

Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses

Jay Belsky,^{1,2,3} Michael Pluess,⁴ and Keith F. Widaman⁵

¹Department of Human Ecology, University of California, Davis, CA, USA; ²Department of Special Education, King Abdulaziz University, Jeddah, Saudi Arabia, USA; ³Department of Psychological Sciences, Birkbeck University of London, London, UK; ⁴Institute of Psychiatry, Kings College, London, UK; ⁵Department of Psychology, University of California, Davis, CA, USA

Background: Most gene-environment interaction (GXE) research, though based on clear, vulnerability-oriented hypotheses, is carried out using exploratory rather than hypothesis-informed statistical tests, limiting power and making formal evaluation of competing GXE propositions difficult. **Method:** We present and illustrate a new regression technique which affords direct testing of theory-derived predictions, as well as competitive evaluation of alternative diathesis-stress and differential-susceptibility propositions, using data on the moderating effect of DRD4 with regard to the effect of childcare quality on children's social functioning. **Results:** Results show that (a) the new approach detects interactions that the traditional one does not; (b) the discerned GXE fit the differential-susceptibility model better than the diathesis-stress one; and (c) a strong rather than weak version of differential susceptibility is empirically supported. **Conclusion:** The new method better fits the theoretical 'glove' to the empirical 'hand,' raising the prospect that some failures to replicate GXE results may derive from standard statistical approaches being less than ideal. **Keywords:** Gene-environment interaction, diathesis-stress, differential susceptibility, child care, DRD4.

Introduction

For the past decade virtually all psychiatric gene-environment (GXE) interaction research has been guided, implicitly if not explicitly, by the diathesis-stress model of environmental action (Zubin & Spring, 1977) which stipulates that some individuals are more susceptible to the negative consequences of adverse experiences than others (Burmeister, McInnis & Zollner, 2008). The expectation, then, is that individuals carrying 'risk alleles' or 'vulnerability genes' (e.g., short 5-HTTLPR, DRD4-7 repeat) will function more poorly than those with different genotypes under conditions of contextual adversity (e.g., child maltreatment, negative life events).

Yet even in the face of guiding theory and strong predictions, the statistical methods used to evaluate GXE are exploratory in character. Using hierarchical regression and/or ANOVA techniques, the interaction of environmental factor and genetic moderator is evaluated after taking into account the main effects of each. Notably, this approach does not formally evaluate whether a detected interaction is consistent with the implicit or explicit diathesis-stress theorizing that motivated the research. This is why, following the detection of a significant interaction, investigators routinely (a) conduct additional regression analyses affording comparisons of simple slopes linking predictor to outcome separately for the putative vulnerable and resilient subgroups, or (b)

compare means of subgroups defined in terms of genotype and environment (i.e., vulnerable genotype/risky environment, vulnerable genotype/benign environment, non-vulnerable genotype/risky environment, non-vulnerable genotype/benign environment). To determine whether findings are consistent with diathesis-stress, visual inspection of plotted simple slopes with or without additional comparison of group means is undertaken.

Being exploratory in nature, these approaches completely disregard the a priori predictions on which they are based. Given well-known statistical challenges to testing interactions (Aiken & West, 1991; McClelland & Judd, 1993), including need for larger sample sizes than when evaluating main effects, statistical methodology should not further undermine statistical power and thus the detection of interactions. This would seem especially the case in studies of GXE interaction given widespread appreciation that even significant main effects of candidate genes – in genotype-phenotype studies – are likely to account for only a very small amount of variance in almost all phenotypes of interest to developmental and behavioral scientists.

A perhaps even more significant limitation of standard *exploratory* approaches is that they do not provide a *direct means of testing competing predictions* derived from alternative theoretical frameworks. This latter limitation is particularly acute now that a theoretical alternative to the diathesis-stress model has been proposed (Belsky & Pluess, 2009; Ellis, Boyce, Belsky, Bakermans-

Conflict of interest statement: No conflicts declared.

Kranenburg, and van IJzendoorn, 2011) and applied to the study of GXE interactions (Belsky et al., 2009), first by Bakermans-Kranenburg and van IJzendoorn (2006), but more recently by many others (for summary, see Belsky & Pluess, in press). This differential-susceptibility framework stipulates that some individuals are more susceptible not just to negative environmental influences but to positive ones as well (i.e., ‘for better *and* for worse’, Belsky, Bakermans-Kranenburg, and van IJzendoorn, 2007). For this reason, Belsky et al. (2009) suggested that some ‘vulnerability genes’ might be better conceptualized as ‘plasticity genes.’

This alternative to the diathesis-stress model of environmental action led Belsky et al. (Belsky, Bakermans-Kranenburg, et al. 2007) to delineate statistical criteria for distinguishing GXE (and other person-X-environment) interactions that reflected differential-susceptibility from ones reflecting diathesis-stress. Kochanska, Kim, Barry and Philibert (2011) proposed adding to Belsky, Bakermans-Kranenburg, et al. (2007) more or less traditional regression approach (with some modifications) the evaluation of regions of significance (Aiken & West, 1991; Preacher, Curran & Bauer, 2006). And more recently Roisman et al. (2012) have advanced even more rigorous statistical criteria for testing differential susceptibility. In all cases, however, these efforts require that an exploratory test of GXE (or person-X-environment) interaction prove significant and thus function as a ‘screen’ before different forms of interaction are evaluated.

To address limitations of such exploratory approaches to testing GXE interactions, we present a confirmatory method developed by Widaman et al. (2012) that explicitly evaluates alternative theoretical models while maximizing statistical power by aligning analyses with hypotheses of interest. Of note is that the approach illustrated in this paper is by no means restricted to testing either GXE, as it will be here, or any other person-X-environment interaction; indeed, it can be used in virtually any and all tests of statistical interactions in which there are competing hypotheses as to the form the interaction might take (e.g., father involvement X paternal sensitivity in predicting infant-father attachment security; Brown, Mangelsdorf & Neff, 2012). The method systematically varies the number of parameters included in a regression equation in order to contrast alternative conceptual frameworks, most notably parameters specifying where on the continuum of environmental measurement regression lines reflecting the association between environmental predictor and outcome for different genetic subgroups will cross. Whereas diathesis-stress theorizing predicts an *ordinal interaction* with regression lines crossing at or above the most positive observed value for the measured environment (Figure 1A), differential-susceptibility theorizing predicts a *disordinal* or cross-over interaction, with regression lines

crossing somewhere within the range of values of the measured environment (Figure 1B). Comparisons of proportion of variance explained by alternate regression models then determine whether one model provides a better fit to the data and therefore a better explanation for the observed phenomenon than do other models.

The confirmatory approach to testing GXE interactions is buttressed by prior work showing that reliance on omnibus tests from exploratory methods may often obscure significant findings aligned with a priori hypotheses. Hale (1977), Rosnow and Rosenthal (1989), and Burchinal and Clarke-Stewart (2007) offered cogent examples of the problematic outcomes of omnibus tests. If an omnibus test has more than one *df*, essentially nil values associated with one or more *df* can combine with significant values associated with other *df* to result in a negatively biased omnibus test value. To avoid the negative bias, a confirmatory approach to testing a priori hypotheses is recommended, and the Widaman et al. (2012) approach does so for GXE research.

To illustrate this method, we present analyses of children’s social competence and behavior problems that test an interaction between a genetic polymorphism, the 7-repeat (7R) allele of the dopamine receptor D4 (*DRD4*), and exposure to varying quality of childcare, using data from the large-scale NICHD Study of Early Child Care (NICHD Early Child Care Research Network., 2005a). This is a particularly interesting issue because (a) prior analyses of these

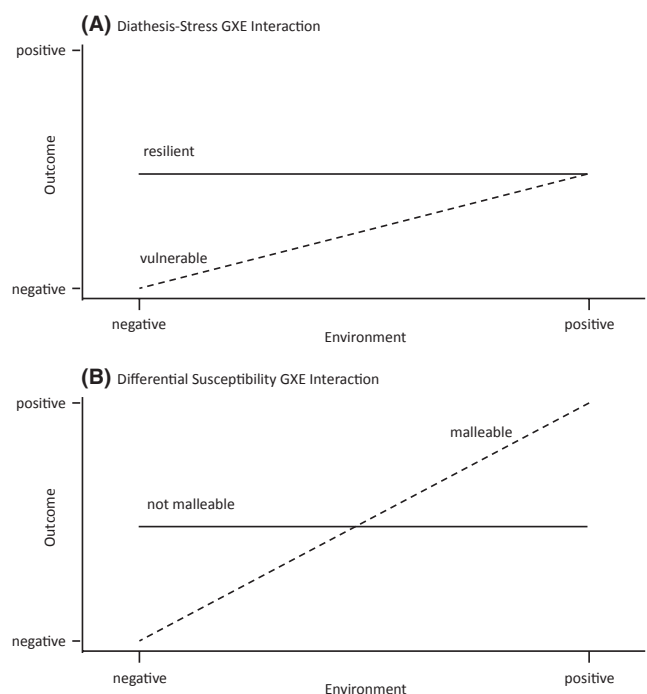


Figure 1 Plots of idealized results under (A) the diathesis-stress GXE model and (B) the differential susceptibility GXE model [figures based on Bakermans-Kranenburg and van IJzendoorn (2007)]

data failed to reveal main effects of childcare quality (Belsky, Vandell, et al. 2007; NICHD Early Child Care Research Network., 2005b); (b) a 10-study meta-analysis showed that variation in genes related to dopamine signaling in the brain influence children's sensitivity to both sensitive/responsive and harsh/unresponsive parenting (Bakermans-Kranenburg & van IJzendoorn, 2011); (c) experimental evidence indicated that children carrying the DRD4-7R allele benefited more than others from an intervention fostering skilled parenting (Bakermans-Kranenburg, van IJzendoorn, Pijlman, Mesman & Juffer, 2008); and (d) the measure of quality of childcare used in the current inquiry is similar to parenting measures in the just-cited work. We thus predicted that the GXE interaction would take the form of differential-susceptibility: children carrying DRD4-7R would exhibit (a) the most social competence and fewest behavior problems under conditions of high-quality childcare and (b) the least social competence and most behavior problems under conditions of low-quality childcare.

Therefore, we conducted a comparative analysis of a number of alternative GXE models. These alternatives represent 'weak' and 'strong' versions of diathesis-stress and differential-susceptibility propositions. The strong(er) version of each framework presumes that children not carrying DRD4-7R are not affected at all by quality of childcare, whereas the weak(er) version of each presumes that those without DRD4-7R are affected by childcare quality, but to a lesser degree than those carrying DRD4-7R. Based on prior research (e.g., Bakermans-Kranenburg & van IJzendoorn, 2011) and theoretical syntheses of research (e.g., Belsky & Pluess, 2009), we predicted that the strong differential-susceptibility model would provide the best, most parsimonious representation of the data for both social skills and behavior problems. That is, we predicted that the GXE interaction would resemble Figure 1B, with the slope for the non-malleable group fixed at zero. We would reject the strong differential-susceptibility model if the weak differential-susceptibility model provided improved fit to the data or if either the weak or strong diathesis-stress models provided comparable fit to the data with still fewer parameter estimates.

Methods

Participants

The NICHD Study of Early Child Care recruited 1364 families through hospital visits shortly after the birth of a child in 1991 at 10 US locations (for detailed description of recruitment procedures and sample characteristics see NICHD Early Child Care Research Network., 2001). The current analysis includes 441 cases on whom genetic data were available. Informed consent was secured at each

data collection delineated below, with a special consent form for the collection of saliva for purposes of assaying genes.

Measures

Genetic analyses. DNA were obtained from buccal cheek cells when children were 15 years of age. Children with at least one 7R allele ($n = 95$; 21.5%) were distinguished from those with both alleles shorter than 7R ($n = 346$; 78.5%).

Childcare quality. Quality of care was measured using observational assessments of caregiver behavior conducted in childcare arrangement at ages 6, 15, 24, 36, and 54 months. Unconditional linear growth curves were fit across repeated composite ratings of caregiver behavior and individual intercepts were estimated reflecting the degree to which caregiving was sensitive, stimulating, positive and neither neglectful nor negative in character. The sample mean, $M = 2.83$ ($SD = 0.24$), and median, $Md = 2.82$, were near the middle of the range (2.10–3.38) observed in the study.

Social competence. Teacher-reported social competence in 1st grade (~6 years old) was assessed with the Social Skills Questionnaire from the Social Skills Rating System (Gresham & Elliott, 1990). Raw scores were standardized. The sample mean, $M = 104.23$ ($SD = 13.19$), and median, $Md = 104.00$, were near population values.

Behavior problems. The Child Behavior Checklist Teacher Report Form (Achenbach, 1991) evaluated externalizing behavior in 1st grade. Raw scores were standardized, based on normative data for children of the same age. The sample mean, $M = 49.96$ ($SD = 8.40$), and median, $Md = 49.00$, were close to population values.

Data analysis

Exploratory and confirmatory analytic approaches were used.

Exploratory. Childcare quality was the environmental variable X , and a dummy variable D demarcated gene group (0 = absence of 7R, and 1 = presence of 7R). The standard multiple regression model can be written as:

$$Y = B_0 + B_1X + B_2D + B_3(X \cdot D) + E \quad (1)$$

where Y is the dependent variable, B_0 the intercept, B_1 and B_2 regression slopes for main effects of environment (X) and genes (D), respectively, B_3 the regression coefficient for the product variable ($X \cdot D$) and represents the difference in slope on X for the '7R present' group relative to the '7R absent' group, and E is a stochastic error term.

Equation 1 is fit once excluding the product term ($X \cdot D$), testing partial (or simultaneous) main effects (Model 1). With the product term added (Model 2), a significant increase in the squared multiple correlation, R^2 , provides evidence for a GXE interaction. To determine whether the GXE interaction is consistent with diathesis-stress or differential-susceptibility models, simple slopes for the two groups can be computed and plotted, and a point estimate of the cross-over point C can be calculated as $C = -B_2/B_3$ (Aiken & West, 1991).

Confirmatory. Following Widaman et al. (2012), we re-parameterized the regression model for a dichotomous gene polymorphism, allowing a-priori testing of alternative forms of the GXE interaction, as:

$$Y : \begin{cases} D = 0 & Y = B_0 + B_1(X - C) + E \\ D = 1 & Y = B_0 + B_3(X - C) + E \end{cases} \quad (2)$$

Here C is the point on X at which the slopes for the two gene groups cross. If the point estimate of C is within the range of values on X observed in a study, the interaction tested is disordinal, consistent with differential-susceptibility. Conversely, if the point estimate of C is greater than or equal to the most positive point on X in the study, the interaction is ordinal, reflecting diathesis-stress. Importantly, because C is a parameter in the model, the point estimate of C is accompanied by a SE , so a confidence interval (CI) for the cross-over point can be calculated, giving more information on the likely range for the population value of this key parameter.

Equation 2 is a four-parameter equation (i.e., B_0 , B_1 , B_3 , and C). C is the point on X at which predicted values for the two groups cross-over or converge, and B_0 is the estimated Y score at the cross-over point. B_1 is the slope for the environmental variable X for the non-7R group ($D = 0$), and B_3 is the comparable slope for the 7R group ($D = 1$). If C falls within the range of X and the GXE interaction is disordinal and if the slope for the non-7R group is fixed at zero (i.e., $B_1 = 0$), the model in Equation 2 is consistent with strong differential-susceptibility in which the non-7R group is unaffected by the environment. This is the model we hypothesized a priori would hold for both child outcome variables we analyzed, and we term this Model 3a.

But, patterns of fixed and free parameters in Equation 2 can be re-specified to represent other a priori models that might explain the data virtually as well as or better than Model 3a. For example, relaxing the constraint that $B_1 = 0$ leads to Model 3b, the weak differential susceptibility model. If the slope for the non-7R group differs significantly from zero, Model 3b should explain significantly more variance than does Model 3a. In such a situation, one has a statistical justification to reject the more parsimonious Model 3a in favor of the better fitting

Model 3b. Or, if one retained the constraint that the non-7R group is unaffected by the environment and added a constraint that the cross-over point fell at the highest value for the environment observed in the sample, or $C = \max(X)$, the model would be consistent with predictions under the strong diathesis-stress model (Model 3c). Finally, if one allowed both B_1 and B_3 to differ from zero, but fixed the cross-over point at the highest value observed in the sample, or $C = \max(X)$, the model would conform to the weak diathesis-stress model, Model 3d.

Models 3b and 3c represent models with one more and one less parameter, respectively, than our preferred Model 3a, so can be tested to determine if these two models lead to a significant increase or decrease in explained variance, respectively, when compared to Model 3a. In addition, all three restricted models (Models 3a, 3c, and 3d) are nested within Model 3b, so can be compared against the fit of Model 3b. Strong-differential-susceptibility and weak-diathesis-stress models (3a, 3d) each have three parameter estimates, so cannot be tested statistically against one another. But, as competing three-parameter models, the one with the higher R^2 is the one that provides a better representation of the data. We also used the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to evaluate model fit. For both the AIC and BIC, lower values indicate better fit of a model to data. AIC and BIC values are especially useful for evaluating non-nested models (e.g., Models 3a and 3b), which cannot be compared using statistical tests.

The use of the re-parameterized Equation 2 thus affords direct determination of whether the GXE interaction is disordinal, as posited by differential-susceptibility, or ordinal, consistent with diathesis-stress. If differential susceptibility best fits the data, point and interval estimates of the cross-over point are obtained. Further, it provides a means of determining whether a strong or weak version of a model best fits the data. Widaman et al. (2012) provide programming details regarding how to fit these confirmatory models in SAS, SPSS, and R packages; they are available upon request.

Results

First, we conduct tests of GXE interactions for social competence and behavior problems using the standard exploratory approach to analyses. Then, we contrast, for each outcome, strong and weak forms of the differential-susceptibility and diathesis-stress models to determine which provided the best, most parsimonious fit to the data.

Standard exploratory analysis

Social competence. Table 1 shows that Model 1, with main effects of environment and gene group,

was fit to data on social competence and had an $R^2 = .023$, with environment effect significant, $B_1 = 7.12$ ($SE = 2.57$), $p < .01$, but not the gene main effect, $p = .17$. Adding the GXE interaction to the equation, Model 2, resulted in an increase in R^2 , or ΔR^2 , of .0175, which was significant, $p < .01$, as was the coefficient for the GXE product term itself, $B_3 = 17.51$ ($SE = 6.23$), $p = .005$. Critically, parameter estimates for Model 2 do not afford a direct determination of the nature of the interaction (i.e., ordinal vs. disordinal), for which follow-up testing is usually undertaken.

Behavior problems. Table 2 shows that Model 1 produced a non-significant R^2 of .0117, $p = .08$, with neither partial main effect proving significant ($ps > .09$). The GXE interaction produced a non-significant $\Delta R^2 = .0057$, $p = .11$ (Model 2). These results of the standard approach yield the conclusion of 'no significant GXE interaction,' thereby terminating consideration of genetic moderation of an environmental effect.

Strong vs. weak differential susceptibility

Versions of Equation 2 reflecting strong and weak differential susceptibility were next fit, first for social competence and then for behavior problems.

Social competence. The strong differential-susceptibility model (3a) stipulates that children without the

7R allele would be unaffected by quality of child care, but that children with the 7R allele would be positively affected by child care quality. As shown in Table 1, Model 3a explained a significant amount of variance in social competence, $R^2 = .0374$, $p < .0005$. The estimated cross-over point C fell close to the sample mean on child care, $\hat{C} = 2.75$ ($SE = 0.08$), and the CI fell completely within the range of child care, 95% CI of $\hat{C}[2.59, 2.91]$. Thus, Model 3a provides strong support for the strong differential susceptibility model for social competence.

Relaxing the constraint that $B_1 = 0$ leads to Model 3b, the weak differential-susceptibility model. As seen in Table 1, Model 3b explained a very small and non-significant amount of additional variance over that explained by Model 3a, $\Delta R^2 = .0032$, $p = .23$. Therefore, we find no statistical basis for rejecting the parsimonious Model 3a in favor of the more highly parameterized Model 3b, lending support in favor of the strong differential susceptibility model as a more optimal representation of the data.

Behavior problems. Contrary to results using the typical exploratory approach to analyses, the fit of the strong differential susceptibility model, Model 3a, to the behavioral problems data yielded a significant $R^2 = .0167$, $p = .025$ (see Table 2). The cross-over point C was estimated fairly close to the sample mean on child care, $\hat{C} = 2.64$ ($SE = 0.17$), and the CI fell entirely within the range of child care, 95% CI of $\hat{C}[2.30, 2.98]$. Thus, Model 3a provides support for

Table 1 Results for alternate regression models for social competence

Standard parameterization			Re-parameterized regression equation				
Parameter	Gene (G) and environment (E) main effects: Model 1	Main effects and GXE interaction: Model 2	Parameter	Differential susceptibility		Diathesis-Stress	
				Strong: Model 3a	Weak: Model 3b	Strong: Model 3c	Weak: Model 3d ^a
B_0	84.48 (7.26)	94.92 (8.11)	B_0	104.5 (0.68)	104.3 (0.91)	105.1 (0.68)	109.1 (1.54)
B_1	7.12 (2.57)	3.41 (2.87)	B_1	0.00 (-) ^b	3.41 (2.87)	0.00 (-) ^b	7.45 (2.59)
B_2	2.04 (1.48)	-47.95 (17.9)	C	2.75 (0.08)	2.74 (0.09)	3.38 (-) ^b	3.38 (-) ^b
B_3	-	17.51 (6.23)	B_3	20.92 (5.54)	20.92 (5.53)	0.73 (2.58)	7.16 (3.40)
R^2	.0231	.0406	R^2	.0374	.0406	.0002	.0189
F	5.15	6.12	F	8.46	6.12	0.08	4.19
df	2, 435	3, 434	df	2, 435	3, 434	1, 436	2, 435
p	.006	<.001	p	<.0005	<.001	.77	.02
F vs. 1	-	7.90	F vs. 3a	-	1.41	16.81	-
df	-	1, 434	df	-	1, 434	1, 435	-
p	-	.005	p	-	.23	<.0001	-
			F vs. 3b	1.41	-	9.80	9.13
			df	1, 434	-	1, 434	2, 434
			p	.23	-	.002	<.0001
AIC	2232.4	2226.6	AIC	2225.9	2226.6	2240.6	2234.3
BIC	2244.7	2242.9	BIC	2238.2	2242.9	2248.8	2246.6

AIC, Akaike information criterion; BIC, Bayesian information criterion.

Tabled values are parameter estimates, with their standard errors in parentheses. F vs. 1 stands for an F test of the difference in R^2 for Model 2 versus Model 1. F vs. 3a stands for an F test of the difference in R^2 for a given model versus Model 3a.

^aAlthough not reported in the text, results for this model are provided for completeness.

^bParameter fixed at reported value; SE is not applicable, so is listed as (-).

Table 2 Results for alternate regression models for behavior problems

Standard parameterization			Re-parameterized regression equation				
Parameter	Gene (G) and environment (E)		Parameter	Differential susceptibility		Diathesis-stress	
	main effects: Model 1	Main effects and GXE interaction: Model 2		Strong: Model 3a	Weak: Model 3b	Strong: Model 3c	Weak: Model 3d ^a
B_0	57.11 (4.78)	53.24 (5.36)	B_0	50.36 (0.45)	50.57 (0.72)	50.10 (0.45)	48.59 (1.02)
B_1	-2.40 (1.68)	-1.02 (1.89)	B_1	0.00 (-) ^b	-1.02 (1.89)	0.00 (-) ^b	-2.82 (1.70)
B_2	-1.62 (0.98)	17.18 (11.9)	C	2.64 (0.17)	2.61 (0.21)	3.38 (-) ^b	3.38 (-) ^b
B_3	-	-6.59 (4.14)	B_3	-7.61 (3.68)	-7.61 (3.69)	1.08 (1.70)	-1.36 (2.24)
R^2	.0117	.0174	R^2	.0167	.0174	.0009	.0072
F	2.59	2.57	F	3.72	2.57	0.40	1.58
df	2, 438	3, 437	df	2, 438	3, 437	1, 439	2, 438
p	.08	.054	p	.025	.054	.52	.21
F vs. 1	-	2.53	F vs. 3a	-	0.29	7.04	-
df	-	1, 437	df	-	1, 437	1, 438	-
p	-	.11	p	-	.59	.008	-
			F vs. 3b	0.29	-	3.65	4.53
			df	1, 437	-	2, 437	1, 437
			p	.59	-	.03	.03
AIC	1884.3	1883.7	AIC	1882.0	1883.7	1887.1	1886.3
BIC	1896.5	1900.1	BIC	1894.3	1900.1	1895.2	1898.6

AIC, akaike information criterion; BIC, Bayesian information criterion.

Tabled values are parameter estimates, with their standard errors in parentheses. F vs. 1 stands for an F test of the difference in R^2 for Model 2 versus Model 1. F vs. 3a and F vs. 3b stand for F tests of the difference in R^2 for a given model versus Model 3a and Model 3b, respectively.

^aAlthough not reported in the text, results for this model are provided for completeness.

^bParameter fixed at reported value; SE is not applicable, so is listed as (-).

the strong differential susceptibility model for the behavior problems data.

Once again, relaxing the constraint that $B_1 = 0$ leads to Model 3b, the weak differential-susceptibility model. As seen in Table 2, Model 3b explained a very small and non-significant amount of added variance over that explained by Model 3a, $\Delta R^2 = .0007$, $p = .59$. In line with results for social competence, we find no statistical basis for rejecting the parsimonious Model 3a in favor of the more highly parameterized Model 3b, lending support in favor of the strong differential susceptibility model as the optimal representation of the behavior problems data.

Differential susceptibility vs. diathesis-stress

Given the better fit of the strong differential susceptibility model relative to the weak version of the model, we proceeded to fit strong and weak versions of the diathesis-stress model to allow comparative evaluation of the fit of these models.

Social competence. The strong diathesis-stress model is nested within the strong differential-susceptibility model, by fixing the cross-over point to be at the maximum value of X observed in the study; here, X was child care quality, and the maximal value observed was 3.38. As shown in Table 1, this model, Model 3c, had a very small level of explained

variance, $R^2 = .0002$, $p = .77$, and fit significantly worse than Model 3a, $p < .0001$.

Relaxing the constraint that $B_1 = 0$ leads to Model 3d, the weak diathesis-stress model. As seen in Table 1, Model 3d had a modest increase in explained variance over Model 3c, but still had a relatively low level of explained variance, $R^2 = .0189$, $p = .02$. As noted above, Models 3d and 3a cannot be compared statistically because they have the same number of parameter estimates. However, Model 3a, the strong differential-susceptibility model, explained about twice as much variance as Model 3d with the same number of estimates and had substantially lower values of both AIC and BIC than did Model 3d. Further, Model 3a had the lowest (i.e., best) levels of AIC and BIC of all six regression models fit to the data, lending clear support to Model 3a as the preferred representation of data.

Problem behavior. The pattern of model comparisons for problem behavior was very similar to that for social competence. Specifically, as shown in Table 2, the strong diathesis-stress model, Model 3c, had a very small level of explained variance, $R^2 = .0009$, $p = .52$, and fit significantly worse than Model 3a, $p < .05$.

Relaxing the constraint that $B_1 = 0$ resulted in Model 3d, the weak diathesis-stress model. As seen in Table 2, Model 3d had a modest increase in explained variance over Model 3c, but still had a

relatively low level of explained variance, $R^2 = .0072$, $p = .21$. As noted above, Models 3d and 3a cannot be compared statistically because they have the same number of parameter estimates. However, Model 3a, the strong differential-susceptibility model, explained over twice as much variance as Model 3d with the same number of estimates. Once again, the AIC and BIC values for Model 3a were noticeably lower than comparable values for Model 3d as well as for the remaining models fit to the data. This pattern of results provides clear and consistent support for Model 3a as the preferred representation of data.

Plotting predicted values under differential susceptibility

Plots of predicted values provide a very useful way of displaying and interpreting data with an interaction. The plot of predicted social competence scores is shown in Figure 2A, where the malleable group does have systematically lower levels of competence than the non-malleable group below the cross-over point. However, above the cross-over point, the malleable group shows consistently higher performance as a function of higher levels of child care quality.

The plot of predicted values for behavior problems is shown in Figure 2B. Below the cross-over point, the malleable group exhibits worse (i.e., higher) levels of behavior problems as a function of lower levels of child care quality. But, above the cross-over point, the malleable group shows improvements in behavior problems with higher levels of quality care.

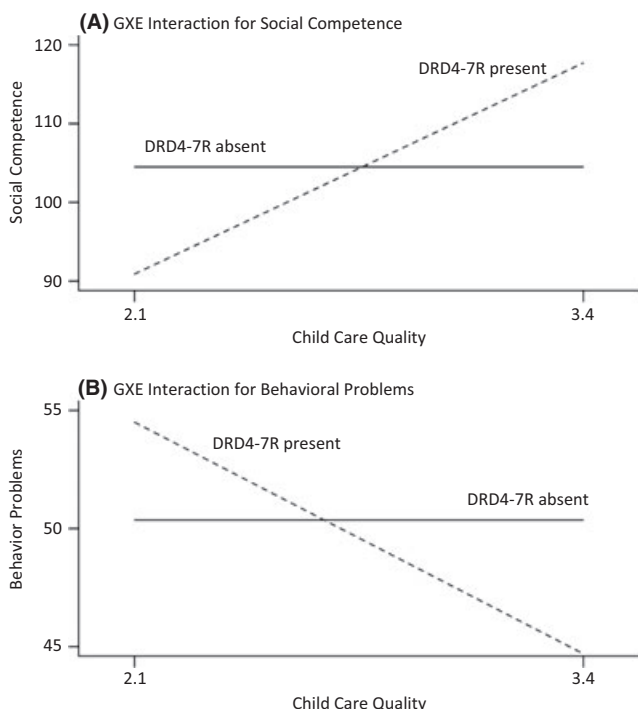


Figure 2 Plots of predicted values as a function of child care quality under the strong differential susceptibility GXE model for (A) social competence and (B) behavior problems [figures based on Bakermans-Kranenburg and van IJzendoorn (2007)]

Discussion

Gene-environment interaction research has grown dramatically over the past decade, as scholars have discussed and debated whether and how the genetic make-up of individuals moderates their susceptibility to environmental influences (Risch et al., 2009; Uher & McGuffin, 2010). Most of this work has been informed by the diathesis-stress view that some individuals are genetically more vulnerable, succumbing to adverse environmental effects, than are others, though the differential-susceptibility perspective offers an alternative, evolutionary-inspired view: those most susceptible to negative influences are simultaneously most likely to benefit from positive ones. Either way, it remains the case that all tests of GXE have been exploratory – and thereby conservative – in nature, rather than optimally sensitive to the theoretical propositions guiding the research.

Here we illustrated an alternative approach in which GXE data are evaluated in terms of competing theoretical propositions. Rather than requiring a significant GXE interaction derived from an exploratory analysis to serve as the gateway for determining the form of the interaction, we show that interactions having different forms, inspired by alternative theoretical frameworks, can be directly tested – and compared. The fact that the findings reported herein indicate, consistent with other work on the effects of children's caregiving environments (Bakermans-Kranenburg & van IJzendoorn, 2011; Bakermans-Kranenburg et al., 2008), that children carrying the DRD4-7R allele prove sensitive to higher and lower quality childcare in terms of their social adjustment, whereas those lacking this allele prove insensitive to childcare quality (see Figure 2A,B), is no doubt interesting. The main contribution of this analysis, however, is the demonstration that the Widaman et al. (2012) confirmatory approach for evaluating interactions can be used to address alternative GXE hypotheses specifically and directly contrast the fit of models that regard a particular genotype as a 'risk' allele versus a 'plasticity' allele.

However, the Widaman et al. (2012) confirmatory approach is not restricted to contrasting such 'vulnerability' and 'susceptibility' hypotheses. It is, in fact, a flexible approach that can be fit to test GXE hypotheses other than diathesis-stress and differential susceptibility. For example, Pluess and Belsky (2012) recently introduced a new concept – *Vantage Sensitivity* – which refers to individual differences in response to exclusively positive experiences, as reflected in the 'bright side' of differential susceptibility. According to this framework, individuals who benefit disproportionately from positive supportive experiences (e.g., psychotherapy, positive life events) manifest 'vantage sensitivity', whereas individuals failing to respond positively to the same experience are referred to as showing 'vantage resistance' – with

no differences between individuals expected in response to negative experiences. This vantage-sensitivity-interaction hypothesis can easily be tested and contrasted with other GXE hypotheses – using the Widaman et al. (2012) approach – by fixing the cross-over point to the minimum end of the environmental variable and, in case of strong vantage sensitivity, constraining the slope of the vantage resistant group to zero. Also important to note is that Widaman et al. (2012) showed that their approach works when genetic variables have more than just two categories or are measured continuously. Indeed, it is applicable not just to the study of GXE, but to investigation of all sorts of interactional effects.

One cannot but wonder whether some of the debate surrounding the existence and detection of GXE interactions in the extant literature may have proven different if methodological ‘gloves’ had been better fit to empirical ‘hands.’ Consider the possibility, for example, that even if some of the non-replication of the 5HTTLPR-X-life-stress interaction predicting depression resulted from limited measurement (Uher & McGuffin, 2010), some of it may also have been a product of limited statistical analysis. Perhaps some of what are considered null findings would mirror what we found regarding problem behavior: When tested in the traditional, exploratory manner, no GXE interaction is discerned; yet when tested in a theoretically-informed manner, evidence of GXE interaction emerges. Certainly some will contend that if GXE interactions are so sensitive to subtleties of analysis, then they are insufficiently robust to be of interest, much less importance. Others will certainly disagree. This is a debate worth having, but to have well informed debate empirical tests themselves should fit the hypotheses being tested.

Consider further the fact that so many GXE designs measure neither a full range of environments nor a full range of psychological functioning, from

positive to negative (Belsky et al., 2009). Instead, some adverse condition (e.g., negative life events) is contrasted with its absence and some pathological outcome (e.g., depression) is evaluated versus its absence. When coupled with the perhaps all-too-conservative, exploratory approach to evaluating GXE, there would seem to be grounds for questioning the conclusion of Risch et al. (2009), based on the meta-analysis of a single GXE interaction and outcome, dismissing the entire GXE paradigm. For too long, perhaps, the debate about GXE interaction has focused on sample size and measurements. We contend that insufficient attention has been paid to the theoretical foundations of GXE inquiry (e.g., diathesis-stress vs. differential susceptibility vs. vantage sensitivity) (Belsky & Pluess, 2009; Ellis, et al. 2011; Pluess & Belsky, 2012), the range of environments and outcomes measured (Belsky, Bakermans-Kranenburg, et al. (2007), and the statistical procedures adopted for evaluating the important notion that individuals differ, for genetic reasons, in their susceptibility to environmental effects.

Acknowledgements

Preparation of this paper was supported in part by the Robert M. and Natalie Reid Dorn Professorship, UC Davis (Jay Belsky), grant HD 064687 from the National Institute of Child Health and Human Development (Rand D. Conger, PI) and grant DA 017902 from the National Institute of Drug Abuse and the National Institute of Alcohol Abuse and Alcoholism (Rand D. Conger, Richard W. Robins, and Keith F. Widaman, Joint PIs).

Correspondence

Jay Belsky, Department of Human Ecology, Program in Human Development and Family Studies, University of California, Davis, Davis, CA 95616, USA; Email: jbelsky@ucdavis.edu

Key points

- Extensive research highlights gene-X-environment interaction (GXE) when it comes to accounting for variation in the extent to which individuals are affected by particular developmental experiences and/or environmental exposures.
- Virtually all research on GXE relies on exploratory methods of statistical analysis, even though most such work is based on explicit – and directional – predictions, typically consistent with diathesis-stress theorizing. Because there is now a well-recognized alternative to this prevailing model of environmental action, one known as differential susceptibility, there is a need for statistical methods that test competing models, which is what is offered here.
- Discovering ‘what works for whom’ is a central goal of clinically relevant research. Methods such as those presented can advance this cause by determining whether certain individuals are more susceptible to both positive and negative developmental experiences and environmental exposures or simply just more vulnerable to adversity, as long presumed.

References

- Achenbach, T. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile*. Burlington, VT: Author.
- Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Bakermans-Kranenburg, M.J., & van IJzendoorn, M.H. (2006). Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Developmental Psychobiology*, *48*, 406–409.
- Bakermans-Kranenburg, M.J., & van IJzendoorn, M.H. (2007). Research review: Genetic vulnerability or differential susceptibility in child development: the case of attachment. *Journal of Child Psychology and Psychiatry*, *48*, 1160–1173.
- Bakermans-Kranenburg, M.J., & van IJzendoorn, M.H. (2011). Differential susceptibility to rearing environment depending on dopamine-related genes: New evidence and a meta-analysis. *Development and Psychopathology*, *23*, 39–52.
- Bakermans-Kranenburg, M.J., van IJzendoorn, M.H., Pijlman, F.T., Mesman, J., & Juffer, F. (2008). Experimental evidence for differential susceptibility: Dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Developmental Psychology*, *44*, 293–300.
- 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.
- Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry*, *14*, 746–754.
- Belsky, J., & Pluess, M. (in press). Beyond risk, resilience and dysregulation: Phenotypic plasticity and human development. *Development and Psychopathology*.
- Belsky, J., & Pluess, M. (2009). Beyond diathesis-stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*, 885–908.
- Belsky, J., Vandell, D.L., Burchinal, M., Clarke-Stewart, K.A., McCartney, K., & Owen, M.T. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701.
- Brown, G.L., Mangelsdorf, S.C., & Neff, C. (2012). Father involvement, paternal sensitivity, and father-child attachment security in the first 3 years. *Journal of Family Psychology*, *26*, 421–430.
- Burchinal, M.R., & Clarke-Stewart, K.A. (2007). Maternal employment and child cognitive outcomes: The importance of analytic approach. *Developmental Psychology*, *43*, 1140–1155.
- Burmeister, M., McInnis, M.G., & Zollner, S. (2008). Psychiatric genetics: Progress amid controversy. *Nature Reviews. Genetics*, *9*, 527–540.
- 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.
- Gresham, F., & Elliott, S. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Hale, G.A. (1977). On use of ANOVA in developmental research. *Child Development*, *48*, 1101–1106.
- Kochanska, G., Kim, S., Barry, R.A., & Philibert, R.A. (2011). Children's genotypes interact with maternal responsive care in predicting children's competence: Diathesis-stress or differential susceptibility? *Development and Psychopathology*, *23*, 605–616.
- McClelland, G.H., & Judd, C.M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.
- NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Applied Developmental Psychology*, *22*, 457–492.
- NICHD Early Child Care Research Network. (2005a). *Child care and child development: Results of the NICHD study of early child care and youth development*. New York: Guilford Press.
- NICHD Early Child Care Research Network. (2005b). Early child care and children's development in the primary grades: Follow-up results from the NICHD study of early child care. *American Educational Research Journal*, *42*, 537–570.
- Pluess, M., & Belsky, J. (2012). Vantage sensitivity: Individual differences in response to positive experiences. *Psychological Bulletin*, in press.
- Preacher, K.J., Curran, P.J., & Bauer, D.J. (2006). Computational tools for probing interactions in multiple linear regressions, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448.
- Risch, N., Herrell, R., Lehner, T., Liang, K.Y., Eaves, L., Hoh, J., ... & Merikangas, K.R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *JAMA*, *301*, 2462–2471.
- Roisman, G.I., Newman, D.A., Fraley, R.C., Haltigan, J.D., Groh, A.M., & Haydon, K.C. (2012). Distinguishing differential susceptibility from diathesis-stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, *24*, 389–409.
- Rosnow, R.L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.
- Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Mol Psychiatry*, *15*, 18–22.
- 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.
- Zubin, J., & Spring, B. (1977). Vulnerability – A new view of schizophrenia. *Journal of Abnormal Psychology*, *86*, 103–126.

Accepted for publication: 4 February 2013

Published online: 25 April 2013