SUPPLEMENT 1

Validity Check of Task Manipulations

Emotional Valence. Beginning with valence ratings, scores for the full sample differed according to both the valence condition ($F_{1,149} = 434.81, p < .001, \eta^2 = 0.75$) and the regulation condition ($F_{1,149} = 8.01, p < .005, \eta^2 = 0.06$). The interaction of valence and regulation conditions was significant, consistent with the interpretation that the experience of each emotion valence (negative and positive) was greater in the suppression than the expression of an emotion, particularly for negative emotions ($F_{1,149} = 21.95, p < .001, \eta^2 = 0.13$). Analysis of the simple effects confirmed that the four conditions differed as indicated by our labeling of the conditions: higher (more positive) ratings for positive than negative emotion (all $p < .001$). In addition, higher (more positive) ratings for induction than suppression were also observed (all $p < .001$) (see Table S1, available online). Also reassuringly, the attention-deficit/hyperactivity disorder (ADHD), ADHD-low-prosocial, and control groups were similar with respect to Self-Assessment Manikin (SAM) valence rating scores, as none of the interactions involving group status were significant (all $F < 1.0$, all $p > .10$), suggesting that interpretation of changes in autonomic reactivity were not confounded by differences in self-reported valence appraisal of the task conditions. Furthermore, groups did not differ with respect to valence during any of the task conditions (all $F < 1.0$, all $p > .10$) (see Table S1, available online).

Emotional Arousal. The arousal rating scores showed a similar picture (see Table S1, available online). There was a significant main effect for arousal scores between positive and negative segments ($F_{1,149} = 107.31, p < .001, \eta^2 = 0.43$). In addition, the interaction of valence and regulation conditions was significant, $F_{1,149} = 8.35, p < .005, \eta^2 = 0.06$). However, the suppression versus induction difference was not significant ($F_{1,149} = 1.52, p = .22, \eta^2 = 0.01$), and none of the interactions involving group status were meaningful (all $F < 1.0$) (see Table S1, available online) for the task condition descriptive data according to group. Specifically, groups did not differ with respect to arousal level during any of the task conditions (all $F < 1.0$, all $p > .10$) (Table 2). Thus, on the whole, these data suggest that the task manipulations were effective and valid across all groups.

Validity Check Based on Task Habituation and Order Effects. Task habituation and order effects were evaluated by comparing physiological measures across the resting and neutral baseline conditions. Repeated-measures analysis of variance (ANOVA) indicated that respiratory sinus arrhythmia (RSA), interbeat interval, and pre-ejection period (PEP) differed across the baselines (all $F > 8.76, p < .01$), suggesting systematic changes in physiology when comparing rest to the demands imposed

TABLE S1 Self-Assessment Manikin (SAM) Scores Across Task Conditions, by Group.

Variable	Control (n = 75)	ADHD: ADHD Only (n = 54)	ADHD: ADHD-Low-Prosocial (n = 21)
SAM Valence/Pleasure			
Baseline 1	3.39 (0.67)	3.54 (0.79)	3.45 (0.81)
Negative induction	2.44 (1.17)	2.48 (1.31)	2.58 (1.08)
Negative suppression	1.73 (0.88)	1.74 (1.07)	1.91 (1.05)
Baseline 2	3.38 (0.71)	3.45 (1.22)	3.46 (0.81)
Positive induction	4.47 (0.81)	4.44 (1.13)	4.68 (0.67)
Positive suppression	4.64 (0.77)	4.67 (0.89)	4.79 (0.42)
SAM Intensity/Arousal			
Baseline 1	2.04 (0.72)	2.31 (0.80)	2.15 (0.76)
Negative induction	3.51 (1.25)	3.61(1.45)	3.37 (0.89)
Negative suppression	3.20 (1.14)	3.09 (1.46)	3.19 (1.11)
Baseline 2	1.93 (0.91)	2.23 (0.81)	1.85 (0.72)
Positive induction	2.11 (1.24)	2.20 (1.48)	2.11 (1.29)
Positive suppression	2.32 (1.36)	2.48 (1.63)	2.16 (1.26)

Note: For this table, no significant group differences were observed. Each row represents next time point in design. Time increasing down table for the repeated-measures design. ADHD = attention-deficit/hyperactivity disorder.

by attending to and orienting to a neutral task, as expected. Second, a polynomial repeated-measures ANOVA for the full sample revealed that the linear effect of time on RSA across all task conditions was significant ($F_{1,149} = 9.82$, $p = .002$, $\eta^2 = 0.06$), as was the quadratic effect ($F_{1,149} = 8.24$, $p = .00$, $\eta^2 = 0.05$) and the cubic effect ($F_{1,149} = 32.92$, $p < .001$, $\eta^2 = 0.18$). This is consistent with participants responding to the task manipulations and inconsistent with a habituation effect. Simple examination of the means (Figures 1 and 2) confirms that neither PEP nor RSA simply decreased across task conditions. The Appendix lists the exact means for specialists wishing to make detailed comparisons.

A different pattern was observed for PEP: here, no significant effect was seen for the linear, quadratic, or cubic effect (all $F < 1.0$, all $p > .05$). This suggests there was no change in PEP across task conditions and that group main effects of PEP would be the more appropriate focus of results. Interbeat interval and respiration rate displayed a similar consistency across conditions all linear, quadratic, and cubic effects (all $F < 2.3$, all $p > .05$).

In addition, group comparisons revealed that during the neutral and resting baselines there were no significant differences in RSA, heart rate, or respiration rate (all $F < 1.0$, all $p > .10$) (see Table S1, available online). However, groups did differ with respect to PEP during the neutral baseline only, during which the ADHD and control groups showed higher arousal than the ADHD + callous/unemotional traits (CU) group ($F_{2,148} = 3.7$, $p < .05$ (Figure 1). We conclude that habituation and order effects were trivial relative to task manipulation effects, which further supports the results of our primary analyses and hypothesis testing.

Validity Check Based on Overall Physiological Responding. In addition, our hypothesis was that there would be specific responses in sympathetic and parasympathetic systems. However, an alternative explanation for the PEP and RSA results might be that they are caused by global fluctuations in the participants' overall physiology, such as interbeat interval or respiration rate, rather than isolated actions of sympathetic or parasympathetic systems. If so, the results should emerge similarly if we simply look at interbeat interval or respiration rate.

To test this possibility, the main effects and interactions for interbeat interval and respiration rate were examined using the same $2 \times 2 \times 3$ repeated-measures ANOVA. The main effects of task valence (positive/ negative affect) and regulation (suppression/induction) for interbeat interval were significant ($F_{1,149} = 5.392$, $p < .05$, $\eta^2 = 0.04$) and $F_{1,149} = 21.555$, $p < .001$, $\eta^2 = 0.13$), respectively, with interbeat interval changing in the task-appropriate direction. However, none of the interactions by group were significant (all $F < 2.10$ and $p > .10$). In addition, there were no significant effects with respect to respiration rates (all $F < 3.10$, $p > .10$). Thus, the groups did not differ on general, multi-determined, physiological parameters, suggesting that it was appropriate to interpret the PEP and RSA effects as specific indexes of sympathetic and parasympathetic activity, respectively, as intended.