


## Regular Article

# Infant temperament, early-childhood parenting, and early-adolescent development: Testing alternative models of Parenting × Temperament interaction

Xiaoya Zhang , Kristina Saylor, Sarah Hartman and Jay Belsky

Department of Human Ecology, University of California, Davis, USA

### Abstract

Here we evaluate whether infant difficult temperament (6 months) functions as a vulnerability or more general plasticity factor when investigating effects of early-childhood parenting (8–42 months) on both positive and negative early-adolescent socioemotional development (age 8–11 years). Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC,  $N = 14,541$ ) and a re-parameterized model-testing approach to distinguish alternative person × environment conceptual models, results indicated that temperament × parenting interacted in predicting externalizing (i.e., hyperactivity, conduct problems), but not other behavior (i.e., emotional symptoms, peer problems), in a (weak) differential susceptibility manner. While more and less supportive parenting predicted, respectively, fewer and more behavior problems, it did so more strongly for children who were more difficult as infants.

**Keywords:** temperament-X-parenting interaction, differential-susceptibility, diathesis-stress, re-parameterized model comparison, ALSPAC

(Received 17 March 2020; revised 9 November 2020; accepted 11 November 2020; First Published online 15 January 2021)

Humans are a highly plastic species, with experiences and exposures, especially early in life, influencing individual differences in development (e.g., Bornstein, 1989). This is not to say that genetic factors are not influential, only that they are not the only source of individual differences. Two widely embraced models of environmental influence raise the prospect that there are actually individual differences in developmental plasticity and environmental sensitivity. The classic diathesis–stress framework stipulates that some individuals are more vulnerable than other (resilient) individuals to the negative effects of contextual adversity, be that for genetic, physiological, or temperamental reasons (Monroe & Simons, 1991; Zuckerman, 1999). An alternative to this “vulnerability” model is that of differential susceptibility; it stipulates that individuals vary more generally in their developmental plasticity such that some are not only at particular risk when exposed to environmental adversity, but also disproportionately likely to benefit from supportive or enriched developmental contexts – or even just benign ones (Belsky, 1997a,b, 2005; Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007; Belsky & Pluess, 2009, 2013; Ellis, Boyce, Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2011). In other words, instead of some individuals being only especially susceptible to the negative effects of adversity, differential susceptibility

thinking posits that such individuals are also particularly susceptible to the benefits of support and enrichment. Weak and strong versions of both models can be distinguished. Strong models reflect the fact that some individuals are highly susceptible, whereas others are not at all; weak versions stipulate that all individuals are susceptible, but some more so than others.

In child development research that focusses on vulnerability, one individual difference characteristic that is significant is temperament, especially difficult temperament in infancy (Rothbart & Bates, 2006). Both negative emotionality, reflecting frequency, intensity and proneness to distress, and difficult temperament, which typically includes additional challenging behaviors (e.g., high activity, low adaptability), have been a long-standing focus of research on Person × Environment interaction. Furthermore, there is repeated evidence that negative emotionality/difficult temperament in infancy and early childhood operates as a vulnerability factor, consistent with diathesis–stress thinking (e.g., Belsky & Pluess, 2011; Kochanska & Kim, 2012; Rothbart & Bates, 2006). There is also evidence from other inquiries, including some by the very same investigators, that this phenotypic characteristic functions as a more general plasticity factor, in line with differential susceptibility thinking (Belsky, 2005; Kochanska, Aksan, & Joy, 2007; Pluess & Belsky, 2009; Van Zeijl et al., 2007). Perhaps even more notable are the results of intervention studies showing that highly difficult/negatively-emotional infants and young children benefit more than others from efforts to promote well-being early in life, again consistent with the differential susceptibility model (Van den Bergy & Bus, 2014; Cassidy, Woodhouse, Sherman, Stupica, & Lejuez, 2011). All this should not be read to imply that different

**Author for Correspondence:** Xiaoya Zhang, Department of Human Ecology, University of California, Davis, CA 95616; E-mail: [xyazhang@ucdavis.edu](mailto:xyazhang@ucdavis.edu)

**Cite this article:** Zhang X, Saylor K, Hartman S, Belsky J (2022). Infant temperament, early-childhood parenting, and early-adolescent development: Testing alternative models of Parenting × Temperament interaction. *Development and Psychopathology* 34: 784–795, <https://doi.org/10.1017/S0954579420002096>

© The Author(s), 2021. Published by Cambridge University Press on behalf of Development and Psychopathology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

facets of temperamental difficulty early in life are always related to the same developmental outcomes (Eisenberg et al., 2010; Rothbart & Bates, 2006), after all, there is evidence that while early inhibition forecasts later internalizing problems (e.g., Prior, Smart, Sanson, & Oberklaid, 2000), limited self-regulation predicts later externalizing ones (e.g., Mathiesen & Prior, 2006). What needs to be appreciated, though, is that the issue under consideration in the current report is not any direct links between early temperament and a future phenotype, but the role of early temperament in moderating environmental effects.

In light of work indicating that negative emotionality or difficult temperament can function as a vulnerability or a more general plasticity factor, the recent meta-analysis of Slagt and associates (2016) of Temperament  $\times$  Parenting interaction research proves informative. Upon analyzing studies in which parenting was operationalized positively (e.g., warmth, positive control) or negatively (e.g., hostility, negative control), they discovered that when difficult temperament was measured in infancy, it operated as a general plasticity factor, moderating positive and negative parenting effects in, respectively, a “for-better-and-for-worse” manner (Belsky et al., 2007); yet when it was assessed after 12 months of age it functioned only as a diathesis, amplifying the adverse effects of negative parenting. There would seem to be evidence, then, that negative emotionality/difficult temperament can function as a general plasticity factor, consistent with differential susceptibility theorizing, at least when assessed in infancy. As the investigations included in the meta-analysis revealed, most temperament  $\times$  parenting research has focused on either negative parenting (and its absence) or positive parenting (and its absence), but rarely both. The current paper aims to extend existing research by relying on a bipolar parenting construct ranging from highly negative to highly positive, as recommended by Belsky and Pluess (2009). This is important because children who prove most and least susceptible to the beneficial effects of positive parenting may – or may not – also prove differentially susceptible to the adverse effects of negative parenting.

To summarize, it seems notable in light of theory and research on variation in sensitivity to parenting that one phenotypic moderator of the association between environmental exposure and developmental functioning is negative emotionality/difficult temperament manifested early in life. What remains unclear, given the focus of most investigations on either positive or negative parenting, is whether difficult temperament in infancy functions as a vulnerability or more general susceptibility factor. In the current report, we thus seek to extend research on the role of difficult temperament in moderating parenting effects by taking advantage of a large, UK population study known as ALSPAC (i.e., Avon Longitudinal Study of Parents and Children; [www.alspac.bris.ac.uk](http://www.alspac.bris.ac.uk)). In particular, we model this paper on the study by Belsky and Pluess (2011) on the role of temperament in predicting adolescent externalizing behavior and test the interaction of infant difficult temperament and parenting in early childhood in predicting indices of both positive socioemotional development (i.e., prosocial behavior) and negative socioemotional development (i.e., hyperactivity, conduct problems, emotional symptoms, peer problems) in early adolescence (i.e., ages 8–11 years).

Methodologically, we implemented the competitive and confirmatory model-testing analytic approach of Widaman et al. (2012; Belsky, Pluess, & Widaman, 2013; Belsky & Widaman, 2018) that distinguishes diathesis stress from differential susceptibility. It competitively evaluates the fit of both strong and weak versions of each model. Thus, we used this approach with (a) the largest sample studied to date examining the two models of

Person  $\times$  Environment interaction, (b) while implementing a quasi-experimental design focused on the third of children with the most difficult temperament in infancy and the third of children with the least difficult temperament. We proceeded in this manner, given the large size of the ALSPAC sample, because it insured that the two subgroups of children were indisputably different, temperamentally, as infants. Nevertheless, we also conducted sensitivity analyses by testing regions of significance (ROS) using the Johnson–Neyman technique (Johnson & Fay, 1950; Preacher, Curran, & Bauer, 2006) and by analyzing the data with the temperament variable retained in its original, continuous form using the full sample instead of extreme groups.

## Method

### Participants and design

Data for this study come from the ongoing ALSPAC. This cohort study was launched in the early 1990s in order to investigate modifiable influences on offspring’s health and development, among many other topics. Pregnant women were enrolled who had estimated delivery dates between 1 April 1991 and 31 December 1992 in the Avon area of England (Boyd et al., 2013; Fraser et al., 2013). Further details of ALSPAC can be found at <http://www.bristol.ac.uk/alspac/>. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Given the large sample size, we implemented a quasi-experimental design, by dividing children into three, equal-sized, temperamental subgroups, reflecting high, moderate, and low difficulty, using the composite temperament construct described below. This meant focusing only on the third of children in the highest and lowest temperament terciles, thereby insuring that no child was likely miscategorized. Thus, a sample of initially 14,541 mother–child dyads was reduced, following multiple imputation of the composite parenting and temperament constructs described below on all dyads, to 4,849–5,049 dyads with highly difficult infants and 4,951–5,015 dyads with the least difficult infants. These subsample sizes vary across the 20 rounds of imputation of missing data, as also described below. In the sensitivity analysis with the full sample, all 14,541 mother–child dyads were involved in the analyses.

### Measures

Data for this report were collected repeatedly, beginning when children were 6 months of age and extending to the child’s 11th year. Although we were not able to provide reliability statistics for temperament and developmental outcome measures due to lack of the item-level data, these measures have proved reliable and valid in previous studies.

**Difficult temperament.** To measure infant difficult temperament, the moderator construct in this study, mothers completed an adaptation of the Carey Infant Temperament Scale (CITS; Carey & McDevitt, 1978) when their child was 6 months of age. Since we sought to model this paper on Belsky and Pluess’s (2011) prior work, using a much larger sample, we relied on

the same five difficult-temperament-related subscales of CITS which they did: (a) activity (e.g., “shows much bodily movement [kicks, waves, arms] when given an injection”), (b) approach (e.g., “first reaction to any new procedure [first haircut, new medicine, etc.] is objection”), (c) adaptability (e.g., “still wary or frightened of strangers after 15 mins”), (d) mood (e.g., “cries when left to play alone”), and (e) intensity (e.g., “reacts strongly to strangers: laughing or crying”; see specific items in the Supplementary Table S1). Mothers thus rated 51 behavioral descriptors with respect to how often the baby manifested the behavior in question using a 6-point Likert scale, ranging from 1 (*almost never*) to 6 (*almost always*), with higher scores reflecting a more difficult temperament (see Supplementary Table S1 for items in each subscale). In line with our quasi-experimental design, infants scoring in the highest tercile were considered to have a difficult temperament (coded as 1) and those scoring in the lowest tercile were regarded as having an easy temperament (coded as 0).

**Parenting.** To assess parenting, the predictor construct in this inquiry, we drew on previously created measures of parenting from prior ALSPAC work focused on children aged 8, 24, 33, 38, and 42 months. All measurements were based on mother reports of their parenting-related feelings and behavior. Three previously identified and internally consistent parenting constructs (see Table 1 for items and scales for each construct), found to be valid in prior work, reflected (a) supportive versus unsupportive parenting (for consistency sake, labels of constructs are the same as used by prior scholars: Waylen & Stewart-Brown, 2010; e.g., “I really enjoy this child”, “I dislike/hate the mess that surrounds the child”), (b) hostility (Chong et al., 2016; e.g., “mother smacks child during tantrums”, “child is slapped”), and (c) stimulating interaction (Gutman & Feinstein, 2010, e.g., “singing to child”, “showing picture books”). For each construct, higher scores represent, respectively, more supportive parenting, less hostility, and more stimulating interaction. All three subscales, when subject to principal axis factoring with direct oblimin rotation, loaded positively on a single factor (see Table 2). Thus, the three scores were summed to create a composite index of supportive versus unsupportive parenting, with higher composite scores representing more supportive parenting.

**Emotional and behavioral problems and prosocial behavior.** The dependent developmental constructs used in this report were based on teachers’ responses to five subscales of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) when children were in school year 3 ( $M_{age} = 8.33$ ,  $SD_{age} = 3.74$  months) and 6 ( $M_{age} = 11.17$ ,  $SD_{age} = 3.90$  months). One SDQ construct, prosocial behavior, was used to represent children’s positive functioning. The other four SDQ problem-behavior constructs – hyperactivity, emotional symptoms, conduct problems, and peer problems – were each averaged across the two ages and then summed to create the total problem score used in the primary analysis. Each subscale includes five questions scored 0 (*not true*) to 2 (*certainly true*). Higher scores reflect more problems.

### Statistical analysis

Data analyses were conducted using online SAS OnDemand (SAS Institute Inc. 2013). Without evidence showing that the mechanism of missing values of our outcome variables (i.e., prosocial and problem behavior) is missing not at random (MNAR), we assumed the data are missing at random (MAR). Therefore we first used multiple imputation (MI; Rubin, 1987) with 20 imputed data sets to decrease the bias caused by missing values (Gottschall, West, & Enders,

2012). We included Temperament  $\times$  Parenting interaction in the imputation model because previous works suggested that excluding interaction terms introduces more bias in estimation (Tilling, Williamson, Spratt, Sterne, & Carpenter, 2016). In addition, although limited-range variables were included in the MI, we did not restrict the range of imputed values because doing so could result in biased estimates (Rodwell, Lee, Romaniuk, & Carlin, 2014). That is, MI could generate implausible values given its purpose of estimating statistics rather than reproducing the values that would otherwise exist had they not been missing. Results to be reported reflect the pooled results from all 20 imputed data sets.

In our next step, we checked the correlation between temperament and parenting because a significant association could compromise the interpretation of any detected temperament  $\times$  parenting (i.e., suggesting temperament:parenting correlation; Belsky et al., 2007). Because analysis of all 20 imputed data sets, including all cases (i.e., not just those in the highest- and lowest-difficulty groups), revealed a significant temperament-parenting correlation ( $r = -.11$ ,  $p < .001$ ), primary analyses testing the Temperament  $\times$  Parenting interaction was carried out after statistically adjusting the composite parenting variable for the composite temperament variable; this yielded a residual variable that became the parenting construct of interest (see Table 3 for variable correlations after adjusting for parenting variables).

The first and preliminary step before implementing Widaman et al.’s (2012; Belsky et al., 2013) competitive and confirmatory model testing approach was to determine, in a traditional regression model (i.e., main effects followed by interaction), whether the interaction term yielded an  $F$  ratio around 1.0 or more, as this would indicate that there was sufficient variation in the interaction term to proceed with formal model testing (Belsky & Widaman, 2018).

Upon passing the just-mentioned screen, we proceeded to evaluate four re-parameterized models: weak and strong versions of diathesis stress and of differential susceptibility (Belsky et al., 2013; Widaman et al., 2012) with the extreme groups (i.e., children with the most and the least difficult infant temperament). The key difference between diathesis–stress and differential susceptibility models is the placement of crossover point, which should be around the middle of the environmental variable (i.e., parenting) per differential susceptibility model and near the maximum value of parenting per diathesis–stress model. The key difference between weak and strong models, regardless of the paradigm, is whether, in the case of the current inquiry, (a) both the most and least difficult children appear affected by parenting, but with the discerned parenting effect stronger for the more difficult temperamental group (i.e., weak models) or (b) whether only in the case of children with difficult temperaments as infants is there an apparent effect of parenting (i.e., strong models).

Operationally, with the weak differential susceptibility model, all parameters except the crossover point are relaxed, including intercept, slope coefficients for both groups (i.e., children with low and high difficult temperament). The crossover point was fixed at the mean (i.e., 0) of the predictor variable (i.e., parenting) to achieve a better model fit.<sup>1</sup> Further constraining the slope on

<sup>1</sup>Although the original re-parameterization approach proposed by Widaman et al. (2012; Belsky et al., 2013; Belsky & Widaman, 2018) relaxed the crossover point for differential susceptibility models, the estimated crossover points yielded large standard errors in our case. Therefore, at the recommendation of one anonymous reviewer and upon consultation with Keith Widaman, we fixed the crossover point at the mean of the parenting variable for the two differential susceptibility models; this resulted in much-improved models.

**Table 1.** Parenting constructs, items, scales, and ages of measurement

Constructs	Items	Metrics	Measurement ages (months)
Supportive versus unsupportive	<i>1. I really enjoy this child</i>	1 = exact feeling; 4 = never feels	8, 33
	<i>2. I feel confident with my child</i>	1 = exact feeling; 4 = never feels	8, 33
	<i>3. It is a great pleasure to watch my child develop</i>	1 = exact feeling; 4 = never feels	8, 33
	<i>4. Having this child makes me feel fulfilled</i>	1 = exact feeling; 4 = never feels	8, 33
	5. I would have preferred that we had not had this baby/ child when we did	1 = exact feeling; 4 = never feels	8, 33
	6. I can't bear hearing the child cry	1 = exact feeling; 4 = never feels	8, 33
	7. I dislike/ hate the mess that surrounds the child	1 = exact feeling; 4 = never feels	8, 33
	8. I feel I have no time to myself	1 = exact feeling; 4 = never feels	8, 33
Hostility <sup>a</sup>	1. Child is slapped	1 = everyday; 5 = never	24
	2. Mother shouts at child when naughty	1 = everyday; 5 = never	24
	<i>3. Mother shouts at child when naughty</i>	1 = never; 5 = daily	42
	<i>4. Mother smacks child during tantrums</i>	1 = never; 5 = daily	42
Stimulating interaction	<i>1. Singing to the child</i>	1 = often; 4 = never	38, 42
	<i>2. Showing picture books</i>	1 = often; 4 = never	38, 42
	<i>3. Playing with toys</i>	1 = often; 4 = never	38, 42
	<i>4. Cuddling</i>	1 = often; 4 = never	38, 42
	<i>5. Physically playing</i>	1 = often; 4 = never	38, 42

Notes. Italicized items have been reverse coded.

<sup>a</sup>For sake of consistency, higher scores of hostility construct represent less hostility.

parenting for children with the least difficult infant temperament to be 0 leads to the strong version of the differential susceptibility model. Forcing the crossover point to be the maximum value of parenting (i.e., 4.95) leads to the weak version of the diathesis–stress model. Further constraining the slope on parenting for children with the least difficult infant temperament to be 0 leads to the strong version of diathesis–stress model (see Appendix for the equations). We evaluated model fitting according to multiple fit indices (i.e., Akaike information criterion [AIC], Bayesian information criteria [BIC],  $R^2$ ).

## Results

### Primary analysis

Results of preliminary analysis provided support for proceeding to model comparison in the case of the composite measure of total child problems, but not prosocial behavior. Table 4 shows the model fitting results for total behavior problems. Variance explained (i.e.,  $R^2$ ) and fit indices (i.e., AIC, BIC) proved most consistent with the weak version of the differential susceptibility model. Specifically, analysis of variance (ANOVA)-based nested model comparison (i.e., strong models are nested in weak model) revealed that this model explained more variance and yielded a good fit to the data compared with other models. This, coupled with the significant effect of parenting on both

children with histories of the most and the least temperamental difficulty, supported the weak differential susceptibility model, according to Widaman et al. (2012) criteria. Visual inspection of Figure 1a provides additional support for this conclusion.

### Secondary analysis

In an attempt to gain further insight into differential parenting effects on children with the most and least difficult temperament as infants, we decomposed the composite problem behavior outcome and reran the analyses just described using its four components as separate dependent variables. The full model-fitting analysis could only be applied to the hyperactivity and conduct problems components, however, as only these outcomes met the Belsky and Widaman (2018) criteria of the Temperament  $\times$  Parenting interaction in the preliminary regression analysis ( $F \geq 1.0$ ).

For hyperactivity (see Table 5 and Figure 1b) and conduct problems (see Table 6 and Figure 1c), the form of interaction proved similar to that discerned in the case of total behavior problems. Thus, the same statistical conditions as described above provided evidence that the weak differential susceptibility model fit the data best when it comes to predicting hyperactivity and conduct problems. That is, ANOVA-based model comparison revealed a significant increase in explained variance (i.e.,  $R^2$ ) by the weak differential-susceptibility model relative to the nested models (i.e.,



**Table 2.** Factor loadings and reliability for the parenting measures

Measures	Factor loadings	Cronbach's $\alpha$
Parenting		
Supportive versus unsupportive	0.63	0.64 (8 months); 0.67 (33 months)
Hostility	0.32	0.63
Stimulating interaction	0.31	0.67 (38 months); 0.66 (42 months)

strong version of differential susceptibility and diathesis–stress models) and model fit indices (i.e., AIC, BIC) suggested that the weak differential susceptibility model fit the data best.

**Sensitivity analyses**

We conducted two sensitivity analyses. First, we conducted ROS tests when predicting total problems, hyperactivity, and conduct problems. The ROS with respect to the parenting variable are displayed in Figure 2, where the significant regions suggest that children with the most difficult temperament exhibited significantly more (i.e., slope of temperament > 0) and less (i.e., slope of temperament < 0) problem behavior relative to children with the least difficult temperament when they experience less (or not) and more sensitive parenting, respectively.

Secondly, we replicated the primary analyses using the full sample with temperament retained as a continuous variable. Results proved generally consistent with those of our primary analyses based on our quasi-experiment design (see Supplementary Tables S2–S4). That is, differential susceptibility models were preferred over diathesis–stress models, though it proved impossible to distinguish between the weak and strong versions of differential susceptibility in the case of hyperactivity and conduct problems.

**Discussion**

Ever more evidence suggests that effects of parenting – and other environmental exposures and developmental experiences – vary across children, with some more susceptible to influence than others (e.g., Belsky & Pluess, 2009, 2013; Ellis et al., 2011). In line with this emerging understanding, the present study sought to extend work on Temperament  $\times$  Parenting interaction – by focusing on a large UK birth cohort and by evaluating whether temperament-moderated effects of parenting proved consistent with differential susceptibility theorizing or the diathesis–stress model of Person  $\times$  Environment interaction. Of note, the sample size of this study, even upon elimination of the moderate-temperament group in our quasi-experimental design, proved substantially larger than all prior studies focused on the temperament of infants and included in the recent meta-analysis Slagt, Dubas, Deković, and van Aken (2016) of research on Temperament  $\times$  Parenting interaction.

Thus, using data on 6-month temperament, parenting measured across 8–42 months, and teacher assessments of prosocial behavior and total problems in early adolescence, competitive and confirmatory re-parameterization model testing indicated that the moderating effect of temperament on problem behavior proved most consistent with the weak version of the differential

**Table 3.** Descriptive statistics of parenting, temperament, and developmental outcomes

Variable	M	SD	Min <sup>a</sup>	Max	1	2	3	4	5	6	7
1. Temperament	109.98	18.87	26.74	194.80	–	–	–	–	–	–	–
2. Parenting	0.00	1.06	–4.90	4.57	.00	–	–	–	–	–	–
3. Prosocial behavior	7.72	2.29	–1.86	17.07	.01***	.10***	–	–	–	–	–
4. Hyperactivity	2.56	2.63	–8.37	12.87	–.03***	–.14***	–.60***	–	–	–	–
5. Conduct problems	0.87	1.55	–6.82	10.00	–.02***	–.13***	–.58***	.64***	–	–	–
6. Emotional symptoms	1.38	1.80	–6.17	10.00	.04***	–.02***	–.16***	.22***	.20***	–	–
7. Peer problems	1.25	1.73	–6.10	10.00	.01***	–.04***	–.43***	.38***	.41***	.48***	–
8. Total problems	12.53	3.69	4.00	33.00	–.01***	–.10***	–.13***	.72***	.62***	.63***	.41***

\*\*\* $p < .001$ .

<sup>a</sup>The presence of negative minimum values were introduced by multiple imputation.

**Table 4.** Results for alternative models for total behavior problems in early adolescence

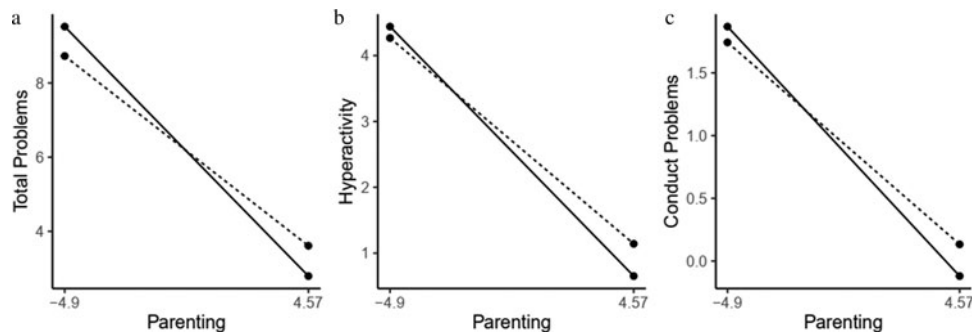
Standard parameterization			Re-parameterized regression equation				
Parameter	Parenting and temperament main effects: Model 1	Effects and interaction of parenting and temperament: Model 2	Differential susceptibility		Diathesis–stress		
			Parameter	Weak: Model 3a	Strong: Model 3b	Weak: Model 3c	Strong: Model 3d
$B_0$	6.08(0.10)***	6.08(0.10)***	$B_0$	6.06(0.07)***	6.06(0.07)***	3.64(0.35)***	5.90(0.10)***
$B_1$	−0.63 (0.08)***	−0.54(0.10)***	$B_1$	−0.71(0.12)***	−0.71(0.12)***	−0.63(0.08)***	−0.09(0.04)*
$B_2$	−0.04(0.16)	−0.04(0.16)	$C$	0.00 (−) <sup>a</sup>	0.00 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>
$B_3$	–	−.17(0.26)	$B_2$	−0.54(0.10)***	0.00 (−) <sup>a</sup>	−0.62(0.08)***	0.00 (−) <sup>a</sup>
$R^2$	.0136	.0138	$R^2$	.037	.029	.036	.019
$F$ versus 1	–	1.33	$F$ versus 3a	–	83.08	–	192.74
$df$	–	1, 86.23	$df$	–	1, 9,912	–	1, 9,912
$p$	–	.25	$p$	–	<.0001	–	<.0001
AIC	34,254.73	34,253.24	AIC	69,853.94	69,853.44	69,905.58	69,903.58
BIC	34,256.73	34,255.25	BIC	69,882.73	69,875.04	69,934.37	69,925.18

Notes.  $B_0$  represents intercept,  $C$  the crossover point. For Model 1,  $B_1$  and  $B_2$  represent the main effects of parenting and temperament (0 = least difficult and 1 = most difficult), respectively. For Model 2,  $B_1$  represents parenting effect for children with the least difficult temperament,  $B_2$  main effect of temperament,  $B_3$  Parenting × Temperament interaction effect (i.e., difference in slope on parenting for children with the most relative to the least difficult temperament). For Model 3a–3d,  $B_1$  and  $B_2$  represent slope on parenting for children with the most and the least difficult temperament, respectively. Values in parentheses are standard errors.  $F$  versus 1 stands for an  $F$  test of the interaction in Model 2.  $F$  versus 3a stands for an  $F$  test of the difference in  $R^2$  for a given nested model versus Model 3a.

AIC = Akaike information criterion; BIC = Bayesian information criteria

<sup>a</sup>Parameter fixed at a certain value; SE is not applicable

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$



**Figure 1.** Parenting  $\times$  Temperament interaction pattern for total problems (a), hyperactivity (b), and conduct problems (c). Solid lines represent children with highly difficult temperament, whereas dotted lines represent children with the least difficult temperament.

susceptibility model when implementing our extreme-group, quasi-experimental design, including only children with the most and least difficult temperaments as infants (in terms of terciles). While extreme-group designs are open to question, they have a long and informative history of use in developmental research on infant temperament (Fox, Henderson, Marshall, Nichols, & Ghera, 2005; Kagan, Reznick, & Snidman, 1987). To be clear, we adopted the extreme-group design for one specific reason – to insure that high- and low-difficulty infants were indisputably different in terms of temperament. As already noted, it seems noteworthy that the sensitivity analyses which included all children with the temperament construct measured continuously yielded results generally consistent with those of our primary, extreme-group approach.

Recall that results indicated that even though greater maternal support forecast significantly fewer behavioral problems for both children with more as well as less difficult temperaments, the predictive power of parenting proved greater in the case of the children who had more difficult temperaments as infants. Just as noteworthy – and due to reliance on a parenting measure that ranged from highly supportive to highly unsupportive – was the discovery (using ROS analysis) that this differential developmental response to parenting of the two groups of children involved both *more and less supportive* parenting. In other words, the children with difficult temperament as infants proved more “susceptible” than the others to both unsupportive and supportive parenting in this work, that is, “for better and for worse” (Belsky et al., 2007). Of note, this result regarding children’s total problem behavior was replicated in sensitivity analyses using the full sample, a continuous rather than categorical parameterization of difficult temperament, and ROS testing (rather than model comparison). It is also noteworthy that early-childhood parenting predicted early-adolescent functioning, a period of substantial development. Nevertheless, we would be remiss in highlighting these findings if we did not make clear that the effects under consideration were small in magnitude and that the great power of a large sample no doubt enhanced our ability to detect them.

When we decomposed the composite index of problem behavior in order to gain further insight into the findings just summarized, evidence indicated that the just-highlighted result pertaining to the total problem-behavior composite appeared driven by the differential effects of parenting on children’s hyperactivity and conduct problems, but not emotional symptoms or peer problems. Theory and empirical evidence suggest that hyperactivity and conduct problems represent externalizing problems, whereas emotional symptoms and peer problems reflect internalizing ones (e.g., Goodman, Lamping, & Ploubidis, 2010). Therefore,

our results would seem to indicate that the weak differential susceptibility effect under consideration only emerged when forecasting externalizing problems (i.e., hyperactivity, conduct problems), not internalizing ones (i.e., emotional symptoms, peer problems). This indicates that children with the most difficult temperament as infants proved somewhat more susceptible to parenting effects when it came to the development of externalizing behavior problems (i.e., hyperactivity, conduct problems), but not internalizing ones (i.e., emotional symptoms, peer problems). Parenting showed no main effect in the case of internalizing problems; only difficult temperament showed main effects in predicting emotional symptoms in particular ( $\beta = 0.09, p = .03$ ).

It is perhaps not surprising that temperament  $\times$  parenting interacted in predicting externalizing problems (i.e., hyperactivity, conduct problems). Such a finding may have emerged because supportive parenting promotes children’s development of self-regulation, thereby undermining the risks of externalizing problems (Eisenberg et al., 2005). In addition, children with a history of difficult temperament may be especially in need of externally imposed positive control in the form of supportive parenting in early childhood (Albers, Beijers, Risken-Walraven, Sweep, & de Weerth, 2016). Consequently, they are at higher risk of externalizing problems when experiencing unsupportive parenting, but at the same time disproportionately likely to exhibit low levels of problem behavior when exposed to supportive parenting. An alternative explanation of such “for-better-and-for-worse” Temperament  $\times$  Parenting interaction is that difficult temperament may serve as a proxy for some genetic markers of susceptibility in developing externalizing problems, which needs to be explored (Belsky et al., 2009).

With regard to the prediction of internalizing problems, our failure to detect the main effects of parenting while chronicling a main effect of temperament in predicting emotional symptoms is consistent with prior work. For example, Stoltz, Beijers, Smeekens, & Deković (2017) documented a predictive effect of 5-year-olds’ negative affectivity but not parenting behavior on internalizing problems of 11-year-olds (Stoltz et al., 2017). Consider also Pettit and associates’ (2008) intergenerational study which found that parental and grandparental history of major depression predicted children’s internalizing problems more strongly than parental partner status and relationship quality (Pettit, Olino, Roberts, Seeley, & Lewinsohn, 2008). Might these similar results be due to a stronger effect of genetics on internalizing than on externalizing problems?

Another possibility of our failure to detect either main effects of parenting or a Temperament  $\times$  Parenting interaction in the case of internalizing problems could have been due to our measure of temperament. Perhaps regulative characteristics, not captured by our

**Table 5.** Results for alternative models for hyperactivity in early adolescence

Standard parameterization			Re-parameterized regression equation				
			Differential susceptibility		Diathesis–stress		
Parameter	Parenting and temperament main effects: Model 1	Effects and interaction of parenting and temperament: Model 2	Parameter	Weak: Model 3a	Strong: Model 3b	Weak: Model 3c	Strong: Model 3d
$B_0$	2.65 (0.05)***	2.65(0.05)***	$B_0$	2.56(0.03)***	2.56(0.03)***	1.17(0.19)***	2.54(0.05)***
$B_1$	−0.36 (0.04)***	−0.33(0.04)***	$B_1$	−0.40(0.05)***	−0.40(0.05)***	−0.34(0.04)***	−0.01(0.02)
$B_2$	−0.17(0.07)***	−0.17(0.07)*	$C$	0.00 (−) <sup>a</sup>	0.00 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>
$B_3$	–	−0.07(0.06)	$B_2$	−0.33(0.04)***	0.00 (−) <sup>a</sup>	−0.38(0.04)***	0.00 (−) <sup>a</sup>
$R^2$	.022	.023	$R^2$	.052	.041	.052	.025
$F$ versus 1	–	1.16	$F$ versus 3a	–	108.93	–	273.90
$df$	–	1, 105.43	$Df$	–	1, 9,912	–	1, 9,912
$p$	–	.28	$P$	–	<.0001	–	<.0001
AIC	18,862.93	18,862.09	AIC	53,540.60	53,502.36	53,742.51	53,740.51
BIC	18,864.93	18,864.10	BIC	53,569.40	53,523.95	53,771.30	53,762.10

Notes.  $B_0$  represents intercept,  $C$  the crossover point. For Model 1,  $B_1$  and  $B_2$  represent the main effects of parenting and temperament (0 = least difficult and 1 = most difficult), respectively. For Model 2,  $B_1$  represents parenting effect for children with the least difficult temperament,  $B_2$  main effect of temperament,  $B_3$  Parenting × Temperament interaction effect (i.e., difference in slope on parenting for children with the most relative to the least difficult temperament). For Model 3a–3d,  $B_1$  and  $B_2$  represent slope on parenting for children with the most and the least difficult temperament, respectively. Values in parentheses are standard errors.  $F$  versus 1 stands for an  $F$  test of the interaction in Model 2.  $F$  versus 3a stands for an  $F$  test of the difference in  $R^2$  for a given nested model versus Model 3a.

AIC = Akaike information criterion; BIC = Bayesian information criteria

<sup>a</sup>Parameter fixed at a certain value; SE is not applicable

\* $p < .05$ . \*\*  $p < .01$ . \*\*\* $p < .001$



**Table 6.** Results for alternative models for conduct problems in early adolescence

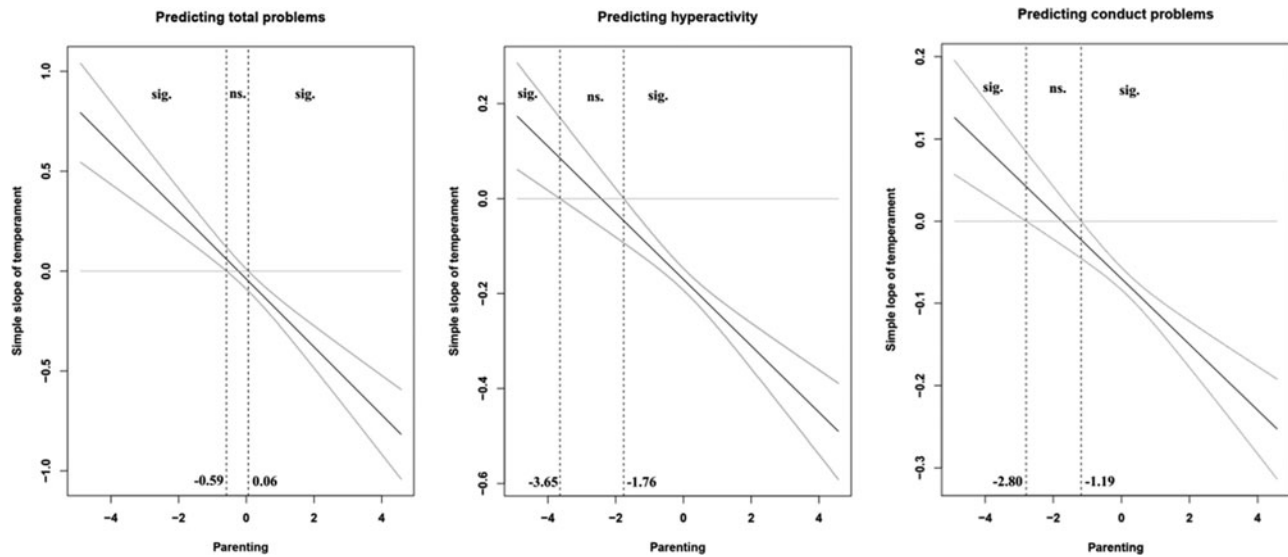
Standard parameterization			Re-parameterized regression equation				
			Differential susceptibility		Diathesis–stress		
Parameter	Parenting and temperament main effects: Model 1	Effects and interaction of parenting and temperament: Model 2	Parameter	Weak: Model 3a	Strong: Model 3b	Weak: Model 3c	Strong: Model 3d
$B_0$	0.91(0.02)***	0.91(0.03)***	$B_0$	0.87(0.02)***	0.87(0.02)***	0.15(0.11)	0.85(0.03)***
$B_1$	−0.19(0.02)***	−0.17(0.03)***	$B_1$	−0.21(0.03)***	−0.21(0.03)***	−0.18(0.02)***	−0.01(0.01)
$B_2$	−0.07(0.04)	−0.07(0.04)	$C$	0.00 (−) <sup>a</sup>	0.00 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>	4.95 (−) <sup>a</sup>
$B_3$	–	−0.04(0.04)	$B_2$	−0.17(0.03)***	0.00 (−) <sup>a</sup>	−0.19(0.02)***	0.00 (−) <sup>a</sup>
$R^2$	.017	.017	$R^2$	.041	.032	.041	.019
$F$ versus 1	–	1.24	$F$ versus 3a	–	97.59	–	231.17
$df$	–	1,142.03	$df$	–	1, 9,912	–	2, 9,912
$p$	–	.48	$p$	–	<.0001	–	<.0001
AIC	8,517.39	8,516.80	AIC	38,997.58	39,152.50	39,443.96	39,441.96
BIC	8,519.39	8,518.81	BIC	39,026.38	39,175.10	39,472.75	39,463.55

Notes.  $B_0$  represents intercept,  $C$  the crossover point. For Model 1,  $B_1$  and  $B_2$  represent the main effects of parenting and temperament (0 = least difficult and 1 = most difficult), respectively. For Model 2,  $B_1$  represents parenting effect for children with the least difficult temperament,  $B_2$  main effect of temperament,  $B_3$  Parenting  $\times$  Temperament interaction effect (i.e., difference in slope on parenting for children with the most relative to the least difficult temperament). For Model 3a–3d,  $B_1$  and  $B_2$  represent slope on parenting for children with the most and the least difficult temperament, respectively. Values in parentheses are standard errors.  $F$  versus 1 stands for an  $F$  test of the interaction in Model 2.  $F$  versus 3a stands for an  $F$  test of the difference in  $R^2$  for a given nested model versus Model 3a.

AIC = Akaike information criterion; BIC = Bayesian information criteria

<sup>a</sup>Parameter fixed at a certain value;  $SE$  is not applicable

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$



**Figure 2.** Regions of significance with respect to parenting for temperament effects on total problems, hyperactivity, and conduct problems. In each of the plot, the region between the two dashed lines represents nonsignificant region (denoted as “ns.”), whereas other regions represent significant regions (denoted as “sig.”). Significant and nonsignificant regions indicate that the children with the most and the least difficulty significantly differ or not, respectively, on problem behavior within a certain range of parenting. Values for the bound of regions with respect to parenting were shown in the plots next to the dashed lines above the horizontal axes. The gray lines represents the 95% confidence interval limits for slope coefficient of temperament.

temperament measure, protect children from risks of low maternal support, as revealed by research chronicling a three-way interaction involving reactive temperament (i.e., fearful inhibition), regulatory temperament (i.e., inhibitory control), and parenting on children’s internalizing problems (Liu, Calkins, & Bell, 2018). This result, coupled with the preceding speculative analysis of our temperament findings – and non-findings – suggests that future research would do well to include measures of both reactive and regulatory temperament when investigating Temperament  $\times$  Parenting interaction and children’s internalizing and externalizing problems. It will be interesting to see if such work replicates the specificity of the Temperament  $\times$  Parenting interactions documented herein.

It seems notable that while an anticipated Temperament  $\times$  Parenting interaction emerged in the case of externalizing problem behavior (i.e., hyperactivity, conduct problems), no such effect was detected in the case of prosocial behavior. This null result is in line with other work. Consider in this regard previous research highlighting different roles of specific parenting behavior interacting with specific temperament characteristics in predicting prosocial behavior (e.g., Augustine & Stifter, 2015). McGinley (2008) observed, for example, that angry/frustrated temperament, but not fearful and shy temperament, mattered, with the same being true of maternal firm discipline but not responsiveness. Clearly, it will be important for future investigations of Temperament  $\times$  Parenting interaction to investigate multiple dimensions of temperament and parenting.

To summarize, infant temperament and parenting interacted in a weak-differential-susceptibility-like manner. Such findings suggest that while parenting interventions could benefit all children when it comes to preventing or ameliorating externalizing (but not internalizing) problems, those with a history of difficult temperament in infancy could be most likely to benefit. Although no Temperament  $\times$  Parenting interaction emerged in predicting prosocial behavior, interventions aiming to promote maternal support may raise children’s prosocial behaviors regardless of infant temperament.

Whatever the strengths of the current work, especially the large sample size, reliance on a competitive and confirmatory model fitting approach, and use of a bipolar measure of parenting (ranging from more supportive to more negative), as well as a quasi-experimental design that insured dramatic differences in infant temperament across the two subgroups of children, this work is not without limits. Most notably, perhaps, is the exclusive focus on prosocial behavior for the “positive” side of development in contrast to multiple indicators of problematic development. Results of Temperament  $\times$  Parenting interaction could have been different had we had more information on children’s competencies (e.g., cooperation, academic achievement; data of these variables and other alternatives were not made available to us). Reliance on maternal reports of parenting is, of course, also limiting. Conceivably, findings could have been different had observational or adolescent reports of parenting been available. Our reliance on a multi-faceted measure of difficult temperament rather than a focus on different components of our composite construct also merits attention. Only future work will be positioned to address such possibilities.

**Supplementary Material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0954579420002096>.

**Acknowledgments.** This publication is the work of the authors who will serve as guarantors for the contents of this paper. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

**Funding Statement.** The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC.

**Conflicts of Interest.** None

## References

- Albers, E. M., Beijers, R., Risken-Walraven, J. M. A., Sweep, F. C. G. J., & de Weerth, C. (2016). Cortisol levels of infants in center care across the first

- year of life: Links with quality of care and infant temperament. *Stress: The International Journal on the Biology of Stress*, 19, 8–17. doi:10.3109/10253890.2015.1089230
- Augustine, M. E., & Stifter, C. A. (2015). Temperament, parenting, and moral development: Specificity of behavior and context. *Social Development*, 24, 285–303. doi:10.1111/sode.12092. doi: 10.1111/sode.12092
- Belsky, J. (1997a). Theory testing, effect-size evaluation, and differential susceptibility to rearing influence: The case of mothering and attachment. *Child Development*, 68, 598–600. doi:10.1111/j.1467-8624.1997.tb04221.x
- Belsky, J. (1997b). Variation in susceptibility to rearing influences: An evolutionary argument. *Psychological Inquiry*, 8, 182–186. doi: 10.1207/s15327965pli0803\_3
- Belsky, J. (2005). Differential susceptibility to rearing influences: An evolutionary hypothesis and some evidence. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind: Evolutionary psychology and child development* (pp. 139–163). New York, NY: Guilford.
- Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science*, 16, 300–304. doi:10.1111/j.1467-8721.2007.00525.x
- Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry*, 14, 746. doi:10.1038/mp.2009.44
- Belsky, J., & Pluess, M. (2009). *Beyond diathesis stress: Differential susceptibility to environmental influences*. *Psychological Bulletin*, 135, 885–908. doi:10.1037/a0017376
- Belsky, J., & Pluess, M. (2011). Differential susceptibility to long-term effects of quality of child care on externalizing behavior in adolescence? *International Journal of Behavioral Development*, 36, 2–10. doi:10.1177/0165025411406855
- Belsky, J., & Pluess, M. (2013). Beyond risk, resilience, and dysregulation: Phenotypic plasticity and human development. *Development and Psychopathology*, 25, 1243–1261. doi:10.1017/S095457941300059X
- 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, 1135–1143. doi:10.1111/jcpp.12075
- Belsky, J., & Widaman, K. (2018). Editorial perspective: Integrating exploratory and competitive-confirmatory approaches to testing person-x-environment interactions. *Journal of Psychology and Psychiatry*, 59, 296–298. doi:10.1111/jcpp.12824
- Bornstein, M. H. (1989). Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin*, 105, 179–197. doi:10.1037/0033-2909.105.2.179
- Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., ... Davey Smith, G. (2013). Cohort profile: The “children of the 90s” – the index offspring of the Avon longitudinal study of parents and children. *International Journal of Epidemiology*, 42, 111–127. doi:10.1093/ije/dys064
- Carey, W. B., & McDevitt, S. C. (1978). Revision of the infant temperament questionnaire. *Pediatrics*, 61, 735–739.
- Cassidy, J., Woodhouse, S. S., Sherman, L. J., Stupica, B., & Lejuez, C. W. (2011). Enhancing infant attachment security: An examination of treatment efficacy and differential susceptibility. *Development and Psychopathology*, 23, 131–148. doi:10.1017/s0954579410000696
- Chong, S. Y., Chittleborough, C. R., Gregory, T., Mittinty, M. N., Lynch, J. W., & Smithers, L. G. (2016). Parenting practices at 24 to 47 months and IQ at age 8: Effect-measure modification by infant temperament. *PLoS One*, 11, 1–16. doi:10.1371/journal.pone.0152452
- Eisenberg, N., Haugen, R., Spinrad, T. L., Hofer, C., Chassin, L., Zhou, Q., ... Liew, J. (2010). Relations of temperament to maladjustment and Ego resiliency in At-risk children. *Social Development (Oxford, England)*, 19, 577–600. doi:10.1111/j.1467-9507.2009.00550.x
- Eisenberg, N., Zhou, Q., Spinrad, T. L., Valiente, C., Fabes, R. A., & Liew, J. (2005). Relations among positive parenting and children’s effortful control and externalizing problems. *Child Development*, 76, 1055–1071. doi: 10.1111/j.1467-8624.2005.00897.x
- Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary neurodevelopmental theory. *Development and Psychopathology*, 23, 7–28. doi:10.1017/S0954579410000611
- Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology*, 56, 235–262. doi:10.1146/annurev.psych.55.090902.141532
- Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., ... Ring, S. (2013). Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42, 97–110. doi: 10.1093/ije/dys066
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the strengths and difficulties questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38, 1179–1191. doi:10.1007/s10802-010-9434-x
- Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research*, 47, 1–25. doi:10.1080/00273171.2012.640589
- Gutman, L. M., & Feinstein, L. (2010). Parenting behaviours and children’s development from infancy to early childhood: Changes, continuities and contributions. *Early Child Development and Care*, 180, 535–556. doi:10.1080/03004430802113042
- Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, 15, 349–367.
- Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*, 58, 1459–1473. doi:10.1111/j.1467-8624
- Kochanska, G., Aksan, N., & Joy, M. E. (2007). Children’s fearfulness as a moderator of parenting in early socialization: Two longitudinal studies. *Developmental Psychology*, 43, 222–237. doi:10.1037/0012-1649.43.1.222
- Kochanska, G., & Kim, S. (2012). Difficult temperament moderates links between maternal responsiveness and children’s compliance and behavior problems in low-income families. *Journal of Child Psychology and Psychiatry*, 54, 323–332. doi:10.1111/jcpp.12002
- Liu, R., Calkins, S. D., & Bell, M. A. (2018). Fearful inhibition, inhibitory control, and maternal negative behaviors during toddlerhood predict internalizing problems at age 6. *Journal of Abnormal Child Psychology*, 46, 1665–1675. doi:10.1008/s10802-018-0419-5
- Mathiesen, K. S., & Prior, M. (2006). The impact of temperament factors and family functioning on resilience processes from infancy to school age. *European Journal of Developmental Psychology*, 3, 357–387.
- McGinley, M. (2008). *Temperament, Parenting, and Prosocial Behaviors: Applying A new Interactive Theory of Prosocial Development* (Doctoral dissertation). Available from ProQuest Dissertations and Theses Database. (UNL No. 3315159)
- Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, 110, 406–425. doi:10.1037/0033-2909.110.3.406
- Pettit, J. W., Olino, T. M., Roberts, R. E., Seeley, J. R., & Lewinsohn, P. M. (2008). Intergenerational transmission of internalizing problems: Effects of parental and grandparental major depressive disorder on child behavior. *Journal of Clinical Child & Adolescent Psychology*, 37, 640–650. doi: 10.1080/15374410802148129
- Pluess, M., & Belsky, J. (2009). Differential susceptibility to rearing experience: The case of childcare. *Journal of Child Psychology and Psychiatry*, 50, 396–404. doi:10.1111/j.1469-7610.2008.01992.x
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448.
- Prior, M., Smart, D., Sanson, A. N. N., & Oberklaid, F. (2000). Does shy-inhibited temperament in childhood lead to anxiety problems in adolescence? *Journal of the American Academy of Child & Adolescent Psychiatry*, 39, 461–468. doi:10.1097/00004583-200004000-00015
- Rodwell, L., Lee, K. J., Romaniuk, H., & Carlin, J. B. (2014). Comparison of methods for imputing limited-range variables: A simulation study. *BMC Medical Research Methodology*, 14, 57. doi:10.1186/1471-2288-14-57

- Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 3, social, emotional and personality development* (6th ed., pp. 99–166). Hoboken, NJ: John Wiley & Sons, Inc.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- SAS Institute Inc. (2013). *SAS/ACCESS® 9.4 Interface to ADABAS: Reference*. Cary, NC: SAS Institute Inc.
- Slagt, M., Dubas, J. S., Deković, M., & van Aken, M. A. G. (2016). Differences in sensitivity to parenting depending on child temperament: A meta-analysis. *Psychological Bulletin*, *142*, 1068–1110. doi:10.1037/bul0000061
- Stoltz, S., Beijers, R., Smeekens, S., & Deković, M. (2017). Diathesis stress or differential susceptibility? *Testing Longitudinal Associations Between Parenting, Temperament, and Children's Problem Behavior. Social Development*, *26*, 783–796. doi: 10.1111/sode.12237
- Tilling, K., Williamson, E. J., Spratt, M., Sterne, J. A. C., & Carpenter, J. R. (2016). Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *Journal of Clinical Epidemiology*, doi:10.1016/j.jclinepi.2016.07.004
- Van den Berg, H., & Bus, A. G. (2014). Beneficial effects of BookStart in temperamentally highly reactive infants. *Learning and Individual Differences*, *36*, 69–75. doi:10.1016/j.lindif.2014.10.008
- Van Zeijl, J., Mesman, J., Stolk, M. N., Alink, L. R. A., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., ... Koot, H. M. (2007). Differential susceptibility to discipline: The moderating effect of child temperament on the association between maternal discipline and early childhood externalizing problems. *Journal of Family Psychology*, *21*, 626–636. doi:10.1037/0893-3200.21.4.626
- Waylen, A., & Stewart-Brown, S. (2010). Factors influencing parenting in early childhood: A prospective longitudinal study focusing on change. *Child: Care, Health and Development*, *36*, 198–207. doi:10.1111/j.1365-2214.2009.01037.x
- 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*, 615–622. doi:10.1037/a0030003
- Zuckerman, M. (1999). *Vulnerability to psychopathology: A biosocial model*. Washington, DC, US: American Psychological Association. doi:10.1037/10316-001

## Appendix

In the exploratory phase, we first ran multiple regression models using parenting and temperament (0 = the least difficult temperament group, and 1 = the most difficult temperament group) to predict problem behavior (i.e., Model 1 in Tables 4–6) based on this equation:

$$Y_i = B_0 + B_1 \text{ Parenting}_i + B_2 \text{ Temperament}_i + e_i \quad (1)$$

where  $Y_i$  is the dependent variable,  $B_0$  the intercept,  $B_1$  and  $B_2$  the main effects of parenting and temperament, respectively.

Then we ran multiple regression models (i.e., Model 2 in Tables 4–6) further including the product of temperament and parenting based on the following equation:

$$Y_i = B_0 + B_1 \text{ Parenting}_i + B_2 \text{ Temperament}_i + B_3 (\text{Parenting}_i \cdot \text{Temperament}_i) + e_i \quad (2)$$

where  $Y_i$  is the dependent variable,  $B_0$  the intercept,  $B_1$  the slope on parenting for easy temperament group,  $B_2$  the effects of temperament,  $B_3$  the effects of the Parenting  $\times$  Temperament interaction, representing the difference in slope on parenting between the most and the least difficult temperament groups.

Next we conducted the re-parameterized model testing (i.e., Model 3a–3d in Tables 4–6) based on the equation below:

$$Y_i: \begin{cases} \text{Temperament} = 1 & Y_i = B_0 + B_1 (\text{Parenting}_i - C) + e_i \\ \text{Temperament} = 0 & Y_i = B_0 + B_2 (\text{Parenting}_i - C) + e_i \end{cases} \quad (3)$$

where  $Y_i$  is the dependent variable,  $B_1$  and  $B_2$  stand for slope on parenting for the most and the least difficult temperament groups, respectively,  $B_0$  the intercept,  $C$  the crossover point. For differential susceptibility models,  $C$  were fixed at 0 – the mean of the predictor variable (i.e., parenting), with  $B_2$  further fixed at 0 for the strong version. For diathesis–stress models,  $C$  were fixed at 4.57 – the maximum value of the predictor, with  $B_2$  additionally fixed at 0 for the strong version.

Finally, we conducted alternative model testing, as a sensitivity analysis, with temperament remained as a continuous variable. This sensitivity analysis was based on the following equation:

$$Y_i = B_0 + B_1 (\text{Parenting}_i - C) + B_3 (\text{Parenting}_i - C) \cdot \text{Temperament}_i + e_i \quad (4)$$

where  $Y_i$  is the dependent variable,  $B_0$  the intercept,  $C$  the crossover point,  $B_1$  the slope on parenting for the least difficult group, and  $B_3$  the difference in slope on parenting between the most and the least difficult temperament groups. For differential susceptibility models,  $C$  were fixed at 0 – the mean of the predictor variable (i.e., parenting), with  $B_1$  further fixed at 0 for the strong version. For diathesis–stress models,  $C$  were fixed at 4.95 – the maximum value of the predictor, with  $B_1$  additionally fixed at 0 for the strong version.