CHILD DEVELOPMENT

Child Development, November/December 2015, Volume 86, Number 6, Pages 1999–2016

Analysis of Behavioral and Emotional Problems in Children Highlights the Role of Genotype × Environment Interaction

Dylan Molenaar
University of Amsterdam

Christel Middeldorp
VU University Amsterdam and Neuroscience Campus Amsterdam (NCA)

Toos van Beijsterveldt
VU University Amsterdam

Dorret I. Boomsma
VU University Amsterdam and Neuroscience Campus Amsterdam (NCA) and VU University Medical Center

This study tested for Genotype × Environment (G × E) interaction on behavioral and emotional problems in children using new methods that do not require identification of candidate genes or environments, can distinguish between interaction with shared and unique environment, and are insensitive to scale effects. Parental ratings of problem behavior from 14,755 twin pairs (5.3 years, *SD* = 0.22) indicated G × E interaction on emotional liability, social isolation, aggression, attention problems, dependency, anxiety, and physical coordination. Environmental influences increased in children who were genetically more predisposed to problem behavior, with ~20% of the variance due to G × E interaction (8% for anxiety to 37% for attention problems). Ignoring G × E interaction does not greatly bias heritability estimates, but it does offer a comprehensive model of the etiology for childhood problems.

The heritability of many aspects of emotional and behavioral problems in children is substantial. The extent to which the genome contributes to differences between children in emotional and behavioral problems has been estimated between 0.4 and 0.8 for a wide range of problems (e.g., Faraone et al., 2005; Geschwind, 2009; Gregory & Eley, 2007; Kan et al., 2013; Nivard et al., 2014; Rhee & Waldman, 2002; Rice, 2009). There is, however, an ongoing debate about the interplay between genotype and environment, involving questions such as "Does the effect of the environment on the development of emotional and behavioral problems depend on the underlying genotype?" and if yes, "to what extent" and "in what direction?" The lack of knowledge about Genotype × Environment (G × E) interaction limits our understanding of how genetic and environmental risk mechanisms underlie child development (Plomin & Rutter, 1998).

Different Theoretical Models

Different theoretical models have been proposed for G × E interplay in child psychopathology (see Rutter, Moffitt, & Caspi, 2006; Shanahan & Hofer, 2005; Vendlinski, Lemery-Chalfant, Essex, & Goldsmith, 2011). The bioecological model (Bronfenbrenner & Ceci, 1994) postulates that the environmental variance will be larger in adverse circumstances, thereby masking genetic effects. If the adverse environmental influences are small (i.e., there is no or little risk or stress) genetic effects are more pronounced resulting in a higher heritability than in the presence of adverse events. The diathesis–stress model (e.g., Rende & Plomin, 1992) predicts that genetic vulnerability (diathesis) in the presence of

The study was supported by the European Research Council Genetics of Mental Illness ERC-230374, the Twin-Family Database for Behavior Genetics and Genomics Studies (NWO-480-04-004), the Longitudinal Survey and Panel Funding (NWO-481-08-011), the NARSAD Brain & Behavior Research Foundation-18633, the Genetic Influences on Stability and Change in Psychopathology from Childhood to Young Adulthood (ZonMW 912-10-020), the Neuroscience Campus Amsterdam (NCA), and the Institute for Health and Care Research (EMGO+). We thank the three anonymous reviewers for their valuable comments and suggestions during the review process of the present article.

Correspondence concerning this article should be addressed to Dylan Molenaar, Psychological Methods, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA, Amsterdam, Netherlands. Electronic mail may be sent to d.molenaar@uva.nl.

 0009-3920/2015/8606-0023
DOI: 10.1111/cdev.12451

environmental stress increases the chance of behavioral problems and that the heritability of the trait will be higher for children in at-risk environments. Finally, in the differential susceptibility model (Belsky et al., 2009), individuals are hypothesized to differ in the degree to which they are susceptible to positive or negative environmental influences due to plasticity genes. Children with a plastic genetic predisposition score higher on problem behavior in the presence of adverse environmental factors, but lower in the presence of protective environment factors, compared to children without this genetic predisposition. It follows that the bioecological model on one hand and the diathesis–stress and differential susceptibility models on the other have differing predictions regarding the direction of the G × E interaction effect.

Mixed Empirical Results

Candidate G × E interaction studies show that it is challenging to detect robust effects. For instance, Caspi et al. (2002) found in a large sample of male children that the risk of developing antisocial behavior after being exposed to maltreatment was mostly increased in individuals with a functional polymorphism in the monoamine oxidase A gene. This result supports the diathesis–stress model and was replicated by Kim-Cohen et al. (2006), but not by Haberstick et al. (2005) or Young et al. (2006).

Other studies investigated G × E interaction in the context of the classical twin design with measured environmental factors. Such studies model genetic influences as latent effects on the phenotype and test whether these genetic effects are moderated by the environmental variables (e.g., Boomsma, de Geus, van Baal, & Koopmans, 1999; Purcell, 2002). However, as with the candidate gene approach, results are mixed. Vendlinski et al. (2011) give examples of G × E interaction studies, which look at the moderation of the heritability of child problem behavior by various environmental risk factors, including negative life events, low socioeconomic status (SES), family conflicts, and parent–child relationship problems. They conclude that evidence is found for G × E interaction. However, the direction of the effect, that is, an increase or a decrease of the heritability in an adverse environment, varies across studies and across traits (see Vendlinski et al., 2011) making a conclusion about the tenability of the bioecological model or the diathesis–stress and differential susceptibility models difficult to draw.

There are several reasons for the discrepancies in results from G × E interaction studies in child problem behavior. First, results depend on the enironmental risk factor being investigated, even when risk factors seem closely related. For instance, Button, Lau, Maughan, and Eley (2008) reported maternal putative discipline to lower heritability of externalizing problems in children, whereas Hicks, South, DiRago, Iacono, and McGue (2009) found mother–child relationship problems to increase heritability of externalizing problems. Second, in studies of measured candidate genes, like the Caspi et al. (2002) study into childhood maltreatment, effect sizes are small, requiring large sample sizes to have adequate power to detect or replicate them. For instance, Duncan and Keller (2011) have shown that the pattern of the currently published results of small and large studies on G × E interaction point to positive publication bias. Third, in psychiatry in general and in the field of child problem behavior research in particular, it has been argued that identifying relevant environments is at least as challenging as identifying relevant genetic polymorphisms, as there are numerous candidates (Loeber & Farrington, 1998; Moffitt, 2005), which need to be measured reliably to ensure acceptable power to detect a G × E interaction (Moffitt, 2005). Fourth, G × E interaction in general can arise due to arbitrary properties of the measurement scale (Eaves, 2006; Eaves, Last, Martin, & Jinks, 1977; Purcell, 2002). In a simulation study, Molenaar and Dolan (2014) illustrate this problem by showing that even mildly skewed phenotypic data (e.g., due to a disproportionally number of easy items) can lead to spurious G × E interactions (see also Tucker-Drob, 2009; Tucker-Drob, Harden, & Turkheimer, 2009).

Interactions With the Family or the Individual

As is evidenced by the aforementioned research, the current G × E interaction research in child problem behavior is often aimed at detailing the nature of the interaction, that is, the research attempts to establish the precise environments and genes that interact. An advantage is that—if successful—results of such studies are very informative as they supply substantial detail about the nature of the interaction. However, as discussed earlier, a disadvantage is that results may depend too much on arbitrary properties of the studies (e.g., how the environment is measured, what measurement instruments are used) and chance fluctuations (i.e., due to small effect sizes and multiple testing issues), which cause

the mixed results that are currently observed in G × E research in child problem behavior.

A key solution to this problem is to look at the problem from a broader perspective. That is, we ask two questions, "How do genes and environment interact in child problem behavior?" and "What are the key characteristics of the environment that put genetically predisposed children at risk to develop some sort of problematic behavioral condition?" Such a focus is more general and less concerned with exact details (i.e., which environments and which genes). To answer these questions, no candidate genes or environments need to be identified. That is, they can both be treated as latent (unmeasured) variables. This has the important advantage that the problems outlined above associated with the identification and measurement of candidate genes and environments are overcome. In addition, as discussed next, by using item-level data to address these questions, the issue of arbitrary measurement properties in the outcome variable is also overcome.

Thus, by focusing on these more general questions, important problems with respect to the mixed results in G × E research in child problem behavior can be overcome. In addition, the results of such more general studies are still informative about the nature of a possible G × E interaction in child problem behavior. That is:

1 The results give an important indication of the direction and shape of the G × E interaction. That is, does the environment promote the genetic effect in children that are genetically at risk, or does it affect children that are genetically protected?
2 The results give an important indication of the nature of the environments that interact with genes. That is, are the interacting environments on the familial level or are they on the individual level?

From the current literature on G × E interaction in child problem behavior, these two questions cannot be answered straightforwardly, as both the nature and the direction of the interaction effect differs across the results from various studies (e.g., as discussed earlier, the variable *maternal putative discipline* protected genetically predisposed children, while the highly related variable *mother–child relationship problems* promoted the effect in genetically predisposed children; see Vendlinski et al., 2011). Answering these two questions is, however, important to our understanding of child problem behavior: First, knowing the nature and direction of the environmental effects in G × E interactions may help the clinician to decide on what therapy to administer to a child with problem behavior. Second, the results can shed a light on which theoretical model for G × E interplay in child problem behavior is tenable: the bioecological model, the diathesis–stress model, or differential susceptibility model. Finally, in future studies, these results may give an important indication to the developmental scholar of what results to expect in furthering their developmental studies using candidate genes and environments. These studies can be set up to be more confirmatory to counter the problem of chance capitalization.

The Present Study

We adopt an alternative to the measured genes and measured environments approaches. In this alternative approach, both the genetic and the environmental sources of phenotypic variance are treated as unmeasured variables (see Molenaar, van der Sluis, Boomsma, & Dolan, 2012; Van der Sluis, Dolan, Neale, Boomsma, & Posthuma, 2006), avoiding problems associated with candidate genes and environments as discussed earlier. That is, by treating the environment and genotype as latent variables, the focus is on the net results of all environments and genotypes that have an effect on the phenotype. In our alternative approach, the item-level data of the behavioral measures are taken into account, which makes the interaction results insensitive to scale properties (Molenaar & Dolan, 2014; Schwabe & van den Berg, 2014). Note that using item-level data are advisable in G × E interaction studies in general (i.e., also when analyzing using candidate genes or environments) and not specifically in the case of unobserved genes and environments.

In the current article, we present the Unmeasured Genotype × Unmeasured Environment methodology and apply it to a range of child problem behaviors. The data consist of maternal and paternal rating on twins from the Netherlands Twin Register (NTR) and include 14,755 twin pairs at the age of 5 years. The focus will be on a broad range of problem behavior as assessed by the short Devereux Child Behaviors (DCB) Rating Scale (Spivack & Spotts, 1966; Van Beijsterveldt, Verhulst, Molenaar, & Boomsma, 2004), which includes emotional liability, social isolation, aggressive behavior, attention problems, dependency, anxiety problems,

and physical coordination. In an analysis that included approximately 50% of the current sample of twin pairs, the heritability estimates of these dimensions varied between 0.5 and 0.8 (Van Beijsterveldt et al., 2004; note that the social isolation scale was not included in this article). The current study aims to reveal whether the heritability estimates are dependent on the environment and whether the results are similar or different for different aspects of problem behavior.

Testing Unmeasured Genotype × Unmeasured Environment Interactions

Under a simple additive model, the phenotypic variance for each phenotypic problem behavior (Y) can be decomposed into variance due to genetic predisposition (the additive genetic factor, A) and due to the environment, which consists of family effects (the common environmental factor, C), and individual effects (the unique environmental factor, E). That is,

$$\sigma^2_Y = \sigma^2_A + \sigma^2_C + \sigma^2_E \quad (1)$$

where Y denotes the phenotypic score on one of the scales for an individual. In this so-called Additive Genetic–Common Environment–Unique Environment (ACE) model (Posthuma et al., 2003), the correlations among the latent genetic and environmental effects are as follows: cor(E_1, E_2) = 0, cor(C_1, C_2) = 1 for all twin pairs, and cor(A_1, A_2) = 0.5 for dyzygotic (DZ) twins and cor(A_1, A_2) = 1 for monozygotic (MZ) twins, where subscripts 1 and 2 denote the two members of a twin pair. Within the ACE model, a G × E interaction can be operationalized as A × E or as A × C (see Jinks & Fulker, 1970; Molenaar et al., 2012; Van der Sluis et al., 2006), that is, an interaction of genes with the familial environment or with the individual environment. That is:

$$\sigma^2_{Y|A} = \sigma^2_{C|A} + \sigma^2_{E|A} = \exp(\gamma_0 + \gamma_1 A) + \exp(\beta_0 + \beta_1 A) \quad (2)$$

where Y|A denotes the phenotypic score given (or conditional on) the genetic predisposition (i.e., the level of A) and exp(.) is the exponential function—$\exp(x) = e^x$.

In the model, a test on an interaction between the genotype and the family environment (A × C) thus concerns a statistical test on the significance of parameter γ_1, and a test on an interaction between the genotype and the individual environment (A × E) concerns a test on the significance of parameter β_1. If the estimate of parameter γ_1 is significantly larger than 0, this indicates that the effect of the family (the variance of C, $\sigma_C{}^2$) increases with increasing genetic predisposition to develop problem behavior (the levels of A). Congruently, if γ_1 is significantly smaller than 0, the effect of the family has a smaller influence on the development of problem behavior for individuals with a larger predisposition for problem behavior. A similar interpretation holds for β_1 in the case of an A × E interaction. Molenaar et al. (2012) stressed that both effects should be taken into account simultaneously as unmodeled A × C interaction may result in an increased Type I error rate to detect an A × E interaction effect and vice versa. Taking both effects into account prevents this problem and increases the statistical power to detect a genuine effect. In addition, the method is insensitive to G × E correlation, which can suggest G × E interaction when not properly accounted for (Purcell, 2002).

Note that the way in which the interactions are interpreted here is not different from the way in which interactions are interpreted in standard regression analysis. Here, we interpret an interaction as "the effect of the environment on the phenotype which depends on an individual's genetic predisposition," where the genetic predisposition is not an observed variable, but a continuous unmeasured variable. This operationalization is consistent with the classical conceptualization of G × E interaction by Jinks and Fulker (1970; see also below).

Departures From Normality

As the model concerns the linear effects of A on the logarithm of the variance of C and E, we refer to this model as the linear interaction model. In the linear interaction model, β_1 statistically describes the possible dependency of $\sigma^2{}_E$ on A (A × E interaction) and γ_1 the possible dependency of $\sigma^2{}_C$ on A (A × C interaction). As A is a latent variable (factor score), we cannot make inferences about the position of an individual subject on this variable. However, from the marginal distribution of the data (i.e., the effects of A "averaged out"), we can make inferences about the presence of G × E interactions in the population. For normally distributed A, C, and E factors, the presence of A × E and A × C interactions implies departures from

bivariate normality in the observed distribution of the Twin 1 and Twin 2 phenotypic scores. As can be seen in Figure 1, the presence of a positive A × C interaction ($\gamma_1 > 0$) results in more variance correlated between twin members in the right tail of the phenotypic distribution (i.e., the bivariate tale becomes longer). In addition, the presence of a positive A × E interaction ($\beta_1 > 0$) results in more variance *un*correlated between twin members in the right tail of the phenotypic distribution (i.e., the bivariate tale becomes wider). Thus, parameter β_1 picks up the departure from normality that is due to A × E interaction and parameter γ_1 picks up the departure from normality that is due to A × C interaction.

Illustration Using Height and Weight as Phenotype

To illustrate the present model, we apply it to data by Osborne (1980); the data contain 130 MZ twin pairs and 107 DZ twin pairs between the ages of 12 and 20. Here, we analyze the height and weight of the twins. We chose height as this variable is likely not subject to G × E interaction. For the weight variable, we expect G × E interaction. The average height of the sample is 65.22 in. (SD = 3.57) with a minimum of 53 and a maximum of 76 in. The average weight of the sample is 122.19 pounds (SD = 24.94) with a minimum of 70 and a maximum of 233 pounds.

The parameter estimates based on the within-zygosity standardized height and weight variables are presented in Table 1. The table displays the parameter estimates for height and the weight with 99% confidence intervals. As can be seen for height, both the A × E interaction (β_1) and the A × C interaction (γ_1) parameters are nonsignificant as judged by their confidence intervals. That is, as expected, there is no G × E interaction underlying height in the Osborne data. For weight, it can be seen that the A × E interaction parameter (β_1) is nonsignificant; however, the A × C interaction parameter (γ_1) is significant with an estimate of 0.94. This indicates that the effect of the family is larger for subjects with a higher genetic predisposition to weigh more. That is, if you have a genetic predisposition, you are more vulnerable to familial influences that make you weigh more (e.g., familial eating habits).

In Figures 2 (height) and 3 (weight), the predicted distribution of the data is displayed for the full model together with the observed data. As can be seen for height (Figure 2), the predicted distribution hardly departs from normality. However, for weight (Figure 3) the predicted distribution departs from normality congruently with an A × C interaction as illustrated in Figure 1c. Some non-normality due to A × E interaction can also be seen (i.e., a reverse effect of Figure 1b as for weight, the β_1 estimate is smaller than 0, while Figure 1b is based on a positive β_1 parameter), although this departure is nonsignificant (as already concluded above).

Curvilinear Effects

Due to the linear nature of the model in Equation 2, only monotone effects of G × E interaction are tested; that is, the effects of E and C are only allowed to be either strictly increasing or decreasing across genetic predisposition, A. To allow for non-monotone effects, the model above can be extended to a curvilinear interaction model:

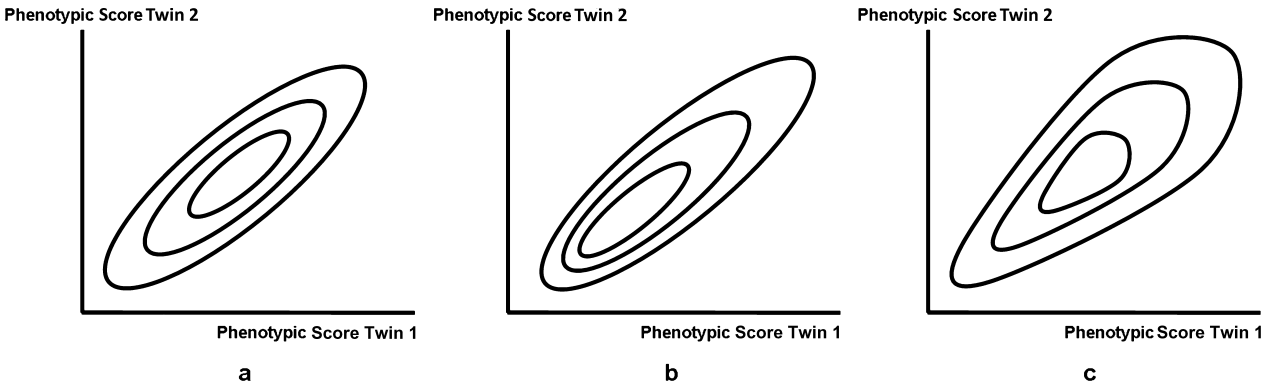

Figure 1. The model implied distribution of the phenotypic scores of Twin 1 and Twin 2 in the case of (a) a standard Additive Genetic–Common Environment–Unique Environment (ACE) model without Genotype × Environment interaction (Equation 1); (b) an A × C interaction with $\gamma_1 > 0$ (Equation 2); (c) an A × E interaction with $\beta_1 > 0$ (Equation 2).

Table 1
Illustration of the Unmeasured Genotype × Unmeasured Environment Methodology on the Weight and Height of 237 Twin Pairs

Phenotype	σ_A	β_0	β_1	γ_0	γ_1
Height	0.87 [0.75, 0.99]	−3.23 [−3.89, −2.76]	−0.16 [−1.12, 1.01]	−1.60 [−2.47, −0.80]	0.43 [−0.26, 0.82]
Weight	0.82 [0.70, 0.92]	−3.49 [−4.21, −2.75]	−0.57 [−1.34, 0.54]	−1.77 [−2.71, −1.00]	0.94 [0.62, 1.30]

Note. The 99% confidence intervals (based on the likelihood profile) are in brackets.

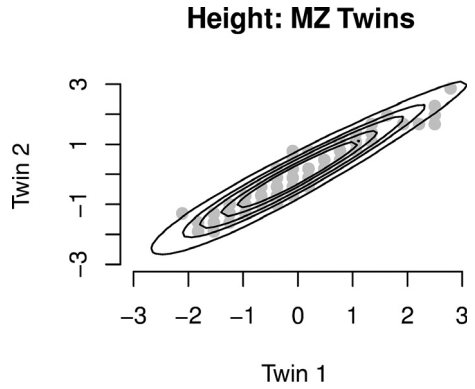

Figure 2. Predicted distribution by the full model (solid lines) and the observed distribution (dots) for the Twin 1 and Twin 2 height scores in the dyzygotic (DZ) and monozygotic (MZ) subsamples.

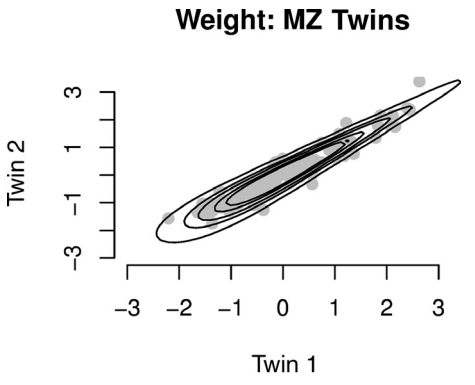

Figure 3. Predicted distribution by the full model (solid lines) and the observed distribution (dots) for the Twin 1 and Twin 2 weight scores in the dyzygotic (DZ) and monozygotic (MZ) subsamples.

$$\sigma^2_{Y|A} = \sigma^2_{C|A} + \sigma^2_{E|A} = \exp(\gamma_0 + \gamma_1 A + \gamma_2 A^2) + \exp(\beta_0 + \beta_1 A + \beta_2 A^2) \quad (3)$$

The two new parameters, β_2 and γ_2 account for possible quadratic effects of the genotype on the familial and individual environment (see Van der Sluis et al., 2006). It is also possible to add higher order effects (i.e., effects of A^3, A^4, etc.). However, doing so will rapidly over parameterize the model and likely pick up sampling fluctuations. Van der Sluis et al. (2006) showed that adding a quadratic effect to the linear interaction is viable in the sense that the quadratic effects can be recovered adequately. Therefore, we include the curvilinear effect of the genotype (i.e., A and A^2) to capture the most important patterns in the data.

Comparison to Conventional Approaches

Jinks and Fulker Regression

To test the interaction between the genotype and the individual environment, that is, an A × E interaction, Jinks and Fulker (1970) proposed to

regress the absolute MZ Twin 1 and MZ Twin 2 differences, $|Y_1 - Y_2|$, on the summed MZ twin scores, $Y_1 + Y_2$, and if desirable, the squared summed scores, $(Y_1 + Y_2)^2$. That is,

$$|Y_1 - Y_2| = a + \beta_1 * (Y_1 + Y_2) + \beta_2 * (Y_1 + Y_2)^2 + \text{residual}.$$

As discussed by Jinks and Fulker (1970), in the case of a linear A × E interaction, parameter β_1 will be significant, where $\beta_1 > 0$ will indicate higher environmental influences for increasing levels of A, and $\beta_1 < 0$ will indicate lower environmental influences for increasing levels of A. Note that this parameter is comparable to parameter β_1 in the current approach. In addition, parameter β_2 above models the quadratic effect of the interaction. Note that this parameter resembles parameter β_2 from the current approach. Van der Sluis et al. (2006) compared the Jinks and Fulker approach to the present approach for MZ twins only and found that the present approach has a slightly higher power.

Extended DeFries and Fulker Regression

Another related approach is the extended DeFries and Fulker regression method (Cherny, Cardon, Fulker, & DeFries, 1992; see also Brant et al., 2013). In this approach it is possible to test for Phenotype × Family interactions. Although this type of interaction is more general than the interaction considered in the present approach (Phenotype × Family vs. Genotype × Family), it is related as a presence of A × C in the data will be detected as a C × Phenotype interaction if A is an important source of variation in the phenotype.

The standard DeFries and Fulker regression method is given by:

$$Y_1 = b_0 + b_1 Y_2 + b_2 R + b_3 Y_2 R + \text{residual}$$

where R is 1 for MZ and 0.5 for DZ twins. As discussed in DeFries and Fulker (1985), in this regression b_1 will be an estimate of the familial environmental effects, σ^2_C, and b_3 will be an estimate of the genetic effect, $\sigma_A{}^2$.

Note that the data need to be doubly entered (each twin member occurs as Twin 1 and as Twin 2 in the data) and that therefore standard errors need to be obtained using a bootstrap procedure.

In Cherny et al. (1992), this approach is extended to yield:

$$Y_1 = b_0 + b_1 Y_2 + b_2 R + b_3 Y_2 R + b_4 Y_2^2 + b_5 Y_2^2 R + \text{residual}.$$

As discussed in Cherny et al. (1992; see also Brant et al., 2013), b_4 models the interaction between the familial environment and the phenotype. Specifically, if $b_4 > 0$ the family effect increases across the phenotype, and if $b_4 < 0$ the family effect decreases across the phenotype. Note that we thus expect a similar sign for parameter b_4 in the extended DeFries and Fulker regression approach as parameter γ_1 in our approach.

Method

DCB Questionnaire

The DCB consists of questions about problem behavior in children rated by the parents. The original DCB consists of 121 items, of which a shortened version was presented by Van Beijsterveldt et al. (2004). The short version includes 42 items that measure seven different aspects of problem behavior in children. The parents are asked to indicate on a 5-point scale whether the statements are applicable. The items of the questionnaire cover the following aspects of problem behavior: emotional liability (5 items, e.g., "Markedly impatient"), social isolation (3 items, e.g., "Quite timid or shy"), aggressive behavior (7 items, "Hits, bites, scratches other children"), attention problems (5 items, e.g., "Jumps from one activity to another"), dependency (5 items, e.g., "Does not want to do things for himself"), anxiety problems (6 items, e.g., "Concern about his physical health"), and physical coordination (5 items, e.g., "Gets dirty and untidy"). See Table 2 of Van Beijsterveldt et al. (2004) for all items within these scales.

Sample

We analyzed data collected by the NTR (Boomsma et al., 2002; Van Beijsterveldt et al., 2013). The twins' parents complete the DCB when the twins are 5 years old. The total sample comprises 9,750 DZ and 5,005 MZ twin pairs. Both paternal and maternal ratings are analyzed. Data are missing for 81 mothers and 1,619 fathers. Analyses are based on estimated phenotypic factor scores—as described below—that are standardized within gender. We note, however, that the pattern of results is similar to the analyses of the unstandardized scores. This is expected as gender

differences are small. Psychometric properties of the DCB in this sample and a more detailed description of the data collection have been reported in Van Beijsterveldt et al. (2004).

Phenotypic Scores

Interaction effects imply non-normality in the data, captured by the parameters γ_1 and β_1. Non-normality also can have other causes, like a disproportional number of easy or difficult items, or a floor or ceiling effect, which can result in spurious interaction effects. This may preclude interpretation of the results in terms of genuine G × E effects (see Eaves, 2006; Eaves et al., 1977; Purcell, 2002). Item scores of the DCB are skewed due to the disproportional use of the lowest answer category, as may be seen in Figure 4 for the item score distribution of the seven items in the Aggressive Behavior subscale.

If G × E interaction is tested either separately on each item of the DCB or on the sum scores of the items, artificial interaction effects may arise (see Molenaar & Dolan, 2014; Schwabe & van den Berg, 2014) if data are assumed to be continuous with equal measurement precision at each point of the scale while in fact the data are ordinal with unequal measurement precision across the scale. Note that analyzing ordinal item data as if they are continuous is not recommended in ACE modeling in general. That is, neglecting the ordinal nature of the data can bias correlations downward (Olsson, 1979) leading to an underestimation of the heritability (see, e.g., Van den Berg, Glas, & Boomsma, 2007). A solution to this problem (Molenaar & Dolan, 2014; see also Schwabe & van den Berg, 2014) is to explicitly take the ordinal item-level properties of the data into account. By doing so, the unequal measurement precision across the scale is taken care of and does not influence the results concerning possible G × E interactions. An example of a well-accepted measurement model for Likert scale item responses is the graded response model (Samejima, 1969),

$$P(X_{ip} = c|\theta_p) = \frac{1}{1 + \exp(-\alpha_i\theta_p - \tau_{ic})} - \frac{1}{1 + \exp(-\alpha_i\theta_p - \tau_{i(c+1)})}. \quad (4)$$

In this model, the probability of answering in a certain answer category [$P(X_{ip} = c \mid \theta)$] is a function of item parameters, that is, the discrimination parameters, α_i, and the threshold parameters, τ_{ic}, and the probability of a response is a function of the phenotypic person parameters, θ_p. As the measurement and phenotypic properties are separated, the phenotypic person parameters, θ_p, can be submitted to the A × C and A × E decomposition in Equations 2 and 3 (i.e., θ_p is substituted for Y). Non-normality due to floor, ceiling, and poor scaling effects are captured by the threshold parameters, τ_{ic}. For instance, in the case of a floor effect as in Figure 4, the first category will be associated with a large threshold parameter (τ_{i1}) accounting for the floor effect. Doing so, the β_1 and γ_1 interaction parameters are unaffected by the properties of the scale. Specifically, in a simulation study, Molenaar and Dolan (2014) show that even in the case of severe floor effects in Likert item scores due to the disproportionate use of the lower answer category, the true values of the interaction parameters β_1 and γ_1 are adequately recovered using the approach above. As the method by Molenaar and Dolan is numerically challenging, we use a two stage procedure to test for A × E and A × C interactions. In the first stage, we fit the measurement model in Equation 4 to the MZ and DZ twin data of all the seven scales of the DCB questionnaire simultaneously using discrete factor analysis (Wirth & Edwards, 2007). Next, we calculate within gender standardized estimates for the phenotypic person parameters (θ_p) for each scale. In the second stage, we analyze these scores in the univariate approach of Molenaar et al. (2012). In this two-stage procedure, the item-level properties are explicitly taken into account by the item parameters in the first stage; thus, the phenotypic person scores will be independent of the properties of the measurement scale. Figure 5 shows the phenotypic person scores of the aggressive behavior scale (of which Figure 4 displayed example items with severe floor effects). To ensure that results presented in this article are not chance findings, we created two cross-validation samples to see whether the results would hold, by randomly splitting the total sample into two independent subsamples.

A disadvantage of two-stage procedures is that standard errors of the estimated phenotypic person parameters (θ_p) in the first stage are neglected in the model estimation of the second stage. This will result in an underestimation of the standard errors of the parameters in the second stage. However, because we have a very large sample size (for the ratings of one parent we have 14,755 records containing 72 observations each), we assume that standard errors of the θ_p estimates can be reasonably neglected. We checked the results of the

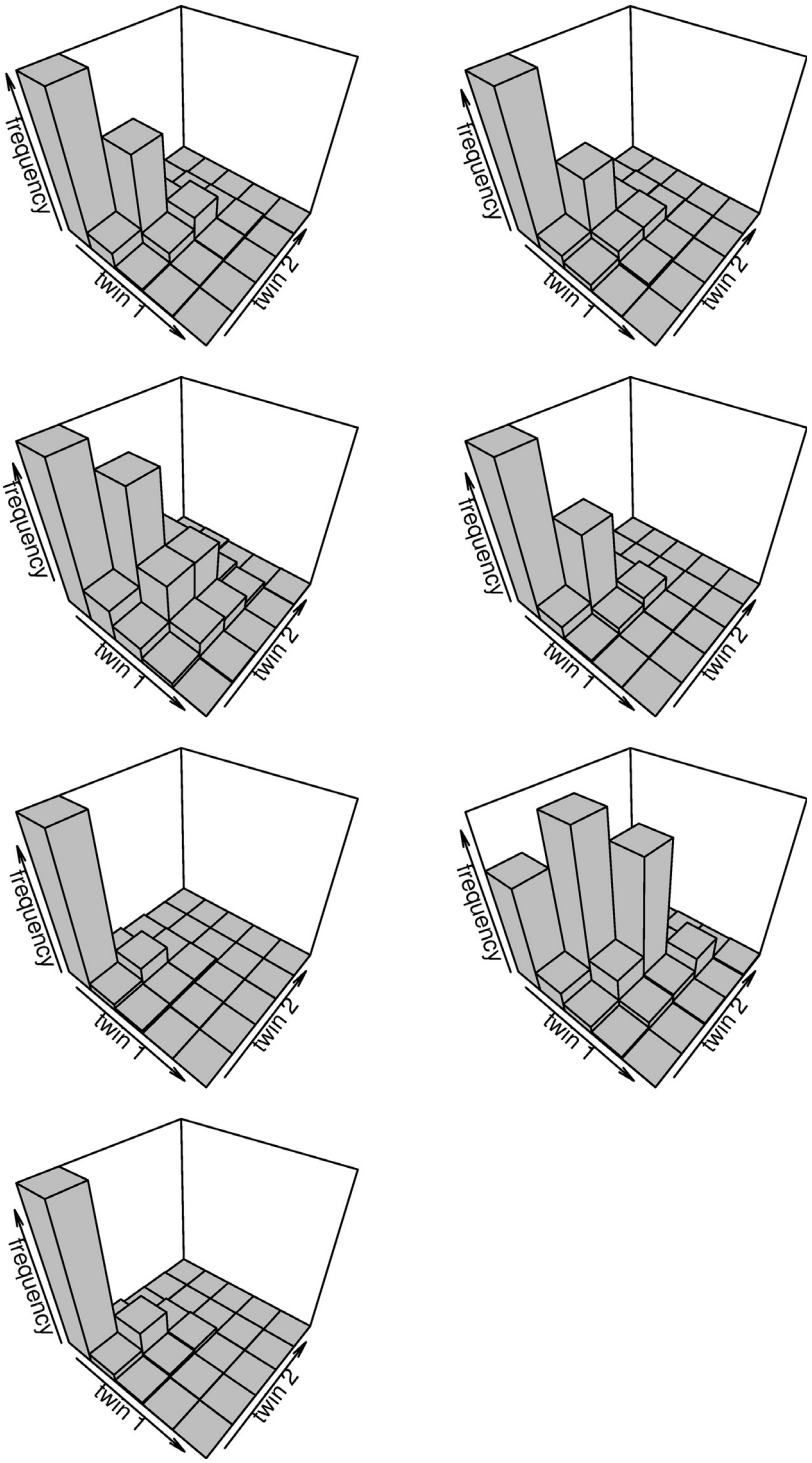

Figure 4. Distribution of the item scores of the Aggressive Behavior subscale for the mother ratings in the monozygotic (MZ) subsamples.

two-stage procedure for the linear interaction model in Equation 2 with the item-level-based procedure (Molenaar & Dolan, 2014) for the maternal ratings in the first cross-validation sample. The pattern of results were the same (results are available upon request). We therefore conclude that the results of the two-stage procedure are reliable and that the A × E and A × C effects are robust to the exact method used. In Mplus (Muthén & Muthén, 2007), we used the full seven-factor structure of the DCB

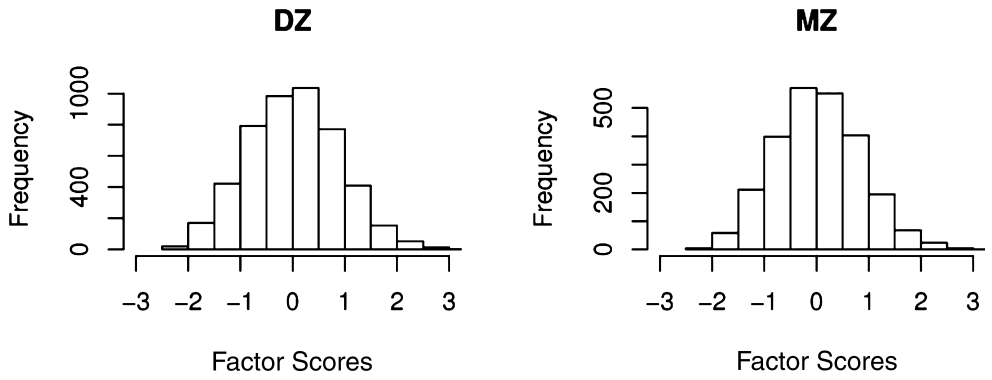

Figure 5. Distribution of the factor scores of the Aggressive Behavior subscale for the mother ratings in the dyzygotic (DZ) and monozygotic (MZ) subsamples. For each pair, one twin member is randomly chosen.

on the item scores of MZ and DZ Twin 1 and Twin 2 members to obtain estimates for θ_p. Discrimination and threshold parameters were constrained to be equal across MZ and DZ twins and across Twin 1 and Twin 2 members. On the factor scores, the A × C and A × E interactions were tested using Mx (Neale, Boker, Xie, & Maes, 2006). As the sample size is very large, we base our model selection on Akaike's information criterion (AIC; Akaike, 1974) and not on confidence intervals or likelihood ratio tests (which are sensitive to large sample sizes).

Results

Three interaction models were considered (a) a linear interaction model with A × E, (b) a linear interaction model with A × C, and (c) a full linear interaction model with both A × E and A × C. In addition, we fit the ACE model without interactions as a baseline model. The models were compared in terms of their goodness of fit as indicated by the AIC fit statistic. For all scales, both A × E and A × C interactions are present (see Table 2) except for the social isolation scale, for which only an A × E interaction was found. Results were highly congruent across the mother and father ratings and across validation samples (labeled 1 and 2 in the table). We also considered the Bayesian information criterion (BIC) fit index, which gave the same pattern of results.

As linear A × E and A × C interactions are present in the data, we proceeded by adding quadratic terms to the linear interactions. This resulted in the following curvilinear models: (a) a curvilinear interaction for A × E (and only a linear interaction for A × C), (b) a curvilinear interaction for A × C (and only a linear interaction for A × E), and (c) a full model in which both A × C and A × E are curvilinear interactions. As can be seen in Table 2, the AIC fit index indicated for all scales, all cross-validation samples, and both the mother and father ratings that the full curvilinear model fits the best. Again the BIC fit index gave the same pattern of results.

In Table 3, parameter estimates and heritabilities are displayed for the full curvilinear model including both A × C and A × E interactions, and the traditional ACE model without A × C and A × E interactions (see the Appendix for details on how we obtained the heritability from the parameter estimates). The results are highly congruent across the mother and father ratings and the validation samples (again labeled 1 and 2 in the table) with homogeneous heritability estimates and A × C and A × E parameter estimates (β_1, β_2, γ_1, and γ_2). With respect to the A × C and A × E effects, the interaction parameter estimates for β_1, β_2, γ_1, and γ_2 all have the same sign. Figure 6 summarizes the A × E interaction results in the full curvilinear model and Figure 7 the A × C interaction results in the full curvilinear model. For both graphs, we used the results from the first validation sample of the mother ratings. For all scales, the effect of the individual, σ^2_E, is first increasing for increasing levels of genetic predisposition, A, with a peak near 1.5–2.5, after which σ^2_E is decreasing. The family effect, σ^2_C, is first increasing for increasing levels of A with a peak near −1, after which σ^2_C is decreasing. Thus, σ_E^2 peaks above the average A level and σ_C^2 peaks below the average A level. We, therefore, conclude that the effect of the individual environment is stronger for higher levels of genetic predisposition and that the effect of the family is weaker for higher levels of genetic predisposition.

Table 2
Akaike's Information Criterion for the Different Models as Applied to Either the Mother or Father Ratings in Validation Samples 1 or 2

	Linear interaction				Curvilinear interaction		
	None	A × E	A × C	Full	A × E	A × C	Full
Aggressive behavior							
Mother							
1	8,316	8,123	7,675	**7,598**	7,435	7,400	***7,243***
2	8,137	8,016	7,675	**7,621**	7,528	7,446	***7,328***
Father							
1	6,867	6,761	6,239	**6,187**	6,010	6,008	***5,840***
2	6,742	6,637	6,225	**6,193**	6,092	6,071	***5,962***
Attention problems							
Mother							
1	10,557	10,496	10,359	**10,326**	10,319	10,219	***10,038***
2	10,746	10,711	10,581	**10,567**	10,565	10,515	***10,365***
Father							
1	9,355	9,315	9,251	**9,202**	9,187	9,121	***9,026***
2	9,347	9,307	9,216	**9,166**	9,150	9,080	***8,903***
Anxiety							
Mother							
1	8,164	8,063	7,763	**7,698**	7,632	7,540	***7,477***
2	8,192	8,077	7,692	**7,636**	7,572	7,524	***7,446***
Father							
1	6,578	6,538	6,222	**6,158**	6,069	6,029	***5,938***
2	6,498	6,436	6,079	**6,005**	5,951	5,898	***5,838***
Dependency							
Mother							
1	7,835	7,778	7,542	**7,429**	7,419	7,248	***7,197***
2	8,271	8,225	8,102	**8,025**	8,012	7,880	***7,816***
Father							
1	7,116	7,018	6,777	**6,646**	6,644	6,480	***6,430***
2	6,937	6,936	6,614	**6,484**	6,473	6,294	***6,230***
Emotional liability							
Mother							
1	8,914	8,797	8,553	**8,458**	8,419	8,367	***8,289***
2	8,880	8,765	8,426	**8,337**	8,277	8,244	***8,124***
Father							
1	7,386	7,364	6,995	**6,881**	6,828	6,793	***6,725***
2	7,147	7,119	6,765	**6,677**	6,619	6,581	***6,480***
Physical coordination							
Mother							
1	8,996	8,874	8,678	**8,656**	8,655	8,570	***8,551***
2	9,397	9,172	8,974	**8,934**	8,931	8,814	***8,775***
Father							
1	7,827	7,791	7,558	**7,499**	7,476	7,371	***7,297***
2	7,845	7,711	7,425	**7,392**	7,383	7,298	***7,255***
Social isolation							
Mother							
1	9,822	8,952	**8,058**	8,058	7,767	8,024	***7,742***
2	9,782	8,721	**7,815**	7,817	7,537	7,795	***7,504***
Father							
1	8,790	8,183	**7,515**	7,517	7,232	7,501	***7,216***
2	8,483	7,771	**7,085**	7,086	6,805	7,060	***6,773***

Note. The curvilinear A × E and A × C interaction models include all linear interactions. For instance, the A × E curvilinear interaction model includes the linear and quadratic A × E effect, together with the linear A × C effect. For both the linear interaction model and the curvilinear interaction model, the smallest fit index is given in boldface. The overall smallest fit index is in italic.

To illustrate how the observed data fit the model predictions, see Figure 8 for the predicted distribution by the full curvilinear model and the observed data distribution for the mother ratings of the Aggressive Behavior scale. As can be seen, the distributions are well predicted with departures from normality, consistent with the A × E and A × C effects illustrated in Figure 1.

Table 3 also summarizes the proportion of the heritability in the traditional ACE model that is due to unmodeled A × C and A × E interactions (p_{h^2}). The proportions of heritability due to unmodeled A × C and A × E interactions are roughly around 0.20 with a minimum of 0.08 for the Anxiety scale and a maximum of 0.37 in the Attention Problems scale. These results thus indicate that, on average, heritability as estimated in an additive model decreases by around 20% when G × E interaction is added to the model.

Conventional Approaches

To show how the present analysis compares to the more conventional approaches discussed earlier, we applied the Jinks and Fulker regression approach and the extended DeFries and Fulker regression approach to the data of all scales in Sample 1 of the mother ratings. As noted above, the Jinks and Fulker regression method is only suitable for the MZ twin data, and the extended DeFries and Fulker regression approach requires the data to be doubly entered, which makes it necessary to bootstrap the standard errors.

The results are in Table 4. For the Jinks and Fulker regression, the parameters of interest are β_1 and β_2 as they correspond directly to the β_1 and β_2 parameters of our curvilinear model. That is, β_1 models the linear component of the A × E interaction (parameter β_1) and β_2 models the quadratic component of the A × E interaction (parameter β_2). As can be seen from the table, these parameters of the Jinks and Fulker regression are in the same direction as in our results (see Table 3). The exact parameter values are different as these are on a different scale. In addition, in the Jinks and Fulker approach, the quadratic component is not always significant as judged by its standard error. This may be because the power is smaller for the Jinks and Fulker approach as this method only uses the MZ twin data.

For the extended DeFries and Fulker regression approach, the parameter of interest is b_4 as this parameter models a linear Family × Phenotype interaction (Phenotype × C). As discussed earlier,

Table 3
Parameter Estimates, Heritability in the Full Curvilinear Interaction Model (h_{full}^2), Heritability in the Traditional Additive Genetic–Common Environment–Unique Environment (ACE) Model (h_{ACE}^2), and the Proportion of the Heritability That Is Due to Genotype × Environment Interaction (p_{h^2}) of the Mother and Father Ratings in Validation Samples 1 or 2

	Curvilinear interaction model								Traditional ACE model				
	Parameter estimates								Parameter estimates				
	σ_A	β_0	β_1	β_2	γ_0	γ_1	γ_2	h_{full}^2	σ_A	β_0	γ_0	h_{ACE}^2	p_{h^2}
Aggressive behavior													
Mother													
1	0.61	−2.67	2.61	−0.64	−0.63	−0.66	−0.37	0.36	0.71	−1.55	−1.25	0.50	0.28
2	0.65	−2.53	2.30	−0.58	−0.72	−0.68	−0.39	0.41	0.73	−1.64	−1.30	0.54	0.24
Father													
1	0.54	−2.76	2.72	−0.71	−0.50	−0.64	−0.35	0.29	0.67	−1.65	−1.03	0.45	0.36
2	0.61	−2.62	2.35	−0.59	−0.59	−0.45	−0.26	0.36	0.62	−1.62	−0.87	0.39	0.08
Attention problems													
Mother													
1	0.76	−1.77	2.26	−0.81	−1.80	−2.07	−0.87	0.54	0.86	−1.17	[a]	0.70	0.23
2	0.80	−1.74	2.14	−0.78	−1.89	−2.02	−0.84	0.58	0.84	−1.12	[a]	0.68	0.15
Father													
1	0.74	−1.49	1.64	−0.58	−1.86	−1.58	−0.49	0.53	0.86	−1.19	[a]	0.71	0.25
2	0.70	−1.86	2.75	−1.07	−1.40	−1.81	−0.79	0.45	0.86	−1.21	[a]	0.71	0.37
Anxiety													
Mother													
1	0.62	−2.31	1.83	−0.39	−0.69	−0.71	−0.42	0.39	0.66	−1.50	−1.08	0.44	0.11
2	0.59	−2.28	2.06	−0.50	−0.65	−0.54	−0.28	0.34	0.61	−1.44	−0.95	0.38	0.11
Father													
1	0.58	−2.46	2.09	−0.52	−0.57	−0.55	−0.29	0.33	0.60	−1.62	−0.81	0.36	0.08
2	0.55	−2.55	2.03	−0.43	−0.50	−0.57	−0.33	0.31	0.58	−1.63	−0.77	0.34	0.09
Dependency													
Mother													
1	0.73	−2.68	1.83	−0.36	−0.84	−0.71	−0.36	0.51	0.82	−1.84	−1.77	0.67	0.24
2	0.73	−2.53	1.92	−0.43	−0.88	−0.76	−0.38	0.50	0.85	−1.79	−2.19	0.72	0.31
Father													
1	0.66	−2.79	2.07	−0.40	−0.60	−0.74	−0.39	0.41	0.80	−1.79	−1.56	0.63	0.35
2	0.67	−2.78	2.04	−0.41	−0.66	−0.76	−0.38	0.43	0.80	−1.86	−1.54	0.63	0.32
Emotional liability													
Mother													
1	0.76	−2.20	1.72	−0.38	−1.28	−0.88	−0.35	0.56	0.83	−1.55	−2.27	0.69	0.19
2	0.71	−2.18	2.03	−0.55	−1.09	−0.77	−0.29	0.49	0.80	−1.51	−1.90	0.63	0.22
Father													
1	0.66	−2.25	1.80	−−0.41	−0.92	−0.59	−0.19	0.43	0.74	−1.59	−1.37	0.54	0.20
2	0.59	−2.43	2.30	−0.61	−0.63	−0.66	−0.32	0.34	0.67	−1.57	−1.08	0.45	0.24
Physical coordination													
Mother													
1	0.88	−2.26	1.27	−0.20	−1.89	−0.95	−0.42	0.72	0.92	−1.77	−3.13	0.80	0.10
2	0.83	−2.37	1.81	−0.37	−1.46	−0.89	−0.40	0.63	0.91	−1.65	−3.90	0.80	0.21
Father													
1	0.83	−2.33	1.73	−0.45	−1.39	−1.03	−0.50	0.64	0.88	−1.71	−2.74	0.76	0.16
2	0.79	−2.48	1.80	−0.36	−1.20	−0.69	−0.33	0.59	0.90	−1.75	−2.97	0.78	0.24
Social isolation													
Mother													
1	0.63	−1.93	2.79	−0.90	−1.26	−0.25	−0.15	0.40	0.75	−1.17	−2.09	0.57	0.30
2	0.62	−2.03	2.80	−0.84	−1.28	−0.27	−0.11	0.39	0.77	−1.22	−2.23	0.60	0.35
Father													
1	0.69	−−1.91	2.76	−0.95	−1.38	−0.38	−0.17	0.46	0.77	−1.20	−2.10	0.58	0.21
2	0.71	−2.09	3.07	−1.09	−1.37	−0.33	−0.14	0.48	0.77	−1.32	−1.89	0.59	0.19

Note. See the Appendix for more details about how we calculated heritability from the parameter estimates; p_{h^2} is calculated as $p_{h^2} = 1 - h^2_{ACE}/h^2_{full}$.
[a]For the Attention Problems scale, there is no C component in the traditional ACE model.

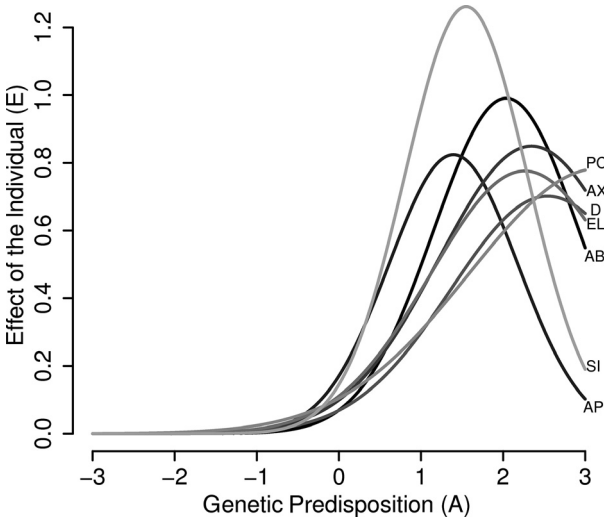

Figure 6. Results concerning A × E interaction. The variance of E as a function of the additive genetic factor A. EL = emotional liability; SI = social isolation; AB = aggressive behavior; AP = attention problems; D = dependency; AX = anxiety problems; PC = physical coordination. The graph is based on the results of the first validation sample of the mother ratings.

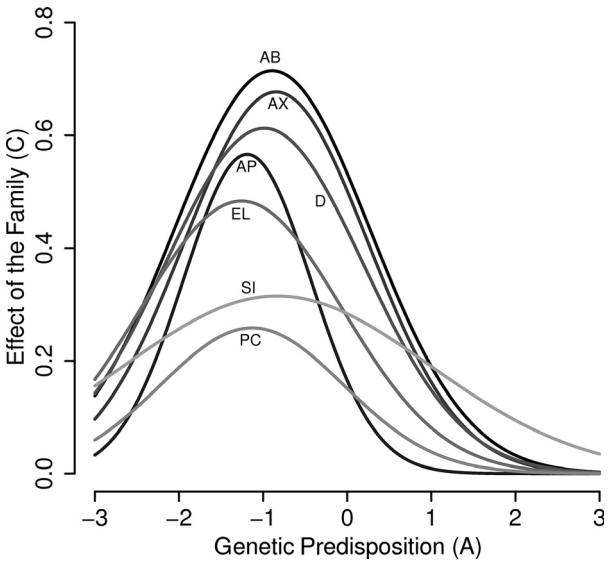

Figure 7. Results concerning A × C interaction. The variance of C as a function of the additive genetic factor A. EL = emotional liability; SI = social isolation; AB = aggressive behavior; AP = attention problems; D = dependency; AX = anxiety problems; PC = physical coordination. The graph is based on the results of the first validation sample of the mother ratings.

we expect the b_4 parameter to pick up A × C interactions given that A is an important source of individual differences in the phenotype. That is, b_4 should correspond to γ_1 from our approach. In the DeFries and Fulker regression approach, curvilinear

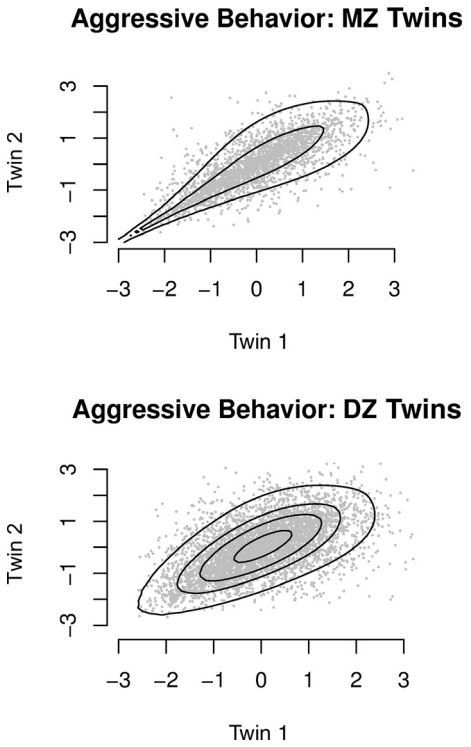

Figure 8. Predicted distribution by the full curvilinear model (solid lines) and the observed distribution (dots) for the mother ratings of the Aggressive Behavior scale in the dyzygotic (DZ) and monozygotic (MZ) subsamples.

effects are not taken into account (i.e., this form is not considered; see Cherny et al., 1992). As can be seen from the table, all b_4 parameter estimates are in the same direction as γ_1 in our study. Note that b_1 is negative in the case of attention problems and physical coordination, which suggests a negative σ^2_C. As the extended DeFries and Fulker regression does not involve explicit parameter constraints, this can happen in practice if the family effect is small or absent in the data (as is the case for these two scales; see Table 3). Taken together, in the two more conventional analyses, we found similar effects as in the analysis using the current approach.

Discussion

Explaining variation among children in their behavioral problems is at the core of child development: Why do some children thrive and others do not? The standard quantitative genetic model offers a strong approach to address the sources of variation, but its application has been criticized for focusing on main effects rather than on G × E interactions. The method we propose, though it needs relatively

Table 4
Parameter Estimates (SEs) *for the Conventional Approaches Applied to the Mother Ratings of Validation Sample 1*

	Jinks and Fulker		Extended DeFries and Fulker				
Scale	β_1	β_2	b_1	b_2	b_3	b_4	b_5
Aggressive behavior	0.15 (0.02)	−0.02 (0.02)	0.37 (0.03)	−0.07 (0.03)	0.28 (0.02)	−0.13 (0.03)	0.05 (0.02)
Attention problems	0.14 (0.02)	−0.09 (0.02)	−0.33 (0.03)	−0.04 (0.04)	0.72 (0.03)	−0.17 (0.03)	0.07 (0.02)
Anxiety	0.15 (0.02)	−0.02 (0.02)	0.36 (0.02)	−0.07 (0.03)	0.30 (0.02)	−0.18 (0.03)	0.08 (0.02)
Dependency	0.03 (0.02)	−0.02 (0.02)	0.20 (0.03)	−0.06 (0.03)	0.45 (0.02)	−0.12 (0.03)	0.07 (0.02)
Emotional liability	0.13 (0.02)	−0.04 (0.02)	0.19 (0.03)	−0.14 (0.03)	0.40 (0.02)	−0.18 (0.03)	0.09 (0.02)
Physical coordination	0.13 (0.02)	−0.04 (0.02)	−0.07 (0.03)	−0.08 (0.03)	0.61 (0.02)	−0.12 (0.03)	0.05 (0.02)
Social isolation	0.30 (0.02)	−0.11 (0.02)	0.22 (0.03)	−0.01 (0.03)	0.35 (0.02)	−0.20 (0.04)	0.06 (0.02)

Note. The standard errors for the extended DeFries and Fulker regression approach are bootstrapped.

large data sets for its application, offers a breakthrough in the debate about the importance of G × E interaction.

As argued in this article, treating the genotype and environment as unmeasured variables has the advantage that the results do not depend on arbitrary measurement properties of genes and environment, while we can still identify the direction of the interaction and the nature of the environments that interact with genes (family or individual level). If we consider the results for the effects of the family and the individual, an interesting pattern is revealed. The individual and familial influences have different nonmonotonic relations with genetic predisposition. Specifically, it appears that individual environmental influences increase with increasing genetic predisposition where the strongest effect is evident above the average predisposition. At the highest levels of genetic predisposition, the environmental influences decrease. The familial environment shows a similar pattern but, interestingly, with the strongest effect below the average predisposition, thus in the opposite direction.

As our results indicate a nonmonotonic G × E interaction on child problem behavior, the results do not fully support either the bioecological, diathesis–stress, or differential susceptibility models as these models predict a monotonic relationship between the environment and genetic predisposition. However, the diathesis–stress model and differential susceptibility model generally describe the interaction of the genotype with the unique environment best, with more environmental variance at the upper range of genetic predisposition. However, the diathesis–stress model and the differential susceptibility model do not account for the decrease in unique environment at the extreme upper range of the genetic scores. For the common environment, the bioecological model generally describes the interaction of the genotype with the environment best, with more environmental variance at the lower range of genetic predisposition. However, the bioecological model does account for the decrease in common environment at the extreme lower range of the genetic scores.

Interestingly, from the review of Vendlinski et al. (2011) it appears that some studies into the Unmeasured Genotype × Measured Environment interaction on child problem behavior are consistent with the present results, while others are not. Specifically, results concerning the interaction of unmeasured genotype with family dysfunction (Button, Scourfield, Martin, Purcell, & McGuffin, 2005), maternal putative discipline (Button et al., 2008), and SES (Tuvblad, Grann, & Lichtenstein, 2006)—which we label as *familial* environment—all support the bioecological model, as in the present study. However, other studies investigating paternal putative discipline (Button et al., 2008), maternal putative discipline (Lau & Eley, 2008), and parental closeness (Miles, Silberg, Pickens, & Eaves, 2005) support the diathesis–stress model. For the *individual* environmental factors, most studies from the overview of Vendlinski et al. (2011) are consistent with the present results in that these studies found that variables such as best friend alcohol use (Harden, Hill, Turkheimer, & Emery, 2008) and stressful life events (Hicks, South, et al., 2009) interact with genotype in accordance with the diathesis–stress model. One study found for measures of antisocial peer affiliation and mother–child problems that a bioecological model was tenable, which is not consistent with the present results concerning the unique environment (Hicks, DiRago, Iacono, & McGue, 2009). Also, Middeldorp et al. (2014) found

that a bioecological model was tenable for child care and SES, while a diathesis–stress model was tenable for birth cohort interaction. Given that the current study investigates average effects, the results do not rule out opposite G × E interaction effects.

Testing for a G × E interaction is challenging, because of the requirement of large sample sizes and the risk of unwarranted conclusions due to scale dependencies. Therefore, the present results are encouraging as we accounted for scale properties by using newly developed methodology that has shown to be insensitive to scale properties, and found the same direction of the effect for all scales, in two cross-validation samples, and for both maternal and paternal ratings. These results, based in part on independent samples (the two cross-validation samples) and in part on samples that are not independent but have different raters assessing the child's behavior, indicate genuine G × E interactions on problem behavior in childhood.

We show that heritability is larger in the traditional ACE model that does not explicitly model G × E interaction. The bias is moderate for most phenotypes, but to obtain a deeper understanding of the etiology of childhood behavioral and emotional problems, it is important to decompose heritability into main and interaction effects.

The models used in this article do not take into account the possibility of Family × Individual Environment interactions (C × E). However, as shown by Molenaar et al. (2012), the presence of such an interaction has a relatively small effect on the results if the additive genetic factor is the largest source of variation in the data. Given that most of the scales in the present study have substantial heritabilities, Family × Individual Environment interactions are unlikely to affect the results. For the Problem Behavior and Anxiety scales, heritability is around 0.35. However, as the results of these scales are highly similar to the other scales, conflation by the Family × Individual Environment interactions is improbable.

Although in the present analysis we accounted for the most serious confounding factors in G × E interaction, there may be alternative explanations for the effects found. That is, the presence of a nonlinear G × E correlation may have conflated by the G × E interactions. For instance, if children with a high genetic predisposition seek out environments that stimulate aggressive behavior, but the relation between the two is nonlinear, a G × E interaction may result. Note that, as discussed earlier, a linear G × E interaction is not problematic. Another alternative explanation for the G × E interactions may arise if the twin sample is unrepresentative of the population (e.g., higher phenotypes are underrepresented). In such a case, skewness in the data will arise, which may be detected as G × E interaction. Note that the shortcomings discussed above are not unique to the present methodology; they are problematic for G × E interactions in general (e.g., the measured environment methodology of Purcell, 2002).

It is of interest to reflect upon the extent to which these results may have an implication for gene-finding studies and the ongoing debate about "missing heritability." We found that when G × E interactions are not accounted for, heritability estimates could be biased upward (between 1% and 34%). These are, in general, relatively small effects compared to the main effects of genotype. Therefore, the strategy of first identifying genetic effects and then identifying the interplay with the environment still seems to be the most promising. This particularly holds for environmental effects that are common in the population, since the effect of the genotype will also be detected when the G × E interaction is not modeled (Munafo, Durrant, Lewis, & Flint, 2009).

To conclude, G × E interaction effects are significantly present for problem behaviors in 5-year-old children. These results and the outcomes of similar future analyses at other ages could be used to guide studies investigating the interactions between specific genotypes or specific environments. Given the relatively small interaction effects, as compared to the main genetic effects, gene-finding studies will not greatly benefit from including these effects in the first stage. However, G × E interaction analyses are warranted in follow-up studies of established genetic variants to further elucidate the effects.

References

Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers: Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity. *Molecular Psychiatry*, *14*, 746–754. doi: 10.1038/mp.2009.44

Boomsma, D. I., de Geus, E. J., van Baal, G. C. M., & Koopmans, J. R. (1999). A religious upbringing reduces the influence of genetic factors on disinhibi-

tion: Evidence for interaction between genotype and environment on personality. *Twin Research*, *2*, 115–125. doi:10.1375/twin.2.2.115

Boomsma, D. I., Vink, J. M., van Beijsterveldt, C. E. M., De Geus, E. J. C., Beem, L., & Mulder, E. J. C. M. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research*, *5*, 1–6. doi:10.1375/twin.5.5.401

Brant, A. M., Munakata, Y., Boomsma, D. I., DeFries, J. C., Haworth, C. M., . . . Hewitt, J. K. (2013). The nature and nurture of high IQ an extended sensitive period for intellectual development. *Psychological Science*. doi:10.1177/0956797612473119

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective—A bioecological model. *Psychological Review*, *101*, 568–586. doi:10.1037/0033-295X.101.4.568

Button, T. M., Lau, J. Y. F., Maughan, B., & Eley, T. C. (2008). Parental punitive discipline, negative life events and gene–environment interplay in the development of externalizing behavior. *Psychological Medicine*, *38*, 29–39. doi:10.1017/S0033291707001328

Button, T. M., Scourfield, J., Martin, N., Purcell, S., & McGuffin, P. (2005). Family dysfunction interacts with genes in the causation of antisocial symptoms. *Behavior Genetics*, *35*, 115–120. doi:10.1007/s10519-004-0826-y

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., . . . Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851–854. doi:10.1126/science.1072290

Cherny, S. S., Cardon, L. R., Fulker, D. W., & DeFries, J. C. (1992). Differential heritability across levels of cognitive ability. *Behavior Genetics*, *22*, 153–162. doi:10.1007/BF01066994

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics*, *15*, 467–473. doi:10.1007/BF01066239

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene by environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049. doi:10.1176/appi.ajp.2011.11020191

Eaves, L. J. (2006). Genotype × Environment interaction in psychopathology: Fact or Artifact? *Twin Research and Human Genetics*, *9*, 1–8. doi:10.1375/twin.9.1.1

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, *30*, 1–42. doi: 10.1111/j.2044-8317.1977.tb00722.x

Faraone, S. V., Perlis, R. H., Doyle, A. E., Smoller, J. W., Goralnick, J. J., Holmgren, M. A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*, 1313–1323. doi:10.1016/j.biopsych.2004.11.024

Geschwind, D. H. (2009). Advances in autism. *Annual Review of Medicine*, *60*, 367–380. doi:10.1146/annurev.med.60.053107.121225

Gregory, A. M., & Eley, T. C. (2007). Genetic influences on anxiety in children: What we've learned and where we're heading. *Clinical Child and Family Psychology Review*, *10*, 199–212. doi:10.1007/s10567-007-0022-8

Haberstick, B. C., Lessem, J. M., Hopfer, C. J., Smolen, A., Ehringer, M. A., Timberlake, D., & Hewitt, J. K. (2005). Monoamine oxidase A (MAOA) and antisocial behaviors in the presence of childhood and adolescent maltreatment. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *135*, 59–64. doi:10.1002/ajmg.b.30176

Harden, K. P., Hill, J. E., Turkheimer, E., & Emery, R. E. (2008). Gene–environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. *Behavior Genetics*, *38*, 339–347. doi:10.1007/s10519-008-9202-7

Hicks, B. M., DiRago, A. C., Iacono, W. G., & McGue, M. (2009). Gene–environment interplay in internalizing disorders: Consistent findings across six environmental risk factors. *Journal of Child Psychology and Psychiatry*, *50*, 1309–1317. doi:10.1111/j.1469-7610.2009.02100.x

Hicks, B. M., South, S. C., DiRago, A. C., Iacono, W. G., & McGue, M. (2009). Environmental adversity and increasing genetic risk for externalizing disorders. *Archives of General Psychiatry*, *66*, 640–648. doi:10.1001/archgenpsychiatry.2008.554

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, mava, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*, 311–349. doi:10.1037/h0029135

Kan, K. J., Dolan, C. V., Nivard, M. G., Middeldorp, C. M., van Beijsterveldt, C. E., Willemsen, G., & Boomsma, D. I. (2013). Genetic and environmental stability in attention problems across the lifespan: Evidence from the Netherlands twin register. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 12–25. doi:10.1016/j.jaac.2012.10.009

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., & Moffitt, T. E. (2006). MAOA, maltreatment, and gene–environment interaction predicting children's mental health: New evidence and a meta-analysis. *Molecular Psychiatry*, *11*, 903–913. doi:10.1038/sj.mp.4001851

Lau, J. Y. F., & Eley, T. C. (2008). Disentangling gene–environment correlations and interactions on adolescent depressive symptoms. *Journal of Child Psychology and Psychiatry*, *49*, 142–150. doi:10.1111/j.1469-7610.2007.01803.x

Loeber, R., & Farrington, D. (1998). *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: Sage.

Middeldorp, C. M., Lamb, D. J., Vink, J. M., Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2014).

Child care, socio-economic status and problem behavior: A study of gene environment interaction in young Dutch twins. *Behavior Genetics*, *44*, 314–325. doi:10.1007/s10519-014-9660-z

Miles, D. R., Silberg, J. L., Pickens, R. W., & Eaves, L. J. (2005). Familial influences on alcohol use in adolescent female twins: Testing for genetic and environmental interactions. *Journal of Studies on Alcohol*, *66*, 445–451. doi:10.15288/jsa.2005.66.445

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene environment interplay in antisocial behaviors. *Psychological Bulletin*, *131*, 533–554. doi:10.1037/0033-2909.131.4.533

Molenaar, D., & Dolan, C. V. (2014). Testing systematic genotype by environment interactions using item level data. *Behavior Genetics*, *44*, 212–231. doi:10.1007/s10519-014-9647-9

Molenaar, D., van der Sluis, S., Boomsma, D. I., & Dolan, C. V. (2012). Detecting specific genotype by environment interactions using marginal maximum likelihood estimation in the classical twin design. *Behavior Genetics*, *42*, 483–499. doi:10.1007/s10519-011-9522-x

Munafo, M. R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene × Environment interactions at the serotonin transporter locus. *Biological Psychiatry*, *65*, 211–219. doi:10.1016/j.biopsych.2008.06.009

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond: Department of Psychiatry, Virginia Commonwealth University.

Nivard, M. G., Dolan, C. V., Kendler, K. S., Kan, K. J., Willemsen, G., van Beijsterveldt, C. E., . . . Boomsma, D. I. (2014). Stability in symptoms of anxiety and depression as a function of genotype and environment: A longitudinal twin study from ages 3 to 63 years. *Psychological Medicine*, *4*, 1–11. doi:10.1017/S00332917140 0213X

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460. doi:10.1007/BF02296207

Osborne, R. T. (1980). *Twins: Black and White*. Athens, GA: Foundation for Human Understanding.

Plomin, R., & Rutter, M. (1998). Child development, molecular genetics, and what to do with genes once they are found. *Child Development*, *69*, 1223–1242. doi:10.1111/j.1467-8624.1998.tb06169.x

Posthuma, D., Beem, A. L., De Geus, E. J., Van Baal, G. C. M., von Hjelmborg, J. B., Iachine, I., & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research*, *6*, 361–376. doi:10.1375/twin.6.5.361

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, *5*, 554–571. doi:10.1375/twin.5.6.554

Rende, R., & Plomin, R. (1992). Diathesis-stress models of psychopathology: A quantitative genetic perspective. *Applied and Preventive Psychology*, *1*, 177–182. doi:10.1016/S0962-1849(05)80123-4

Rhee, S. H., & Waldman, I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin*, *128*, 490. doi:10.1037/0033-2909.128.3.490

Rice, F. (2009). The genetics of depression in childhood and adolescence. *Current Psychiatry Reports*, *11*, 167–173. doi:10.1007/s11920-009-0026-9

Rutter, M., Moffitt, T., & Caspi, A. (2006). Gene–environment interplay and psychopathology: Multiple varieties but real effects. *Journal of Child Psychology and Psychiatry*, *47*, 226–261. doi:10.1111/j.1469-7610.2005.01557.x

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.

Schwabe, I., & van den Berg, S. M. (2014). Assessing genotype by environment interaction in case of heterogeneous measurement error. *Behavior Genetics*, *44*, 394–406. doi:10.1007/s10519-014-9649-7

Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene–environment interactions: Retrospect and prospect. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*, 65–76. doi:10.1093/geronb/60.Special_Issue_1

Spivack, G., & Spotts, J. (1966). *The Devereux Child Behavior (DCB) Rating Scale*. Devon, PA: Devereux Foundation.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*, 1097–1118. doi:10.1037/a0015864

Tucker-Drob, E. M., Harden, K. P., & Turkheimer, E. (2009). Combining nonlinear biometric and psychometric models of cognitive abilities. *Behavior Genetics*, *39*, 461–471. doi:10.1007/s10519-009-9288-6

Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: Gene–environment interaction. *Journal of Child Psychology and Psychiatry*, *47*, 734–743. doi:10.1111/j.1469-7610.2005.01552.x

Van Beijsterveldt, C. E. M., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., Lamb, D., . . . Boomsma, D. I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, *16*, 252. doi:10.1017/thg.2012.118

Van Beijsterveldt, C. E. M., Verhulst, F. C., Molenaar, P. C. M., & Boomsma, D. I. (2004). The genetic basis of problem behavior in 5-year-old Dutch twin pairs. *Behavior Genetics*, *34*, 229–242. doi:10.1023/B:BEGE.0000017869.30151.fd

Van den Berg, S. M., Glas, C. A., & Boomsma, D. I. (2007). Variance decomposition using an IRT measurement model. *Behavior Genetics*, *37*, 604–616. doi:10.1007/s10519-007-9156-1

Van der Sluis, S., Dolan, C. V., Neale, M. C., Boomsma, D. I., & Posthuma, D. (2006). Detecting genotype-environment interaction in monozygotic twin data: Comparing the Jinks & Fulker test and a new test based on marginal maximum likelihood estimation. *Twin Research and Human Genetics*, *9*, 377–392. doi:10.1375/twin.9.3.377

Vendlinski, M. K., Lemery-Chalfant, K., Essex, M. J., & Goldsmith, H. H. (2011). Genetic risk by experience interaction for childhood internalizing problems: Converging evidence across multiple methods. *Journal of Child Psychology and Psychiatry*, *52*, 607–618. doi:10.1111/j.1469-7610.2010.02343.x

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79. doi:10.1037/1082-989X.12.1.58

Young, S., Smolen, A., Hewitt, J., Haberstick, B., Stallings, M., Corley, R., & Crowley, T. (2006). Interaction between MAO-A genotype and maltreatment in the risk for conduct disorder: Failure to confirm in adolescent patients. *American Journal of Psychiatry*, *163*, 1019–1025.

Appendix: Absorption of Unmodeled A × E and A × C Interactions

In the case of the ACE model subject to A × E and A × C interactions, heritability is calculated as:

$$h_{\text{full}}{}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \widetilde{\sigma}_E^2 + \widetilde{\sigma}_C^2}$$

in which $\widetilde{\sigma}_E^2$ and $\widetilde{\sigma}_C^2$ denote the marginal variance of E and C, that is, the conditional variance of E and C (Equation 2 for the linear interaction model and Equation 3 for the curvilinear interaction model) with A marginalized out. For the curvilinear interaction model, these marginal variances can be calculated using

$$\widetilde{\sigma}_E^2 = \int_{-\infty}^{\infty} \exp(\beta_0 + \beta_1 A + \beta_2 A^2)\varphi(A)dA = \frac{\exp\left(\beta_0 + \frac{\beta_1^2}{2-4\beta_2}\right)}{\sqrt{1-2\beta_2}}$$

and

$$\widetilde{\sigma}_C^2 = \int_{-\infty}^{\infty} \exp(\gamma_0 + \gamma_1 A + \gamma_2 A^2)\varphi(A)dA = \frac{\exp\left(\gamma_0 + \frac{\gamma_1^2}{2-4\gamma_2}\right)}{\sqrt{1-2\gamma_2}}$$

where γ_2 and $\beta_2 < .5$ and $\varphi(.)$ is the standard normal density function. In the equations above, the marginal variances for the linear interaction model are obtained for $\gamma_2 = 0$ and $\beta_2 = 0$ for C and E, respectively. In addition, the heritability in the traditional ACE model is calculated as

$$h_{\text{ACE}}{}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \exp(\beta_0) + \exp(\gamma_0)}.$$

In the case of unmodeled A × E or A × C interactions, the interaction effects will be absorbed in $h_{\text{ACE}}{}^2$. To determine which part of $h_{\text{ACE}}{}^2$ is due to A × C and A × E interactions, we determined the proportion with which the heritability decreases or increases when a given interaction effect is added to the model. That is, the proportion of heritability in the traditional model that is due to A × C and A × E interactions is calculated as

$$p_{h^2} = 1 - \frac{h_{\text{full}}^2}{h_{\text{ACE}}^2}.$$