**Supplemental Information**

**Methods**

**Participants.**

Individuals in this longitudinal sample were not a typical community sample but instead are a sample exhibiting higher rates of negative and positive reactivity at 4 months. This sampling procedure was done to enhance the rates of psychopathology in this sample. Previous reports (Fox et al, 2015) have demonstrated that in this sample (as well as others) negative reactivity is associated with increased behavioral inhibition. One open question is whether negative reactivity or positive reactivity (i.e., the groups that are oversampled) is also associated with greater irritability. In this sample reactivity group is not associated with differences in childhood irritability (*p*>.533) or mean ARI (*p*<.510).

**Measures**

**ERN.** A total of 178 longitudinally followed infants came in for the 12-year visit. However, 33 children did not provide useable ERN data because they either because their data had too much artifact, the experimenter was unable to properly attach the EEG net, and/or experimenter error. An additional 18 children were excluded because they provided too few trials of each type. Critically, children who did provide useable data did not differ from those that did not in terms of gender, childhood BI, childhood irritability, social anxiety or 12-year irritability (*p*s>.217). Figure S1 depicts a histogram of the distribution of ERN scores.

**Modeling high stable irritability.** Table S1 provides the model fit parameters for latent class growth analysis. Figure S2 depicts the trajectory classes of childhood irritability.

**ARI as index of irritability.** In addition to the ICC of the ARI being quite small, there was a small but significant correlation between parent and child ARI (*r*(89)=.242, *p*<.022). Furthermore, as an additional check, we examined whether mean ARI was correlated with the CBCL internalizing and externalizing subscales at age 12. Mean ARI was significantly correlated with externalizing at age 12 (*r*(81)=.393, *p*<.001) controlling for internalizing but did not correlate with internalizing controlling for externalizing (p>.888). Correlations for parent and child report of the ARI are presented in the Table S2.

**Associations between childhood irritability and ARI at age 12.** In the main text, we focus our early childhood irritability phenotype on the children who are classified as high stable irritable. The decision to do so is supported by the strong correlation between high stable irritability in childhood and ARI at age 12. Nevertheless, given that other studies have shown that most classes of irritable children are associated with risk for psychopathology, we provide information about these other classes of irritable children. Table S3 depicts the correlations between early childhood irritability classification (probability) scores and ARI at age 12. Results demonstrate that in our sample, the only classification of childhood irritability that is positively associated with 12-year irritability is the high stable irritable class.

**Results**

**Replotting data using Mean +/- 1 Standard Deviation.**

In interpreting the interactions reported in the main text we split the data using tertiles. However, it is often preferable to split data based on the mean and 1 standard deviation above and below the mean. Figures S3 & S4 report our data split using the mean +/- 1 standard deviation method. In our case, we chose to report on even tertile splits rather than the mean and standard deviation approach because our groups were too small upon using the mean +/- 1 standard deviation approach—particularly for the irritability analyses. Nevertheless, we report here how the results change when this analytic approach is taken. To determine the cut-points for this approach we computed the mean ERN magnitude for the entire sample (*M*=-2.2) and one standard deviation above (.15) and below (-4.56) the mean. Individuals who had ERN scores smaller than -4.56 were grouped together as exhibiting a high ERN (n=24, 19% of sample with any ERN data). Individuals who had ERN scores between -4.56 and .15 were categorized as having a moderate ERN (n=84; 66% of sample). Individuals with ERN scores greater than .15 were categorized as having a low ERN (n=19, 15% of sample).

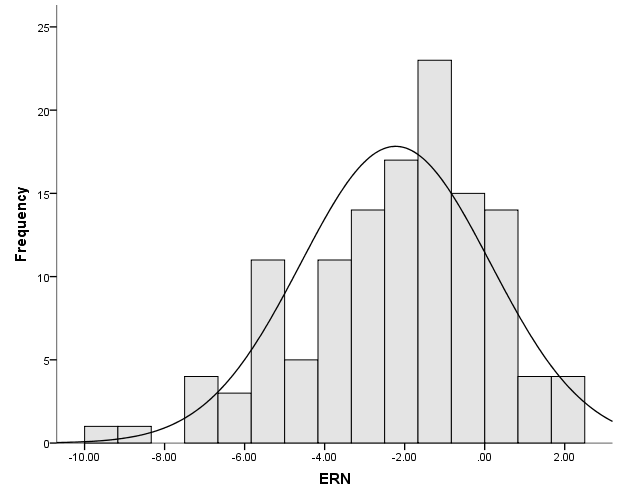
**Behavioral inhibition and social anxiety**. Following this grouping then tested whether the association between BI and social anxiety held for each group. Results indicated that there was a significant association between BI and social anxiety for the large (*r*(22)=.626, *p*<.001) and moderate (*r*(83)=.436, *p*<.001) ERN groups but not for the low ERN group (*p*<.600). These findings converge with those reported in the main text and hold when controlling for childhood irritability. Compare Figure 2 in the Main Text to Figure S3.

**Childhood irritability predicts later ARI.** We then ran the same analysis using the childhood irritability and average ARI scores. Results indicated that there was a significant relation between childhood irritability and 12 year irritability in the moderate ERN group (*r*(42)=.456, *p*<.002) but not in the large ERN group (*p*>.366) or the low ERN group (*p*>.607). These results hold when controlling for behavioral inhibition. Together, these results do not converge with those reported in the main text (Compare Figure 3 in the Main text to Figure S4). We suspect that this may be due to data loss in evaluating groups of individuals who have both irritability phenotype data **and** were categorized as high ERN (*n*=13) or low ERN (*n*=7) using this +/- 1 standard deviation approach of grouping. The data reported in the main text have larger sample sizes per group and thus, we suspect better reflect individual differences in our data set. It is also notable that group size difference between the tertile split and the +/- 1 standard deviation approach is larger in the irritability analyses than the BI analyses.

**Irritability Groups as a multi-categorical IV.** While the probability of being classified as high stable irritable over the course of childhood is a strong reliable predictor of 12 irritability, we could have also run our focal analyses using irritability group (a categorical variable). This approach results in the same overall pattern of results. The association between childhood irritability group and age 12 irritability differed as a function of ERN magnitude (β=.682, Δ*R2* = .227, *F*(3,54) = 6.386, *p*<.001). Interestingly, here we see the same pattern with those exhibiting high stable irritable children with a low and moderate ERN (*p*s<.002) showing heightened irritability at age 12. This pattern is not present in those high stable irritable children with a high ERN. Furthermore, in contrast to the focal analyses reported in the main text, we also see an interaction in the group of children who show low-increasing irritability—such that those with low-increasing irritability who exhibit a high ERN show increased irritability at age 12.

**Using Arcsine transformed scores for childhood irritability**. While our main text reports childhood irritability using non-transformed irritability scores, the results did not change if we instead substituted the arcsine transformed childhood irritability scores. See Table S4.

**Supplemental Analysis Strategy.** To supplement our analyses in the main text, we re-ran all focal analyses using permutation testing (5000 iterations; utilizing the tool Permutation Analysis of Linear Models (PALM); Winkler et al, 2014). Permutation testing further validated the strength of these relations: BI-Social Anxiety (*p*<.0002), stable childhood irritability-ARI (*p*<.0008). Permutation testing further validated the strength of the moderation analyses. The BI x ERN *t*(122)=-2.00, *p*<.022; IRR x ERN *t*(59)=2.93, *p*<.001). By replicating our focal findings using permutation testing, we demonstrate that our results are robust to non-normality and control precisely the Type I error rate.



*Figure S1.* Illustrates the distribution of ERN scores.

*Table S1. Model fit indices for latent class growth analysis of childhood irritability.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Latent Classes Modeled | AIC | BIC | SSBIC | Entropy | VLMR  (*p*) | LMR  *(p)* | BLRT  *(p)* | Smallest class |
| 2 | 3401.583 | 3436.677 | 3404.977 | 0.805 | <.001 | <.001 | <.001 | 31% |
| 3 | 3376.494 | 3422.116 | 3380.906 | 0.732 | 0.3777 | 0.3923 | <.001 | 11% |
| **4** | **3351.609** | **3407.76** | **3357.04** | **0.745** | **0.681** | **0.6856** | **<.001** | **10.50%** |
| 5 | 3312.916 | 3379.594 | 3319.364 | 0.809 | 0.0014 | 0.0019 | <.001 | <1% |
| 6 | 3297.261 | 3374.467 | 3304.727 | 0.813 | 0.0348 | 0.0426 | <.001 | <1% |

Note: Best-fitting model (4 classes; shown in bold) was determined based on considering all model fit indices and interpretability. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; BLRT = Bootstrapped Likelihood Ratio Test; LC = number of latent classes; LMR-LRT = Lo-Mendell-Rubin Adjusted Likelihood Ratio Test; SSABIC = sample size-adjusted BIC; VLMR = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test.

C2 26 (11%)

C3 44 (18%)

C1 138 (56%)

C4 39 (16%)

*Figure S2.* Depicts the trajectory classes of irritability. Lines indicate estimated trajectory of each class. C1= Class 1, etc. Numbers indicate count of individuals in each class and percentages indicate percent of sample assigned to each class.

*Table S2. Pearson Correlation Coefficients Indicating the Relation between parent and child report ARI and Internalizing and Externalizing subscales of CBCL at age 12.*

|  |  |  |
| --- | --- | --- |
|  | Internalizing CBCL age 12 | Externalizing CBCL age 12 |
| Parent ARI | .310\*\* | .448\*\*\* |
| Child ARI | .083 | .290\* |

\*\*\*p<.001, \*\*p<.005, \*p<.01

*Table S3. Correlation coefficients for each irritability classification and ARI at age 12*

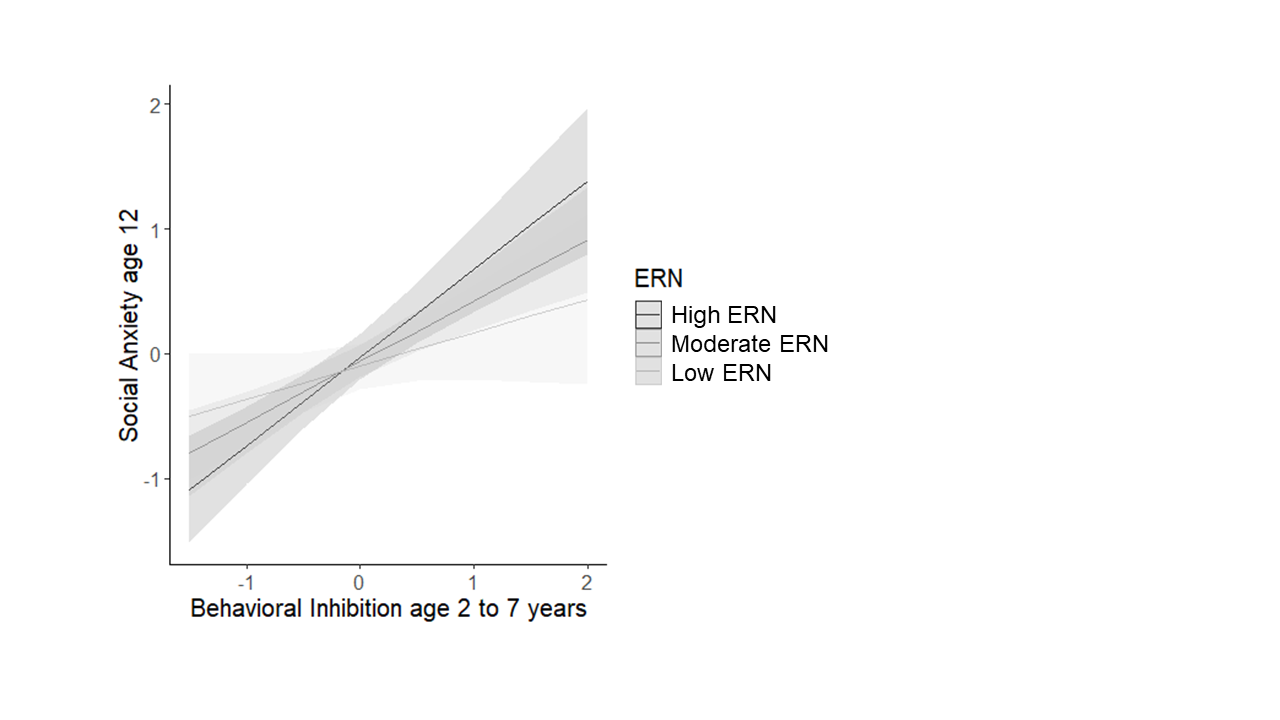
|  |  |
| --- | --- |
| Irritability class | ARI age 12 |
| C1-Low stable irritable | -.282\* |
| C2-Low increasing | .112 |
| C3-High decreasing | -.042 |
| C4- High stable irritable | .350\*\* |

*\*p<.01, \*\*p<.001*

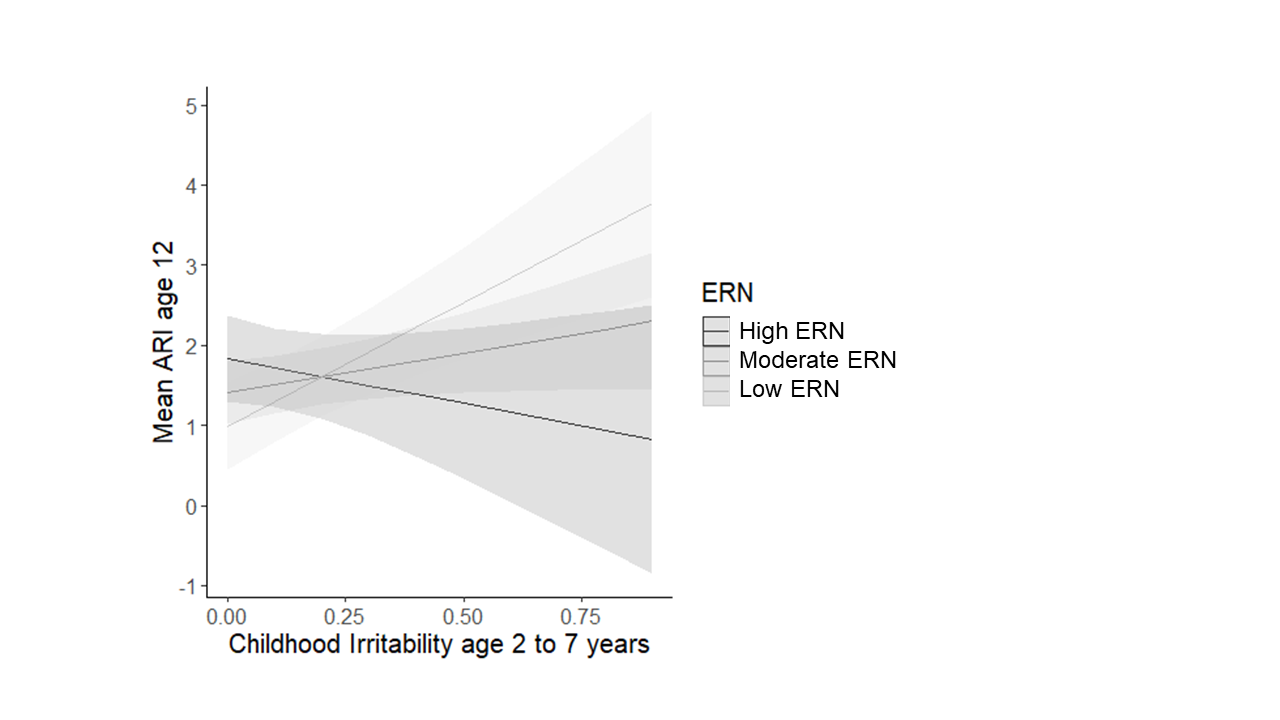
*Table S4.*

Regression results predicting ARI at age 12 (using arcsine transformed childhood irritability scores)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Predictor | β | *SE B* | *t* | *p* | *Fit* |
| 1 | High childhood irritability (Arcsine transformed) | 2.12 | .485 | 4.359 | .0001 | R2=.252, *p*<.001 |
|  | ERN | -.176 | .081 | -2.186 | .033 |  |
|  | High childhood irritability x ERN | .676 | .241 | 2.808 | .007 |  |
| 2 | High childhood irritability (Arcsine transformed) | 2.095 | .488 | 4.293 | .0001 | R2=.259, *p*<.001 |
|  | ERN | -.183 | .081 | -2.247 | .028 |  |
|  | High childhood irritability x ERN | .688 | .242 | 2.844 | .006 |  |
|  | BI | .216 | .285 | .756 | .453 |  |
|  |  |  |  |  |  |  |



*Figure S3.* Depicts the moderating effect of ERN on risk for developing social anxiety. Each line depicts low/moderate/high ERNs as determined by dividing the data using mean +/- 1 standard deviation (SD). Moderate ERN represents mean ERN scores; Low ERN represents scores less than 1 SD below the mean; High ERN represents scores greater than 1 SD above the mean.



*Figure S4*. Depicts the moderating effect of ERN on risk for exhibiting irritability symptoms at age 12. Each line depicts low/moderate/high ERNs as determined by dividing the data using mean +/- 1 standard deviation (SD). Moderate ERN represents mean ERN scores; Low ERN represents scores less than 1 SD below the mean; High ERN represents scores greater than 1 SD above the mean.

Supporting Information References

Fox, N. A., Snidman, N., Haas, S. A., Degnan, K. A., & Kagan, J. (2015). The relations between reactivity at 4 months and behavioral inhibition in the second year: Replication across three independent samples. *Infancy*, *20*(1), 98-114. [**https://doi.org/10.1111/infa.12063**](https://doi.org/10.1111/infa.12063)