Supplemental Text S.1

We followed the steps clearly outlined by Widaman et al. (2012) and Belsky, Pluess, and Widaman (2013). Code adapted from Widaman et al. (2012) is available upon request.

**Standard regression models**

First, we ran four standard regression models for each child outcome: a main-effects model (model 1) and three GxE models for each environmental exposure (models 2-4). All models contained an intercept, the p-factor, all three environmental exposures and five control variables for a total of 10 parameters. The interaction models contained one additional product term for a total of 11 parameters. Simplified notations omitting the extra environmental effects and control variables for the main effects model (equation 1a) and the interaction models (equation 1b) are as follows:

(1a)

(1b)

Where = the outcome of either child externalizing problems or social competence, = the intercept, = the parameter estimate for the binary p-factor variable , = the parameter estimate for the continuous environmental variable of interest , = the parameter estimate of the product term of and , and = the error term.

Second, we calculated a cross-over point on the environmental variable of interest where the predicted p-factor slopes cross from the standard regression models using the equation for all significant interactions visually in line with differential susceptibility theory (Aiken & West, 1991).

**Re-parameterized models**

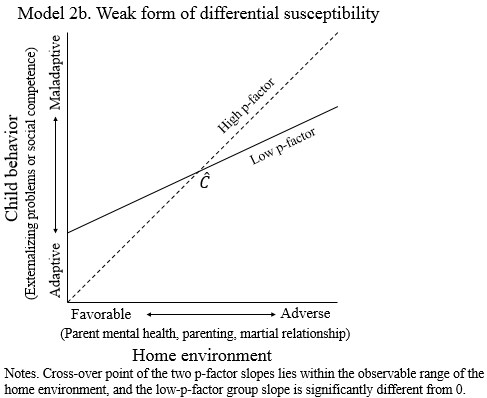
Third, we tested “strong” and “weak” forms of both DST and diathesis-stress for a total of 4 additional models by re-parameterizing the standard regression equations. The main benefit of re-parameterizing the models was to obtain an associated confidence interval for the cross-over point in order to fully assess whether the interaction is truly disordinal in nature. A simplified version of the models is written as:

(2)

Where is the intercept, or the estimated value of at , and and are the slopes of the two p-factor groups on the environmental variable . All other terms in the model were defined above.

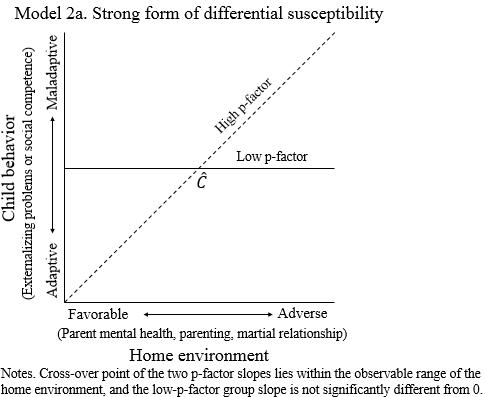
**Weak DST (model 2b)**

We subtracted the cross-over point from the observed value of the environmental variable of interest and re-ran the regression equation.



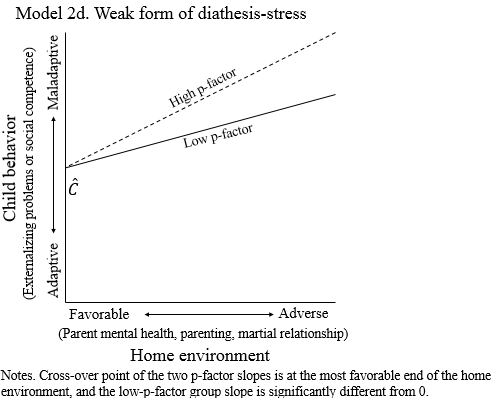
**Strong DST (model 2a)**

As with the weak DST model, we subtracted the cross-over point from the observed value of the environmental variable of interest . We then further re-parameterized the model by setting the slope of the low p-factor group to 0 and re-ran the regression equation. This model represents a more extreme form of DST where the low p-factor group is not at all reactive to their home environment.



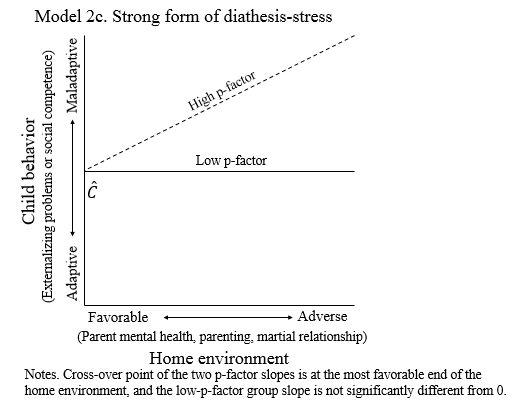
**Weak Diathesis-Stress (model 2d)**

Next, we tested weak diathesis-stress by subtracting the most positive extreme value of (or most adaptive environment) from the observed value of the environmental variable of interest and re-ran the regression equation. In the case of diathesis-stress, we would expect the cross-over point to be at the most extreme value of where both genetic groups would have similar levels of healthy functioning (see figure 1b). If diathesis-stress is the most likely explanation, difference between genetic groups should only appear as the environment becomes more adverse.



**Strong Diathesis-Stress (model 2c)**

Finally, we tested strong diathesis-stress by subtracting the most positive extreme value of (or most adaptive environment) from the observed value of the environmental variable of interest. We then further re-parameterized the model by setting the slope of the low p-factor group to 0 and re-ran the regression equation.



**Analysis of variance F-test**

Fourth, we compared the four regression models using the analysis of variance F-test. The strong DST (2a), weak diathesis-stress (2c) and strong diathesis-stress (2d) are all nested within the weak DST (2b) model, and can be directly tested for decreases in explained variance. Additionally, model 6b has one more parameter than model 2a, and Model 2c has one less parameter than model 6a. Both can be tested for a relative decrease or increase in explained variance, respectively. The strong DST (2a) and weak diathesis-stress (2d) models were not nested and could not be directly compared to one-another using the F-test, so total R-squared and the AIC and BIC were used.

References

Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Belsky, J., Pluess, M., & Widaman, K. F. (2013). Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses. *Journal of Child Psychology and Psychiatry, 54*(10), 1135-1143. doi:10.1111/jcpp.12075

Widaman, K. F., Helm, J. L., Castro-Schilo, L., Pluess, M., Stallings, M. C., & Belsky, J. (2012). Distinguishing ordinal and disordinal interactions. *Psychological Methods*, *17*(4), 615-622.