Supporting Information

McLaughlin et al. 10.1073/pnas.1423363112

SI Materials and Methods

Laboratory-Based Tasks. The peer evaluation task was adapted from a previously validated task used with children as young as age 4 y (1). The task began on a study visit before the physiological reactivity session. Participants were told that they would be playing a game to learn how children choose friends. They were then presented with 30 photographs of children along with brief profiles of each child including three pieces of information: favorite sport, favorite food, and favorite music/band/singer. Participants were told that they would have an opportunity to meet one of the other children in a subsequent visit and were asked to select the 10 children that they were most interested in meeting. Next, participants had their picture taken and provided information on their favorite sport, food, and music/band/singer for their profile. They were told that the other children participating in the study would view their picture and profile and decide whether they wanted to meet the participant. During the test session on a subsequent day, participants were told that each of the 30 children they had previously rated had also seen the participant's photo and profile and decided whether they wanted to meet the participant. Participants were then told that they would learn which of these children wanted to meet them. Trained experimenters delivered feedback about how the participants were ostensibly rated by other children in several phases. The photos of the 30 other children were arranged on two boards, one green and one red. The 10 photos of children that the participant wanted to meet were placed on the green board, and the 20 photos of children the participant did not want to meet were placed on a red board. Children were told that each photo would be moved to a set of two new boards, one green and one red. Photos moved to the green board were children who wanted to meet the participant; photos moved to the red board were children who did not want to meet the participant. First, the experimenter delivered feedback about 5 of the 10 children the participant wanted to meet. Each of these photos were moved to the red board, indicating that these children did not want to meet the participant. Next, the experimenter delivered feedback about 10 of the 20 children the participant did not want to meet. Half of the photos were moved to the green board, and half were moved to the red board. These two rounds were then repeated for the remaining 5 children the participant wanted to meet, followed by the remaining 10 children the participant did not want to meet. Throughout the feedback phase, evaluators viewed a piece of paper on a clipboard before moving each photo to appear as though they were reading the responses of each of the 30 children about the participant. Physiological responses were recorded during the 2-min periods when experimenters provided feedback about the 10 children who the participant wanted to meet (and who, ostensibly, were not interested in meeting the participant). Participants then completed a 5-min recovery period, completed a brief affect questionnaire, and provided a second saliva sample.

Next, participants completed the TSST, which involves three periods. Participants were first told that they would be delivering a speech in front of teachers who would judge their performance. They were then given 5 min to prepare their speech. Participants were also told that their speech would be videotaped and evaluated by additional experts at a later time. Participants were asked to talk about the qualities of a good friend and which of those characteristics they did and did not possess. Next, participants delivered a 5-min speech in front of two evaluators. Evaluators were trained to provide neutral and mildly negative feedback (e.g., appearing bored) during the speech. If participants

were unable to speak for the full 5 min, they were first prompted to continue speaking. If they were again unable to continue, the evaluators asked a series of standardized questions (e.g., "what makes you a particularly good friend?) until the 5 minutes was completed. Finally, participants completed a mental subtraction task out loud in front of the evaluators for 5 min. Specifically, participants were asked to count backwards in steps of seven from a three-digit number and were stopped and asked to start again each time they made a mistake. If participants were unable to do the mental subtraction (i.e., did not get more than two answers correct within the first minute), the experimenters gave them an easier mental subtraction task involving counting backwards by three from a smaller number. After the TSST, participants completed another 5-min recovery period and provided a third saliva sample.

Finally, participants completed a frustration task. They were told that they would be playing a game in which a number appeared on a computer screen, and that they needed to press that same number on the keyboard as quickly as possible. Correct responses provided within the time allowed for a response were followed by a green smiling face and a positive noise; incorrect responses and those that occurred outside the allowable time window were followed by red frowning face and an irritating noise. The task had two distinct phases, each lasting 5 min. During the training phase, participants were given accurate feedback about their performance and had am 800-ms window in which to provide a response. During the test phase, participants had a shortened window in which to respond (700 ms) and received inaccurate feedback about their performance on 30% of trials (i.e., correct responses were given feedback indicating an inaccurate response). Together, this resulted in a high proportion of feedback indicating incorrect responses during the test phase. Following the frustration task, participants provided a fourth saliva sample. The final saliva sample was taken 20 min following the fourth saliva sample.

Physiological Data Scoring. ECG and ICG data were scored by raters unaware of group status. Signals were averaged into 1-min epochs by using Mindware Software (Mindware Technologies). ECG data were scored to provide measures of HR and RSA, a pure measure of parasympathetic nervous system activation. PEP, a measure of SNS activation representing the amount of time that elapses from the beginning of ventricular depolarization to the opening of the aortic valve (electrical systole), was calculated based on the ECG and ICG signals. The Q onset in the ECG and the B onset in the ICG were placed using validated automated scoring algorithms (2, 3) that were visually inspected to ensure accurate placement and adjusted if needed. Stroke volume (SV) was estimated from the dz/dt signal, providing an estimate of the amount of blood ejected from the heart on each cardiac cycle (4).

Sensitivity of Cortisol and DHEA-S Assays. The standard range of cortisol values that can be detected with the cortisol immunoassay used here is 0–110 nmol/L. The standard range of DHEA-S values that can be detected with the immunoassays is 0–12 ng/mL. The range of cortisol and DHEA-S was within the sensitivity bounds for the immunoassays used to assay these markers. The range of cortisol values in our sample was 1.37–18.77 nmol/L at baseline, 2.09–25.29 nmol/L at the second saliva assessment (postpeer-evaluation task), 2.19–39.35 nmol/L at the third saliva assessment (post-TSST), 1.89–28.37 nmol/L at the fourth saliva assessment (postfrustration task), and 2.07–19.36 nmol/L at the fifth saliva assessment (recovery). The range of values for

DHEA-S in our sample was 0.24–9.46 ng/mL at baseline, 0.19–10.1 ng/mL at the second saliva assessment (postpeer evaluation task), 0.22–11.70 ng/mL at the third saliva assessment (post-TSST), 0.20–11.90 at the fourth saliva assessment (postfrustration task), and 0.23–10.70 at the fifth saliva assessment (recovery).

As noted in *Laboratory-Based Tasks* above, our saliva samples were taken ~20 min following the part of each task that was associated with peak physiological and emotional reactivity (the second round of peer rejection in the peer evaluation task, the speech portion of the TSST, and the beginning of test phase of the frustration task). Evidence suggests that cortisol values peak ~20–25 min following a stressor (5) and, as such, it is possible that our saliva assessments missed the peak cortisol concentration following each task.

1. Howarth GZ, Guyer AE, Pérez-Edgar K (2013) Young children's affective responses to acceptance and rejection by peers: A computer-based task sensitive to variation in temperamental shyness and gender. *Soc Dev* 22(1):146–162.
2. Berntson GG, Lozano DL, Chen Y-J, Cacioppo JT (2004) Where to Q in PEP. *Psychophysiology* 41(2):333–337.
3. Lozano DL, et al. (2007) Where to B in dZ/dt. *Psychophysiology* 44(1):113–119.
4. Sherwood A, et al. (1990) Methodological guidelines for impedance cardiography. *Psychophysiology* 27(1):1–23.
5. Kirschbaum C, Pirke K-M, Hellhammer DH (1993) The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28(1-2):76–81.

CONSORT
TRANSPARENT REPORTING of TRIALS

BEIP: Placement at Age 12

Assessed for eligibility (n=187)

Excluded (n=51)
• Genetic syndromes (e.g., Down syndrome)
• Fetal alcohol syndrome
• Microcephaly

Enrollment

Randomized (n=136)

Allocated to Foster Care (n=68)

Allocation

Allocated to Care as Usual (n=68)

Follow-Up

Placement at 12 years (n=56)
29 MacArthur Foster Care
08 Government Foster Care
02 Adopted
12 Reintegrated with Biological Family
05 Institutional Care

Discontinued participation (n=12)

Placement at 12 years (n=58)
20 Institutional Care
14 Government Foster Care
06 Adopted
18 Reintegrated with Biological Family

Discontinued participation (n=10)

Analysis

Analyzed (n= 48)
Physiological data not collected (n= 8)

Analyzed (n= 43)
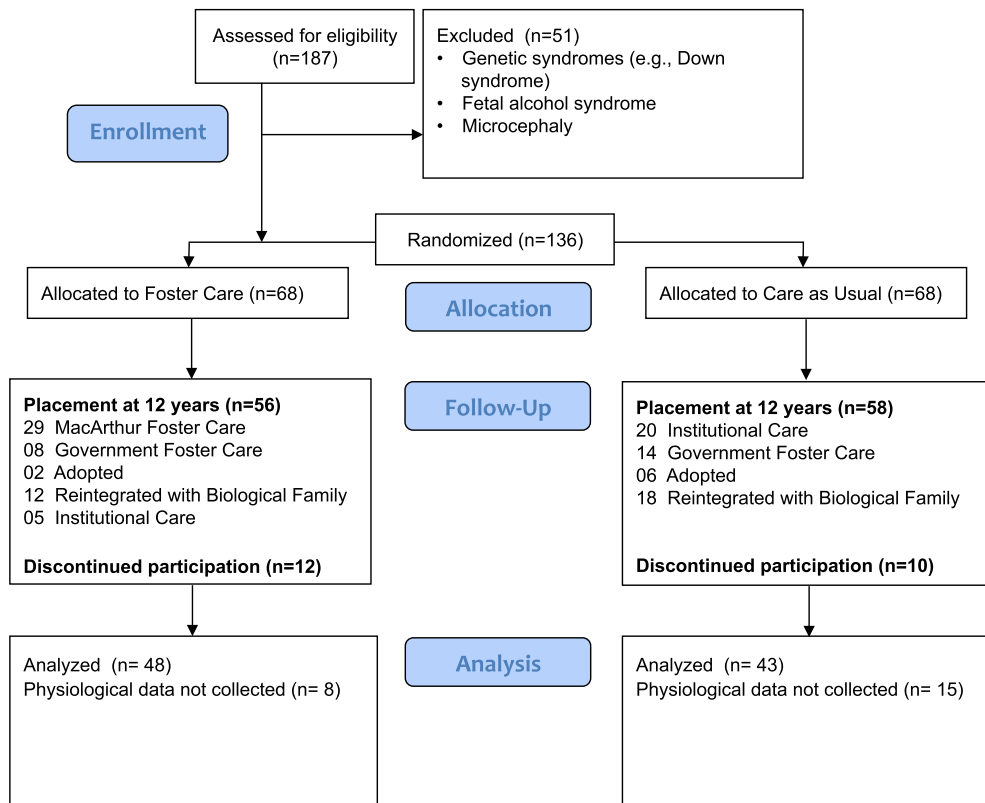Physiological data not collected (n= 15)

Fig. S1.   CONSORT diagram for BEIP. Figure describes the enrollment, randomization, assessment, and analysis of children involved in the randomized controlled trial. Typically developing children in NIG were not part of the randomized controlled trial and, thus, are not included in the figure.
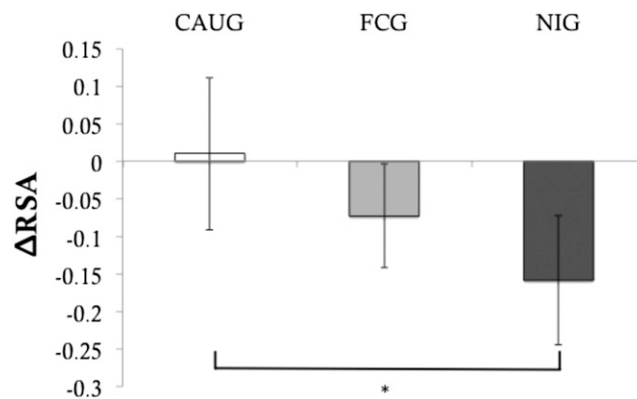
Fig. S2. Group differences in ANS reactivity to frustration. Figure depicts changes in respiratory sinus arrhythmia from the training phase to the test phase of the frustration task. $*P < 0.05$, two-sided test.

Table S1. ANS values at baseline and during each task by group

ANS measure	CAUG M	CAUG SD	FCG M	FCG SD	NIG M	NIG SD	Group difference F	Group difference P value
Baseline								
HR	83.34	(13.36)	78.57	(12.63)	81.81	(10.51)	1.84	0.16
SBP	111.04	(9.98)	112.76	(8.81)	108.28	(7.42)	3.12*	0.047
DBP	57.14	(6.23)	57.57	(5.72)	56.64	(5.19)	0.30	0.74
RSA	6.45	(1.13)	6.72	(1.05)	6.46	(1.08)	0.95	0.39
PEP	93.40	(9.94)	100.70	(9.84)	101.57	(7.63)	10.61*	<0.001
Peer evaluation								
HR	86.92	(13.07)	82.16	(11.87)	86.78	(11.08)	2.36	0.099
SBP	111.50	(8.25)	115.98	(10.04)	113.86	(9.69)	2.41	0.094
DBP	57.86	(5.67)	60.97	(7.05)	61.22	(6.20)	3.50*	0.033
RSA	6.19	(1.03)	6.50	(0.90)	6.29	(0.95)	1.46	0.24
PEP	93.17	(9.90)	99.66	(9.74)	98.72	(8.89)	5.96*	0.003
Speech preparation								
HR	89.85	(13.38)	86.76	(14.47)	91.29	(12.10)	1.31	0.27
SBP	116.71	(11.87)	119.60	(14.08)	116.42	(8.92)	0.92	0.40
DBP	59.81	(8.12)	64.09	(7.31)	62.42	(7.81)	3.02	0.052
RSA	6.29	(1.04)	6.53	(0.93)	6.40	(0.94)	0.69	0.50
PEP	92.03	(10.18)	97.60	(10.93)	97.88	(8.63)	4.61*	0.012
Speech								
HR	91.96	(12.43)	90.92	(15.84)	98.37	(15.24)	3.18*	0.045
SBP	121.84	(15.86)	123.53	(12.35)	129.34	(12.31)	3.07*	0.050
DBP	60.82	(10.04)	63.19	(9.49)	69.43	(9.19)	7.47*	<0.001
RSA	6.38	(0.90)	6.61	(0.85)	6.40	(1.08)	0.79	0.45
PEP	90.82	(9.85)	95.64	(12.73)	93.24	(10.63)	1.94	0.15
Math								
HR	91.64	(12.05)	90.57	(16.57)	99.58	(17.87)	4.47*	0.013
SBP	121.07	(10.94)	124.27	(14.01)	130.02	(10.93)	5.55*	0.005
DBP	62.04	(8.40)	65.00	(9.57)	68.50	(8.79)	4.85*	0.010
RSA	6.33	(0.95)	6.53	(0.97)	6.28	(1.27)	0.69	0.50
PEP	91.68	(9.95)	96.75	(12.16)	92.31	(10.19)	2.83	0.063
Frustration								
HR	86.93	(11.70)	82.01	(12.24)	86.53	(10.57)	2.58	0.079
SBP	114.44	(10.75)	117.02	(11.12)	114.50	(10.07)	0.57	0.57
DBP	55.41	(9.19)	59.66	(8.90)	58.94	(7.54)	2.01	0.14
RSA	5.85	(1.04)	6.33	(1.05)	5.86	(1.11)	3.11*	0.048
PEP	94.52	(9.77)	101.87	(8.70)	101.67	(7.01)	10.33*	<0.001

Table depicts raw values of ANS measures at baseline and during each task. CAUG, care-as-usual group; DBP, diastolic blood pressure; FCG, foster care group; HR, heart rate; M, mean; NIG, never institutionalized group; PEP, pre-ejection period; RSA, respiratory sinus arrhythmia; SBP, systolic blood pressure. $*P < 0.05$, two-sided test.

Table S2. HPA axis values at baseline and during each task by group

HPA axis measure	CAUG M	CAUG SD	FCG M	FCG SD	NIG M	NIG SD	Group difference F	P value
Baseline								
Cortisol	7.08	(3.61)	6.08	(3.62)	6.86	(3.85)	1.00	0.37
DHEA-S	2.43	(1.57)	2.18	(1.83)	2.80	(2.10)	1.36	0.26
Peer evaluation								
Cortisol	7.16	(4.67)	5.45	(2.38)	6.26	(3.02)	4.54*	0.012
DHEA-S	3.07	(2.15)	2.45	(2.12)	2.86	(2.10)	1.36	0.26
Trier Social Stress Test								
Cortisol	7.92	(3.90)	8.36	(6.61)	11.30	(5.57)	5.07*	0.008
DHEA-S	3.13	(2.23)	2.53	(1.16)	2.92	(2.35)	0.84	0.43
Frustration								
Cortisol	7.13	(3.18)	7.55	(5.57)	10.80	(5.51)	6.67*	0.002
DHEA-S	3.01	(2.14)	2.60	(1.30)	2.96	(2.39)	0.43	0.65
Recovery								
Cortisol	6.60	(2.84)	6.35	(3.96)	8.19	(3.84)	2.47	0.088
DHEA-S	2.85	(1.86)	2.54	(1.29)	2.87	(2.15)	0.31	0.74

Table depicts raw values of HPA axis measures at baseline and following each task. CAUG, care-as-usual group; DBP, diastolic blood pressure; FCG, foster care group; HR, heart rate; M, mean; NIG, never institutionalized group. *$P < 0.05$, two-sided test

Table S3. ANS reactivity scores by group

ANS measure	CAUG M	CAUG SD	FCG M	FCG SD	NIG M	NIG SD	Group difference[†] F	P value
Peer evaluation								
HR	3.58	(4.03)	3.37	(3.50)	4.96	(4.41)	2.00	0.140
SBP	1.78	(4.84)	3.17	(8.27)	6.01	(8.22)	3.29*	0.040
DBP	1.06	(4.86)	3.37	(5.47)	4.94	(5.88)	4.82*	0.010
RSA	−0.25	(0.47)	−0.21	(0.56)	−0.17	(0.53)	0.41	0.66
PEP	−0.50	(2.41)	−1.25	(2.84)	−2.85	(4.61)	5.18*	0.007
Speech preparation								
HR	4.50	(5.32)	6.83	(9.31)	7.85	(6.74)	2.11	0.126
SBP	7.58	(11.75)	7.20	(10.88)	9.53	(10.33)	0.55	0.58
DBP	3.09	(7.82)	6.69	(8.24)	7.64	(9.08)	3.07*	0.050
RSA	0.08	(0.80)	−0.12	(1.09)	−0.01	(0.88)	0.48	0.62
PEP	−1.37	(3.07)	−2.53	(5.75)	−4.43	(5.35)	3.73*	0.027
Speech								
HR	12.07	(8.03)	16.74	(12.65)	23.03	(15.01)	7.65*	<0.001
SBP	11.99	(14.86)	15.14	(11.82)	24.24	(15.28)	7.13*	<0.001
DBP	2.70	(10.35)	7.53	(10.28)	14.95	(13.11)	10.82*	<0.001
RSA	−0.03	(0.83)	−0.11	(0.99)	−0.13	(1.18)	0.08	0.92
PEP	−4.23	(5.82)	−7.56	(8.00)	−13.39	(10.42)	11.58*	<0.001
Math								
HR	8.64	(7.77)	12.84	(14.65)	18.65	(13.10)	6.54*	0.002
SBP	12.00	(15.05)	13.80	(12.62)	21.92	(11.54)	5.73*	0.004
DBP	5.59	(8.49)	9.56	(8.40)	14.37	(7.53)	10.63*	<0.001
RSA	−0.11	(0.84)	−0.15	(1.03)	−0.17	(1.01)	0.02	0.98
PEP	−2.10	(5.57)	−5.52	(8.21)	−10.85	(10.31)	11.31*	<0.001
Frustration								
HR	1.29	(2.40)	0.86	(2.00)	1.88	(2.00)	2.79	0.065
SBP	−1.17	(5.75)	−2.40	(11.56)	−1.70	(10.46)	0.10	0.90
DBP	−7.10	(5.15)	−3.96	(7.16)	−4.50	(7.38)	1.34	0.27
RSA	0.09	(0.46)	0.02	(0.34)	−0.15	(0.45)	4.13*	0.018
PEP	0.53	(1.39)	0.28	(1.84)	0.15	(1.48)	0.57	0.57

Table depicts reactivity scores of ANS measures during each task. CAUG, care-as-usual group; DBP, diastolic blood pressure; FCG, foster care group; HR, heart rate; M, mean; NIG, never institutionalized group; PEP, pre-ejection period; RSA, respiratory sinus arrhythmia; SBP, systolic blood pressure. *$P < 0.05$, two-sided test.
[†]Analysis controls for sex.

Table S4. Results of multilevel models examining cortisol and DHEA-S across the study session

HPA axis parameters	Parameter estimate	SE	t value	P value
Cortisol[†]				
Intercept				
CAUG vs. FCG	−0.04	(0.05)	−0.83	0.40
CAUG vs. NIG	1.0	(0.05)	1.95	0.053
FCG vs. NIG	0.13	(0.05)	2.76*	0.006
Slope				
CAUG vs. FCG	0.08	(0.03)	2.43*	0.016
CAUG vs. NIG	0.16	(0.03)	5.34*	<0.001
FCG vs. NIG	0.08	(0.03)	2.74*	0.006
DHEA-S[‡]				
Intercept				
CAUG vs. FCG	−0.07	(0.07)	−1.05	0.30
CAUG vs. NIG	−0.08	(0.07)	−1.20	0.23
FCG vs. NIG	−0.01	(0.06)	−0.15	0.88
Slope				
CAUG vs. FCG	0.01	(0.01)	0.65	0.52
CAUG vs. NIG	−0.01	(0.01)	−0.31	0.76
FCG vs. NIG	−0.01	(0.01)	−0.94	0.35

*$P < 0.05$, two-sided test.

[†]Cortisol values were skewed and log-transformed before analysis. The best-fitting model included linear, quadratic, and cubic effects for time. The intercept and linear slope of cortisol were modeled as random effects; the quadratic and cubic terms were modeled as fixed effects. Group effects were tested with linear contrasts between groups as specified above; group effects were included on the intercept and linear, quadratic, and cubic changes across time.

[‡]DHEA-S values were skewed and log-transformed before analysis. The best-fitting model included linear and quadratic effects for time. The intercept and both linear and quadratic effects of time on DHEA-S were modeled as random effects. Group effects were tested with linear contrasts between groups as specified above; group effects were included on the intercept and both linear and quadratic changes across time.