, When Not Controlling for Maternal Depression, and When Using Right Amygdala Reactivity to Outcome

In order to further establish the robustness of the mediation model supporting diminished amygdala reactivity to reward outcome as a mediator of the relationship between stress evoked cortisol response and depression, we reran our mediation analyses excluding the two children who did not make a button press during the CGT and when not controlling for maternal depression. The

reported results stay the same when maternal depression scores are not included as a covariate, including a significant positive relationship between AUC_g and PFC-S scores (r = .244, p = .04), significant negative relationship between left amygdala reactivity and PFC-S scores (r = -.36, p = .004), and a significant mediation effect of amygdala reactivity (PROCESS Indirect Effect [10,000 bootstrap samples]: .24 (.15), bias corrected 95% CI: .04/.6).

Our mediation analyses also remained significant when the two children who did not make a button press during the CGT were excluded, (PROCESS Indirect Effect [10,000 bootstrap samples]: .21 (.12), bias corrected 95% CI: .04/.55).

Our primary analyses supported left amygdala reactivity to highly salient reward outcomes as an important mediator of the relationship between stress evoked cortisol response and depression severity in preschoolers. To further examine the specificity of our results to amygdala reactivity to highly salient reward outcomes versus amygdala responsivity in general, we carried out an additional mediation model using reactivity scores from the right amygdala region of interest, which exhibited a main effect of time (i.e., similar reactivity to all 3 outcomes) but not an outcome-by-time interaction. The mediation model using the average reactivity score across all outcomes relative to baseline (i.e., the percent signal change score representing a main effect of time) from the right amygdala as the mediator was not significant (PROCESS Indirect Effect [10,000 bootstrap samples]: -.0038 (.06), bias corrected 95% CI: -.14/.1), providing additional support for amygdala reactivity to highly salient outcomes as an important mediator of stress and depression in preschool age children."

Alternative Mediation Model

A growing body of research suggests that early alterations in the developing stress system affect

neural response to reward and increase risk for depression as a result (16). We investigated a mediational model based on this research including stress evoked cortisol response as an independent variable (measured via AUC_g), amygdala reactivity to reward as a mediator variable, depression severity as an outcome variable, and relevant covariates (maternal depression, age in months, gender, CBCL anxiety raw score, CBCL, ODD raw score; please see parent paper and Specificity of Findings to Depression Severity subsection in the supplemental material). As predicted and in line with previous developmental theories and data, the tested mediation model was significant in our preschool aged sample, indicating that attenuated amygdala response to reward mediated the relationship between heightened stress evoked cortisol response and increased depression severity at this very early age (please see results in parent paper). However, given that our measures of stress evoked cortisol response, depression severity, and brain function were measured concurrently, the current data cannot inform directions of causality. As a result, it is possible that alternative mediational relationships between the included variables may be present and fit the current data. In line with this, we tested an additional mediation model based on the growing body of research suggesting a complex and reciprocal relationship between stress and depression (i.e., stress generation (17)). Following this line of thinking we tested two alternative mediation models including one placing stress evoked cortisol response as the independent variable, child depression severity as the mediating variable, and amygdala reactivity to reward as the outcome variable (model 1) and another placing depression as the independent variable, stress evoked cortisol response as the mediator variable, and amygdala reactivity to reward as the outcome variable (model 2). Both mediation models were not significant (Model 1: PROCESS Indirect Effect [10,000 bootstrap samples]: -0.004 (.004), bias corrected 95% CI: -0.02/.0007; Model 2: PROCESS Indirect Effect [10,000 bootstrap samples]: -0.003 (.003), bias corrected 95%

CI: -0.01/.0009).

fMRI Data Acquisition Parameters and Preprocessing Methods

Image acquisition included an initial low-resolution 3D sagittal T1-weighted MP-RAGE rapidly warped to Talairach space (18). This image was then used to provide on-line slice localization for the functional images, placing them as close as possible to the target template. T1 images (TR = 2,400 ms, TE = 3.16 ms, flip angle = 8°, slab = 176 mm, 176 slices, matrix size = 256 x 256, field of view (FOV) = 256 mm, voxel size = 1 x 1 x 1 mm, sagittal plan acquisition) were acquired as part of the structural imaging protocol and used in the transformation of images to a common template space optimized for preschool children (18). The accuracy and validity of this transformation for preschool age children has been demonstrated in previous research (19) and was confirmed through visual inspection for distortions and the accuracy of alignment for key cortical and subcortical landmarks. The functional images were collected with a 12-channel head coil using an asymmetric spin-echo echo-planar sequence sensitive to BOLD contrast (T2*) (TR=2000ms, TE=27ms, FOV=384mm, flip=90°). During each functional run, sets of 32 contiguous axial images with isotropic voxels ($4mm^3$) were acquired parallel to the anterior-posterior commissure plane.

Prior to preprocessing, the first 4 frames of each run were discarded to allow for signal stabilization. The fMRI data were preprocessed and analyzed using in-house Washington University software. Data were reconstructed into images and normalized across runs by scaling whole-brain signal intensity to a fixed value and removing the linear slope on a voxel-by-voxel basis to counteract effects of drift (20). Data was also corrected for head motion using rigid-body rotation and translation correction algorithms (21-23), co-registered to Talairach space using a 12

parameter linear (affine) transformation that included resampling to 3mm cubic, and smoothed using a 6mm FWHM Gaussian filter. Within scan head movement was assessed using output from the rigid-body rotation and translation algorithm. After measuring the translations and rotations in the x, y, and z planes across frames, total root mean square (RMS) linear and angular measures were calculated and used to obtain the average amount of movement in millimeters per frame (i.e., 1 TR) in a given run for each subject (RMS/frame). CGT runs with greater than 1.5mm RMS/frame on average were excluded from further data analysis. Using this criterion, 64 children provided usable CGT data from only 1 of the 2 possible runs. To further reduce any potential effects of head movement on data quality, custom MATLAB (The Mathworks, Natwick, MA) code was used to identify frames with greater than 1mm absolute movement. The identified frames were removed from further data analysis (average percentage of frames removed = 11%). Four children with fewer than 65% of frames remaining after frame-by-frame censoring were not included in subsequent data analyses. Not surprisingly, children who did not provide usable MRI data (n=28; NMRI) were significantly younger on average (mean age 60 [11.5] months) than those who did (n=60; mean age 71 [9] months; MRI). However, groups did not differ in sex (NRMI 16 vs. MRI 33 females; $x^2 = .04$, $p = .85$), maternal BDI-II score (NMRI 7.3[7.8] vs. MRI 8.2[9.4]; $t_{86} = .4$, $p = .69$), or PFC-Scale score (NMRI 18.3[10.7] vs. MRI 15.8[10.5]; $t_{86} = -1.1$, $p = .3$).

Analyses using our a priori reward processing mask and at the whole brain were corrected for multiple comparisons using recommended guidelines (24) addressing recently identified challenges with inflated false positive rates in fMRI studies (25), the Analysis of Functional Neuro-Images (AFNI) 3dFWHMx and 3dClustSim commands were used to determine the combined p-value/cluster size thresholds required to maintain a false positive rate of $p<.05$. Thresholds were $z = 3$ ($p<.001$) and 10 voxels within our reward processing mask (correcting for all ROIs

simultaneously; false positive rate of p<.05 for the whole ROI mask) and z = 3 (p<.001) and 26 voxels for whole-brain analyses (whole-brain false positive rate of p<.05). Large clusters spanning multiple regions identified within our a priori reward processing mask were subsequently partitioned such that peaks of activity were considered separate regions if they were more than 12 mm apart, as measured by a peak-splitting algorithm. Individual time courses within the identified regions were then extracted for subsequent analyses using 6mm spheres centered at the peak voxel coordinates.

fMRI Task

Prior to the child gambling task (CGT), children are asked to choose which of two different candies (M&Ms or Skittles) they would like to play for during the CGT. Once they have chosen their candy, they are introduced to how the CGT is played during their in-person assessment using child friendly language to describe how it is played (e.g., "Your job is to guess if the person hiding behind the question mark is bigger or smaller than you!'), by providing examples of the CGT images and testing understanding of them during instruction ("This is an adult. Are they bigger or smaller than you?"), and providing children a sequence of practice trials (fixation cross, question mark, outcome …) where they play the CGT and demonstrate their understanding of it (on average ~5-10 trials). Each child practices until they can demonstrate they know when to guess and how to make a guess (e.g., press the button for bigger don't press it for smaller). In addition, prior to playing the CGT in the scanner each child is reminded of how the game is played and asked to demonstrate a verbal understanding of how to respond (e.g., Staff: "What do you do when you want to guess bigger?", Child: "Press the button.").

fMRI Whole Brain Results

Whole brain results were significant for a main effect of time in multiple cortical and subcortical regions (see Figure S1 and Table S1). Follow-up analyses found outcome x time effects in parahippocampal gyrus, fusiform gyrus and postcentral gyrus (see Table S1). A large cluster within the right fusiform gyrus was subsequently partitioned such that peaks of activity were considered separate regions if they were more than 12 mm apart, as measured by a peak-splitting algorithm. Additional cortical and subcortical regions were identified following this process (see Table S1). The region matching closest to our a priori region of interest in the left amygdala spanned portions of the amygdala as well as the uncus. Similar to the results reported for the amygdala in the main text, a negative correlation between reactivity within the larger amygdala/uncus ROI (using gain/loss minus neutral difference scores created in a fashion identical to our primary amygdala ROI analyses) and depression severity was found ($r = -.38$, $p = .007$). However, while in the expected direction, the correlation between amygdala/uncus reactivity and stress evoked cortisol response (AUC_g) did not reach significance ($r = -.13$, $p = .17$), suggesting specificity of this relationship to the amygdala.

Table S1: Whole Brain Analyses: Regions identified with main effect of time and outcome x time interaction

Region	Hemisphere	BA	Peak Voxel			Cluster (voxels)	Outcome X Time
			X	Y	Z		
Cerebellum	R		8	-54	-45	60	NS
Parahippocampal Gyrus	R	28	16	-21	-27	59	GL > N
Fusiform Gyrus	L		-58	-9	-27	223	GL > N
Superior Temporal Gyrus	R	38	52	15	-27	93	NS
Rectal Gyrus	L	11	-8	27	-24	35	NS
Putamen	R		16	9	6	27	NS
Precentral Gyrus	R		38	-3	33	47	NS
Inferior Parietal Lobe	R		58	-33	45	43	NS
Post Central Gyrus	L		-34	-33	54	395	GN > L
Fusiform Gyrus*	R		28	-81	-12	18,484	GL > N
Middle Occipital Gyrus		18	26	-83	-4	116	
Cuneus		17	-16	-93	5	113	
Culmen			-37	-47	-30	108	
Declive			-30	-84	-21	110	
Cerebellar Tonsil			40	-61	-38	117	
Lingual Gyrus			-18	-48	2	116	
Culmen			9	-28	-11	116	
Posterior Cingulate		23	-1	-28	18	110	
Middle Temporal Gyrus		37	59	-44	-4	113	
Precuneus		39	41	-64	37	117	
Superior Frontal Gyrus		6	17	16	57	112	
Cuneus		19	5	-84	35	113	
Medial Frontal Gyrus		8	4	40	38	116	
Precuneus		19	-41	-68	39	110	
Precentral Gyrus		44	-42	14	6	113	
Middle Frontal Gyrus		6	-44	1	50	118	
Thalamus			-2	0	5	111	

Region	Hemisphere	BA	Peak Voxel			Cluster (voxels)	Outcome X Time
			X	Y	Z		
Middle Temporal Gyrus		21	-56	-52	0	110	
Middle Frontal Gyrus		8	41	28	38	113	
Paracentral Lobule		31	-3	-9	44	114	
Superior Frontal Gyrus		6	-20	26	54	110	
Amygdala/Uncus			31	-2	-28	107	
Superior Frontal Gyrus		10	20	61	19	110	
Precuneus		7	3	-47	60	109	
Middle Frontal Gyrus		46	-40	41	27	112	
Inferior Temporal Gyrus		20	60	-20	-23	114	
Superior Frontal Gyrus		10	-17	61	14	114	
Amygdala/Uncus			-20	-6	-33	93	
Inferior Temporal Gyrus		21	-65	-6	-18	93	
Superior Temporal Gyrus		38	50	15	-9	93	
Precentral Gyrus		4	56	-19	36	113	

* Peak splitting results using a required cluster distance of 30mm are reported underneath the original cluster.

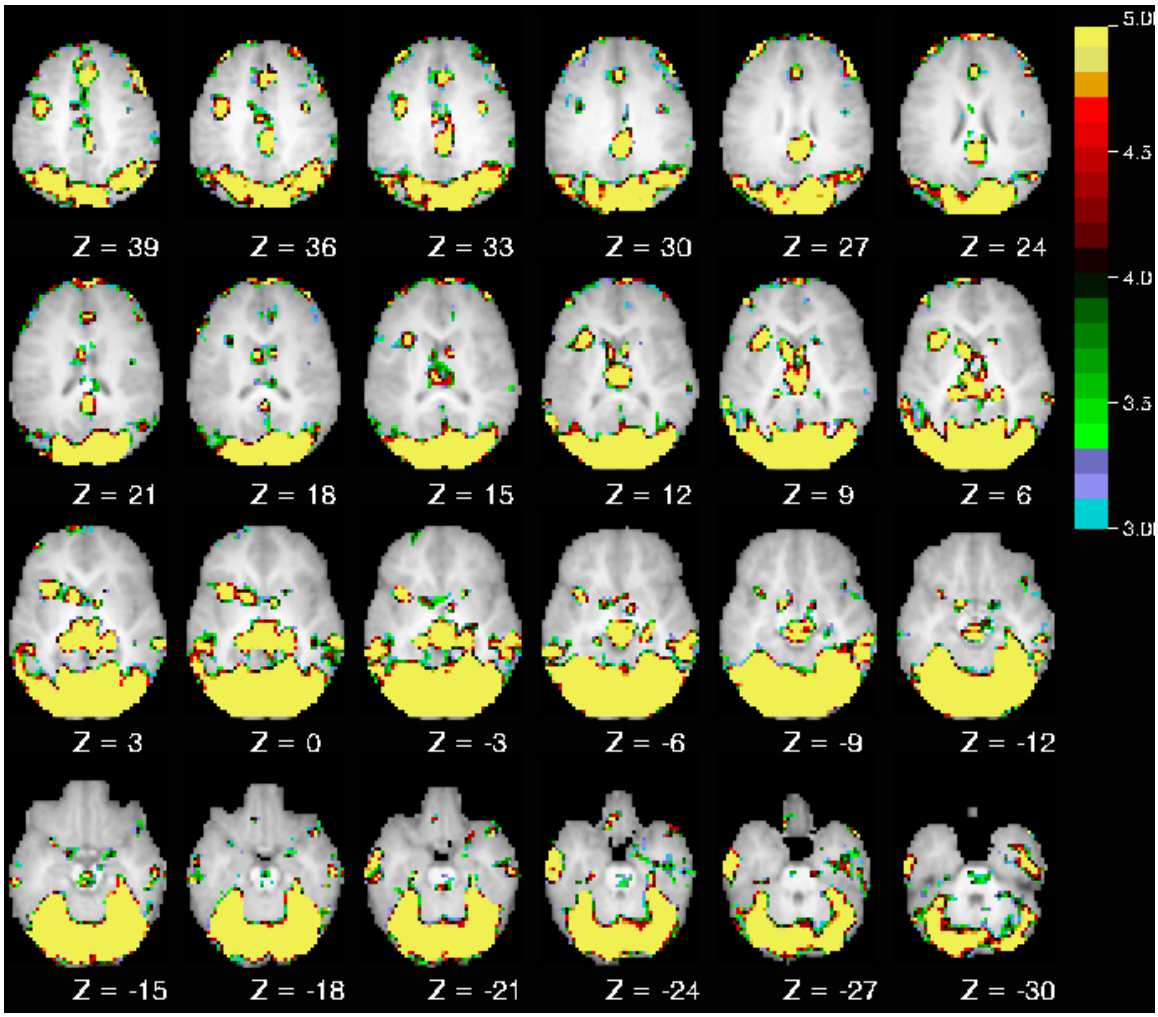

Figure S1: Regions identified with a significant main effect of time during the child gambling task. Colored bar at right represents z-value at the individual voxel level.

Supplemental References

1. Gaffrey MS, Luby JL (2012): Kiddie Schedule for Affective Disorders and Schizophrenia - Early Childhood Version, 2012 Working Draft (K-SAD-EC). St. Louis, MO: Washington University School of Medicine.
2. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. (1997): Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 36:980-988.
3. Birmaher B, Ehmann M, Axelson DA, Goldstein BI, Monk K, Kalas C, et al. (2009): Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children--a preliminary psychometric study. *J Psychiatr Res*. 43:680-686.
4. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Washington, D.C.
5. Luby J, Heffelfinger A, Mrakeotsky C, Hildebrand T (1999): Preschool Feelings Checklist. St. Louis, Missouri: Washington University.
6. Luby J, Lenze S, Tillman R (2012): A novel early intervention for preschool depression: findings from a pilot randomized controlled trial. *J Child Psychol Psychiatry*. 53:313-322.
7. Achenbach TM, Rescorla LA (2001): *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
8. Tolep MR, Dougherty LR (2014): The Conundrum of the Laboratory: Challenges of Assessing Preschool-Age Children's Salivary Cortisol Reactivity. *Journal of Psychopathology and Behavioral Assessment*. 36:350-357.
9. de Weerth C, Zijlmans MA, Mack S, Beijers R (2013): Cortisol reactions to a social evaluative paradigm in 5- and 6-year-old children. *Stress*. 16:65-72.
10. Kryski KR, Smith HJ, Sheikh HI, Singh SM, Hayden EP (2013): HPA axis reactivity in early childhood: associations with symptoms and moderation by sex. *Psychoneuroendocrinology*. 38:2327-2336.
11. Kryski KR, Dougherty LR, Dyson MW, Olino TM, Laptook RS, Klein DN, et al. (2013): Effortful control and parenting: associations with HPA axis reactivity in early childhood. *Dev Sci*. 16:531-541.
12. Pruessner JC, Dedovic K, Khalili-Mahani N, Engert V, Pruessner M, Buss C, et al. (2008): Deactivation of the limbic system during acute psychosocial stress: evidence from positron emission tomography and functional magnetic resonance imaging studies. *Biol Psychiatry*. 63:234-240.
13. LeMoult J, Ordaz SJ, Kircanski K, Singh MK, Gotlib IH (2015): Predicting first onset of depression in young girls: Interaction of diurnal cortisol and negative life events. *J Abnorm Psychol*. 124:850-859.

14. Lopez-Duran NL, Kovacs M, George CJ (2009): Hypothalamic-pituitary-adrenal axis dysregulation in depressed children and adolescents: a meta-analysis. *Psychoneuroendocrinology*. 34:1272-1283.

15. Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH (2003): Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*. 28:916-931.

16. Pechtel P, Pizzagalli DA (2011): Effects of early life stress on cognitive and affective function: an integrated review of human literature. *Psychopharmacology (Berl)*. 214:55-70.

17. Rudolph KD, Hammen C, Burge D, Lindberg N, Herzberg D, Daley SE (2000): Toward an interpersonal life-stress model of depression: the developmental context of stress generation. *Dev Psychopathol*. 12:215-234.

18. Talairach J, Tournoux P (1988): *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging.* Stuttgart: Thieme Medical Publishers.

19. Ghosh SS, Kakunoori S, Augustinack J, Nieto-Castanon A, Kovelman I, Gaab N, et al. (2010): Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age. *Neuroimage*. 53:85-93.

20. Bandettini PA, Jesmanowiz J, Wong EC, Hyde JS (1993): Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*. 30:161-173.

21. Friston KJ, Jezzard P, Turner R (1994): The analysis of functional MRI time series. *Human Brain Mapping*. 1:153-171.

22. Snyder AZ (1996): Difference image versus ratio image error function forms in PET-PET realignment. In: R. Myer VJC, D. L. Bailey, & T. Jones, editor. *Quantification of Brain Function Using PET*. San Diego: Academic Press, pp 131-137.

23. Woods RP, Cherry SR, Mazziotta JC (1992): Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*. 16:620-633.

24. Cox RW, Reynolds RC, Taylor PA (2016): AFNI and Clustering: False Positive Rates Redux. *bioRxiv*.

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