
Disentangling Genetic, Environmental, and Rater Effects on Internalizing and Externalizing Problem Behavior in 10-year-old Twins

Meike Bartels¹, Dorret I. Boomsma¹, James J. Hudziak², Marjolein J. H. Rietveld¹, Toos C. E. M. van Beijsterveldt¹, and Edwin J. C. G. van den Oord³

¹ Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands

² Departments of Psychiatry and Medicine (Division of Human Genetics), Center for Children, Youth and Families, University of Vermont, College of Medicine Burlington, Burlington, Vermont, USA

³ Virginia Institute of Psychiatric and Behavioral Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, Virginia, USA

Previous studies have emphasized the importance of rater issues in studying the etiology of variation in internalizing and externalizing problems in children. Earlier results indicate only moderate agreement between parents, and assume that parents assess a specific aspect of their child's behavior. In comparable samples of younger children, additive genetic effects are the main factor explaining individual differences in both internalizing and externalizing behavior. It is unknown whether this pattern of rater influences and variance decomposition will be consistent in older children. Child Behavior Checklists (Achenbach, 1991), completed by both parents, were collected in a sample of 2956 Dutch 10-year-old twin pairs. The etiology of individual differences in internalizing and externalizing syndromes was examined using a model that corrected for possible rater bias, rater-specific effects and unreliability. The best fitting model suggested that disagreement between the parents is not merely the result of unreliability and/or rater bias, but each parent also provides specific information from his/her own perspective on the child's behavior. Significant influences of additive genetic, shared environmental and unique environmental factors were found for internalizing and externalizing syndromes.

Parental descriptions are often used to collect information about a child's behavioral and emotional problems. A meta-analysis by Achenbach et al. (1987) showed a mean correlation of .60 between maternal and paternal ratings of the same child. The high interparent correlation shows that parents can provide meaningful information about their child's behavior, for if parental ratings would reflect nothing but error the correlations between their ratings would be close to zero. On the other hand, this interparent correlation

is less than perfect. This may be explained by different forms of rater bias and unreliability. Sources of rater bias are stereotyping, employing different normative standards, or having certain response styles (i.e., judging problem behaviors more or less severely). Because these types of bias may differ between raters, they may lead to disagreement between raters. Unreliability can become an important source of disagreement when raters cannot give an accurate description about relevant behaviors. For instance, evidence is found that parents may be relatively insensitive to affective disturbances in children (Angold et al., 1987). Another explanation for the less than perfect interparent correlation is that parents are not assessing exactly the same behavior in their children. It is known that different raters can provide, each from their own perspective, somewhat different but valid and complementary information about the child's functioning (Achenbach et al., 1987). Loeber and colleagues (1989), for instance, found that children's reports on their conduct problems tended to complement the information provided by adults.

It is difficult to draw conclusions about the processes underlying the (dis)agreement between parental ratings on the basis of the parental intercorrelations alone. Genetically informative data are helpful in this respect, allowing, due to their special properties, the evaluation of different hypotheses about the (dis)agreement in parental ratings. Models can be fitted to the data to test whether parental disagreement is caused by unreliability and rater bias, or involves the fact that parents provide specific information about

Received 27 August, 2003; accepted 20 November, 2003.

Address for correspondence: M. Bartels, Department of Biological Psychology, Vrije Universiteit, room 1F 66, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. Email: m.bartels@psy.vu.nl

their children's behavior. A correct representation is not only important from a substantive point of view, but also to obtain more accurate estimates of genetic and environmental effects. For instance, rater bias will cause shared environmental effects to be overestimated and measurement error will magnify the estimate of nonshared environmental effects. The use of multiple raters makes it possible to disentangle these rater effects from variance caused by the child's behavior so that parameter estimates are less biased and have a clearer interpretation.

To study agreement and disagreement between parental ratings, Hewitt et al. (1992) proposed so-called Rater Bias and Psychometric models that combine data of two raters and can be estimated using genetically informative data. The Rater Bias model assumes that parents assess the same behaviors in the child and have a common understanding of the behavioral descriptions. This may apply when both parents are equally confronted with the behaviors shown by the child (for instance at home). Disagreement between the raters is regarded as error, resulting from rater bias and/or unreliability. In addition to assessing similar aspects of the child's behavior, the Psychometric model assumes that each parent assesses specific aspects of the child's behavior. This will occur when the parent observes the child in distinct situations or is exposed to distinct samples of the child's behavior. For instance, the parent who usually brings the child to school may be more familiar with the child's behavior outside the home. Moreover, each parent may interact differently with the child (Achenbach et al., 1987). These unique interactions between a parent and a child may allow each parent to provide additional information about the child's behavior, apart from the information on which they both agree. Disagreement in this model does not merely arise from unreliability and/or rater bias, but also because each parent contributes, from his own perspective, different but valid information on the child's functioning. The psychometric model tests this possibility by examining whether there are significant genetic effects on the unique part of each parent's rating. If the behaviors uniquely rated by the parents are shown to be influenced by the genotype of the child, the parent must have been assessing a "real" but unique aspect of the child's behavior.

A number of quantitative genetic studies have used the Child Behavior Checklist (CBCL; Achenbach, 1991, 1992) to examine genetic and environmental effects on children's problem behaviors (Silberg et al., 1994; Edelbrock et al., 1995; Schmitz et al., 1995; Van den Oord et al., 1996; Zahn-Waxler et al., 1996; Gjone & Stevenson, 1997; Leve et al., 1998; Van der Valk et al., 1998a, 1998b; Hudziak et al., 2000). Yet, only a few studies employed models that incorporated rater differences. Rowe and Kandel (1997) administered the CBCL to mothers and fathers for their oldest two offspring (aged 9 to 17) in 76 families. The

subjects, however, were non-twin siblings rather than twins. Hence, estimation of separate genetic and environmental components of trait variance was impossible. The combination, though, of three informants (mother, father, self-report) and the rating of two children per family, allowed the authors to disentangle rater effects from variance caused by a common understanding of the behavioral description in parents. Their models demonstrated that mother and father ratings contained a substantial individual view component, but parents also assessed similar aspects of the child's behavior. Hewitt and colleagues (1992) fitted Rater Bias and Psychometric models to parental ratings of the Internalizing scale (CBCL) for 983 twin pairs. They found that both for their prepubertal cohort (8 to 11 years) and for their pubertal cohort (12 to 16 years) the Psychometric model fitted the data better than the Rater Bias model. Van der Valk et al. (2001, 2003) also found that the Psychometric model fitted their data significantly better than the Rater Bias model at both ages 3 and 7. Thus these studies indicated that disagreement between parental ratings is partly caused by mothers and fathers assessing different aspects of the child's behavior.

In the present study we fitted Rater Bias and Psychometric models to data for the Internalizing and Externalizing scale of the CBCL. The sample consisted of 2956 Dutch 10-year-old twin pairs. The first aim of this study was to fit Rater Bias and Psychometric models. Results of previous studies in comparable samples of Dutch twins indicated a Psychometric model as best fitting model. Achenbach et al. (1987) observed, however, that the correlation between similar informants (e.g., parents) decreased with age of the child. One explanation is a decrease in the quality of parent ratings. Parents mainly interact with their children in the home environment. However, as children become older other social contexts such as school and the peer group become relatively more important. Consequently, it becomes more difficult for parents to assess problem behaviors in their children. Such a possible decrease could result in a better fit of the Rater Bias model compared to the Psychometric model. Another explanation for lower parental agreement in older children is that parent-child relations become more individual and specialized over the years. For instance, the roles of mother and father may become more differentiated and they may engage in different activities with their children. Such a change would imply that the rater-specific view of the parent increases. This specialization would suggest, in line with the findings at age 3 and 7, that a Psychometric model would fit best at age 10.

A second aim was to use the best fitting model to estimate influences of genetic and environmental components on internalizing and externalizing problem behavior at age 10. Comparison of the results of this study to the results of comparable studies in 3- and 7-year-old Dutch twins gives the opportunity to

disentangle real behavioral development from changes in rater effects. The large sample of twin pairs used provided the power necessary to be able to detect possible small changes.

Material and Method

Subjects

All participants were registered by the Netherlands Twin Registry (NTR), kept by the Department of Biological Psychology at the Vrije Universiteit in Amsterdam. Of all multiple births in the Netherlands, 40–50% are registered by the NTR (Boomsma et al., 1992, 1998). For this study, data from twins from the birth cohorts 1986–1991 were used. Questionnaires were mailed to families within 3 months of the twins' 10th birthday. After 2 to 3 months, reminders were sent and 4 months after the initial mailing, persistent non-responders were contacted by phone. Families whose addresses were not available were included in the nonresponse group. Of the parents who rated their twins' behavior at age 3 (the first assessment age), 60% still participated in the study after a 7-year interval (at age 10). One hundred and forty twin pairs were excluded because either one or both of the children had a disease or handicap that interfered severely with daily functioning at age 10 or at a younger age. Finally, the analyzed sample consists of 2956 mother ratings and 2234 father ratings.

Zygoty was determined for 620 same-sex twin pairs by DNA or blood group polymorphisms. For all other same-sex twin pairs, zygoty was determined using discriminant analysis to relate questionnaire items to zygoty based on blood/DNA typing. Parents were asked how much the twins resembled each other in facial structure, hair color, facial color, eye color, and whether they were ever mistaken for each other by the parents themselves, by family, or by strangers. They were also asked if the twins were as much alike as "two peas in a pod", whether it was difficult for the parents to separate the twins on a recent picture, and how similar the twins' hair structure was. The zygoty was correctly classified by questionnaire in nearly 95% of the cases (Rietveld et al., 2000).

This left a sample of 519 monozygotic males (MZM), 471 dizygotic males (DZM), 618 monozygotic females (MZF), 458 dizygotic females (DZF), and 890 dizygotic opposite sex (DOS) twin pairs. In general, mothers' response rate out-numbered fathers' response rate. Therefore, the data could be further divided into twin pairs for which both mother and father had replied (400 MZM, 347 DZM, 470 MZF, 348 DZF, and 669 DOS) and twin pairs for which only mothers had replied (119 MZM, 124 DZM, 148 MZF, 110 DZF, and 221 DOS). Because of a relative small amount of families from which only fathers replied ($N = 28$) these families were not used in the analyses.

Socioeconomic status (SES) was obtained from a full description of the occupation of the parents when the children were 3 years of age. The level of

occupation was coded according to the system used by Statistics Netherlands (CBS, 1993). The code was based on the mental complexity of the work and ranged from low skilled to scientific work. An earlier comparison of the parental SES distribution with those obtained for the general Dutch population showed a slightly higher frequency of the middle and higher SES groups (for details see Rietveld et al., 2003).

Measures

The Child Behavior Checklist (CBCL 4–18; Achenbach, 1991) was developed for parents to score the behavioral and emotional problems of their 4- to 18-year-old children. It consists of 120 problem items that are scored by the parents on a 3-point scale based on the occurrence of the behavior during the preceding 6 months: 0 if the problem item was not true, 1 if the item was somewhat or sometimes true, and 2 if it was very true or often true. The syndrome scales were composed according to the 1991 profile (Achenbach, 1991). Dutch syndrome scales and comparability with the syndrome scales as developed by Achenbach are reported in Verhulst et al. (1996). In this manual the two broadband scales Internalizing (INT) and Externalizing (EXT) are analyzed. The Internalizing scale consists of the Anxious/Depressed, Somatic Complaints and Withdrawn subscales. The Externalizing scale consists of the Aggressive and Rule Breaking Behavior subscales. For the Internalizing scale, subjects were only included if not more than three items were missing for the Anxious/Depressed scale, and not more than two items were missing for Somatic Complaints and Withdrawn scales. For the Externalizing scale the inclusion criterion was not more than three items missing from the Aggressive and Rule Breaking Behavior scales. This ensured that the two syndrome scales were always composed of all problem behaviors loading on that scale.

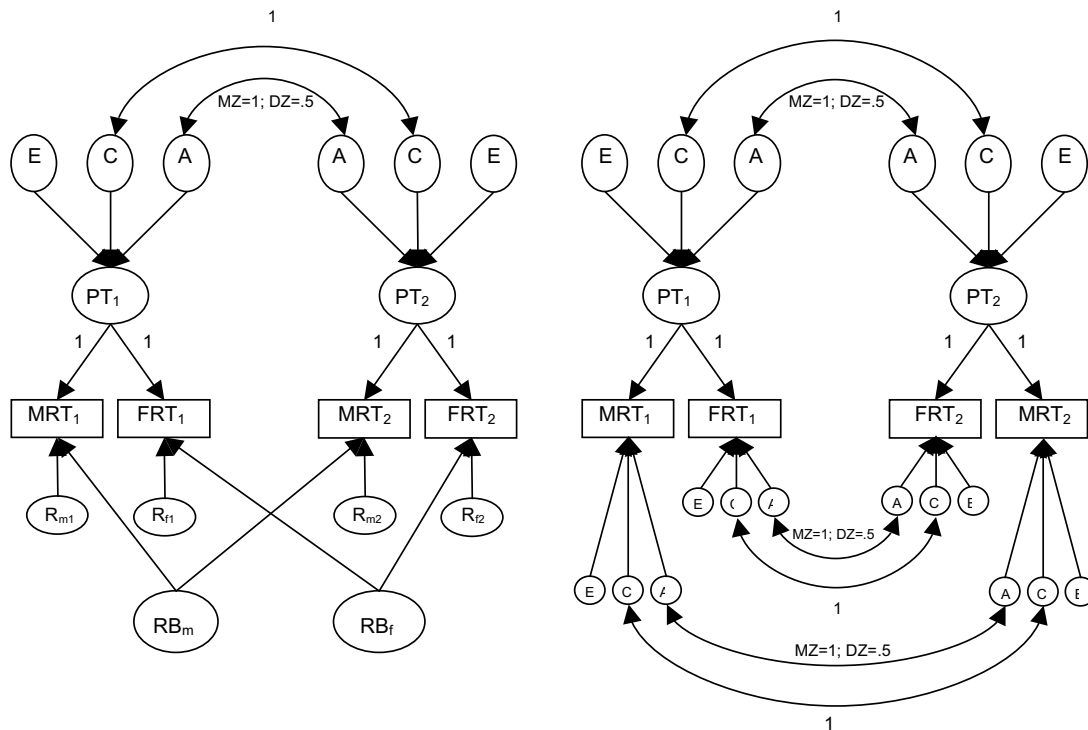
The data were square root transformed to approximate normal distributions that are required for maximum likelihood estimation. After transformation, all skewness and kurtosis indices were between -1.0 and 1.0 , implying that not much distortion is to be expected (Muthén & Kaplan, 1985).

Data Analyses

Descriptive statistics and prevalence of internalizing and externalizing behavior were calculated using SPSS/windows 10. Pearson correlations were used to calculate twin correlations and to calculate the inter-parent correlations. Significance of difference in means for boys and girls were tested using ANOVA for the oldest and the youngest of the twin pair separately. Differences in means based on mother or father ratings were tested using a paired-samples *t* test for the oldest and the youngest of the twin pair separately.

Structural Equation Modeling of Data from Twins Rated by More than One Rater

Data from monozygotic and dizygotic twins were used to decompose the variance in scores on the



The Rater Bias Model

Rater bias model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares: MRT₁, FRT₁, MRT₂, FRT₂) are linear functions of the latent phenotypes of the twins (PT₁ and PT₂), mother's and father's bias (RB_m, RB_f), and residual error (R_{m1}, R_{f1}, R_{m2}, R_{f2}). Latent phenotypes of the twins (PT₁ and PT₂) are influenced by A (additive genetic factors), C (shared environmental factors), and E (nonshared environmental factors).

The Psychometric Model

Psychometric model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares: MRT₁, FRT₁, MRT₂, FRT₂) are linear functions of the latent phenotypes of the twins (PT₁ and PT₂) and rater-specific variance. Latent phenotypes of the twins are influenced by common (i.e., across both parents) A (common genetic factor), C (common shared environmental factor), and E (common nonshared environmental factor). Rater-specific variance is made up of rater-specific (i.e., unique to each parent) A (rater-specific genetic factor), C (rater-specific shared environmental factor and/or rater bias), E (rater-specific nonshared environmental factor and/or measurement error/unreliability).

Figure 1
Models for multiple rater data.

Internalizing and Externalizing scales into a contribution of the additive effects of many genes, environmental influences that are shared by twins (like style of parenting, socioeconomic level, or religion) and environmental influences that are not shared by twins (such as an illness, relationships with peers, or measurement errors). For a summary of the twin method, the various assumptions, and the plausibility of these assumptions see Martin and Eaves (1977), Eaves (1982), Kendler and Eaves (1986), Falconer (1989), and Neale and Cardon (1992).

In the Rater Bias model (Hewitt et al., 1992; Figure 1, left part) the phenotypes of the twins are a function of three common factors underlying the ratings of both mothers *and* fathers: a genetic factor

(A), a shared environmental factor (C), and a non-shared environmental factor (E). In addition to these three common factors, rater-specific factors are modeled: a maternal rater bias factor, a paternal rater bias factor, and residual (unreliability) factors affecting each rating. The influence of the common factors is assumed to be independent of the maternal and paternal rater bias and unreliability factors.

The Psychometric model (Hewitt et al., 1992; Figure 1, right part) also estimates the influence of a genetic (A), a shared environmental (C), and a non-shared environmental factor (E) common to the phenotypes of the twins as rated by both parents. In addition, three rater-specific factors, a genetic (A_{mf}), shared environmental (C_{mf}), and nonshared

environmental factor (E_{mf}) are estimated for the ratings of mother/father. Disagreement between parents in this model can be caused by rater-specific behavioral views, leading to different but valid information of each rater. These rater-specific behavioral views can have their own unique influences, estimated in the rater-specific additive genetic, shared environmental, and nonshared environmental factors. Disagreements can also be caused by rater bias, which will confound the rater-specific shared environmental effects, or by unreliability, which will confound the rater-specific nonshared environmental effects. The three common factors loading on the twins' phenotypes contain only reliable variance, causing the common nonshared environmental factor to contain only pure independent environmental effects (McArdle & Goldsmith, 1990) and the common shared environmental factor to contain only pure shared environmental effects.

Model Fitting

The program Mx (Neale et al., 1999) was used to analyze the data through a simultaneous analysis of the 4×4 variance-covariance matrices in the five zygosity by sex twin groups (MZM, DZM, MZF, DZF, DOS) where both mother and father ratings were available, and the 2×2 variance-covariance matrices in the five zygosity by sex twin groups with only mother ratings. Estimates for male and female twins were allowed to differ. The model describes the observed variance-covariance matrices adequately when the residual variance-covariance matrices are trivially small. A good model is indicated by a low non-significant χ^2 test statistic ($p > .05$). Apart from the χ^2 test statistic, Akaike's Information Criterion ($AIC = \chi^2 - 2 \times \text{degrees of freedom}$) was computed. The lower the AIC the better the fit of the model to the observed data.

The best fitting full model was further examined for possible simplifications. It was tested whether the common and/or rater-specific factors could be removed from the model, whether estimates for boys and girls could be constrained to be the same, and

if the rater-specific factors for mothers and fathers could be constrained to be equal. The only factor that was never dropped from the model was the rater-specific nonshared environmental factor, because measurement errors are estimated in this factor.

Results

Description of the Data

The untransformed mean problem scores, standard deviations and prevalence of internalizing and externalizing behavior at age 10 are given in Table 1. Significance tests showed that boys did receive higher mother and father ratings than girls for the Externalizing scale (mother ratings: $F(1, 2580) = 86.73, p = .00$; father ratings: $F(1, 1967) = 55.66, p = .00$). For this same scale, mothers gave higher ratings to their children than fathers did ($t = 8.826, df = 1956, p = .00$), implying possible rater differences. For the Internalizing scale mothers gave higher ratings to their twin children than fathers did ($t = 13.84, df = 1938, p = .00$), implying possible rater differences for this scale as well. The homogeneity of the variance was tested with Mx (Neale et al., 1999). No differences could be found in the variances of MZM, DZM, MZF, DZF, and DOS for the Internalizing scale. For the Externalizing scale MZM variance is equal to DZM variance and MZF variance is equal to DZF variance, however the variance for boys and girls, both MZ and DZ could not be set equal.

Twin Correlations

Table 2 shows, for both the Internalizing and Externalizing scale, in the first and second columns the correlations between the twins rated by the same rater (mother or father rated both children), and in the third and fourth columns the cross-correlations between the twins each rated by a different rater (mother and father each rated one child). In the fifth and sixth columns the interparent correlations between mothers and fathers are given, both for first and second-born twin. The interparent correlations were comparable for both first and second-born twin

Table 1

Means (Standard Deviations), Sample Sizes, and Prevalence ($T \geq 65$) for the Internalizing and Externalizing Scale, in 10-year-old Twins (per Zygosity)

| | | INTERNALIZING | | | | EXTERNALIZING | | | | | |
|---|-----|---------------|-------------|----------|-------------------|---------------|-------------|-------------|-------------------|-------|-------|
| | | Mother | Father | N(M/F) | Prevalence M F | Mother | Father | N(M/F) | Prevalence M F | | |
| ♂ | MZM | 4.59 (4.60) | 3.50 (3.83) | 1008/796 | 5.5% | 4.8% | 8.67 (7.43) | 7.11 (6.65) | 1029/798 | 11.2% | 11.7% |
| | DZM | 5.18 (5.44) | 4.14 (4.57) | 933/691 | | | 8.03 (7.34) | 6.90 (6.31) | 937/692 | | |
| | DOS | 4.66 (4.97) | 3.56 (4.14) | 877/665 | | | 7.69 (7.10) | 6.51 (6.25) | 885/667 | | |
| ♀ | MZF | 5.12 (5.00) | 3.56 (3.95) | 1216/943 | 9.1% | 7.8% | 5.91 (5.53) | 4.81 (4.94) | 1223/939 | 4.0% | 5.2% |
| | DZF | 5.35 (5.35) | 4.30 (4.86) | 893/688 | | | 5.90 (5.86) | 5.04 (5.13) | 905/696 | | |
| | DOS | 4.76 (4.91) | 3.56 (3.70) | 878/662 | | | 5.34 (5.37) | 4.50 (4.79) | 885/667 | | |

Note: MZM/DZM = Monozygotic/Dizygotic males, MZF/DZF = Monozygotic/Dizygotic females, DOS = Dizygotic opposite sex, N children M/F = number of children for Mothers (M) and Fathers (F).

Table 2

Correlations (Ratings Given by the Same Rater), and Cross-correlations (Ratings Given by Different Raters) Between the Twins and the Interparent Correlations, per Zygosity, for 10-year-olds

| | Internalizing | | | | | | Externalizing | | | | | |
|-----|------------------|-----|------------------|-----|-------------|-----|------------------|-----|------------------|-----|-------------|-----|
| | same rater twins | | different raters | | | | same rater twins | | different raters | | | |
| | M/M | F/F | twins | | interparent | | M/M | F/F | twins | | interparent | |
| | | | M/F | F/M | O | Y | | | M/F | F/M | O | Y |
| MZM | .66 | .66 | .40 | .46 | .64 | .63 | .86 | .86 | .67 | .66 | .77 | .73 |
| DZM | .48 | .49 | .26 | .36 | .67 | .63 | .54 | .57 | .39 | .36 | .70 | .74 |
| MZF | .68 | .74 | .40 | .40 | .57 | .51 | .81 | .83 | .61 | .56 | .71 | .67 |
| DZF | .51 | .62 | .43 | .37 | .71 | .66 | .58 | .57 | .42 | .35 | .70 | .72 |
| DOS | .53 | .53 | .31 | .35 | .67 | .60 | .50 | .56 | .38 | .40 | .78 | .74 |

Note: MZM/DZM = monozygotic/dizygotic males, MZF/DZF = monozygotic/dizygotic females. DOS = dizygotic opposite sex twins. Same rater twins = correlation between the oldest and the youngest twin, rated by M/M = mothers or F/F = fathers. Different raters twins = cross-correlation: either oldest twin rated by mothers and youngest by fathers (M/F) or the other way around (F/M). Different raters interparent: O = correlation between mother and father ratings for the oldest child; Y = idem for the youngest child.

for all zygosity by sex groups. On average, the interparent correlations for the Internalizing scale were .63, and for the Externalizing scale .73.

The correlations between the first and second-born twin both rated by mothers (M/M, first column) and those both rated by fathers (F/F, second column) can be used to obtain a first estimate of the genetic influences (h^2), the shared environmental influences (c^2), and the nonshared environmental influences (e^2) on the total variance. For instance, if we take for the Internalizing scale the first column M/M: the genetic influences for boys can be estimated as $(r_{MZM} - r_{DZM}) \times 2 = (.66 - .48) \times 2 = .36$. Nonshared environmental influences for boys can be estimated as $(1 - r_{MZM}) = (1 - .66) = .34$, and the shared environmental influences for boys can be estimated as $(2 \times r_{DZM}) - r_{MZM} = (2 \times .48) - .66 = .30$. For girls, the correlations between the MZ and DZ twin pairs can be compared in similar ways to obtain a first impression of the genetic and environmental influences.

The obtained results of fitting univariate models (mother and father ratings separately), that estimated three factors: A, C, and E and possible sex differences, were comparable to those expected by comparing the MZ and DZ correlations. Individual differences in both internalizing and externalizing behavior can be explained by additive genetic (INT 37–39%, EXT about 48–69%), shared environmental (INT 29–32%, EXT 16–33%) and nonshared environmental influences (INT 25–32%, EXT 14–19%). Significant sex differences in the strength of genetic and environmental influences on externalizing behavior were found, with higher estimates of genetic influences for boys (full model fitting results are available on request by the author). The sex differences imply only a difference in the strength of the additive genetic effect and no real heterogeneity. Influences of different genes in boys and girls would be represented by lower DOS correlations in comparison to DZ correlations in same-sex twins. In this study, the DOS correlation for

externalizing behavior is not different from the DZ correlations (see Table 2).

Univariate analyses make a decomposition of the total variance in genetic, shared environmental, and nonshared environmental factors. To take rater differences into account, the information from the twin's cross-correlations has to be used. By calculating cross-correlations between mother ratings of oldest twins with father ratings of youngest twins (M/F, third column) or the other way around (F/M, fourth column), one can make a decomposition of the variance on which both kinds of raters agree. The difference between the decomposition of the variance shared between raters (i.e., common view) and the decomposition of the total variance can be used to estimate the genetic, shared environmental, and nonshared environmental influences on the variance uniquely rated by one particular rater (i.e., rater-specific view). For example, take for the Internalizing scale the cross-correlations between mother ratings of oldest twins and father ratings of youngest twins (M/F) for boys. The same comparisons between the r_{MZ} and r_{DZ} can be made to estimate the genetic influences on the variance shared by raters, namely $2 \times (r_{MZM-cross} - r_{DZM-cross}) = (.40 - .26) \times 2 = .28$. Thus we can conclude that the total genetic variance of 36% can be divided into a genetic influence for behaviors that are similarly rated by the parents of 28% and a genetic influence for behaviors that are uniquely rated by mothers of 8%. This shows that genes of the child effect the rater-specific part of the maternal ratings, implying that the parental disagreement is not merely caused by measurement errors but that mothers, in addition to the common view, also assess a valid rater-specific part of their child's behavior. Finding genetic influences for behaviors that are differently rated by mothers and fathers does not seem to be a chance finding, but arises systematically in the data. Additionally, for the father ratings of boys and for the mother and father ratings of girls, both for the

Internalizing and Externalizing scale, similar rater-specific genetic effects were found.

To estimate the environmental influences on the variance shared by raters the interparent correlations (fifth and sixth columns for oldest and youngest twin, respectively) have to be used. Table 2 shows that for the Internalizing scale the interparent correlation (between mothers and fathers of the same child) in the MZM group was .64 for the oldest twin. The cross-correlation (between mothers and fathers of different children) was .40, indicating a nonshared environmental contribution on the variance shared by raters of: interparent correlation - $r_{mzm-cross} = .64 - .40 = .14$. Thus the nonshared environmental influences can be divided into an influence for behaviors that are similarly rated by both parents of 14% and an influence for behaviors that are uniquely rated by mothers of 20% (i.e., 34% - 14%). Shared environmental influences on the variance shared by raters can be estimated as $(2 \times r_{DZM}) - r_{MZM} = (2 \times .26) - .40 = .12$. Taking rater differences into account, the shared envi-

ronmental influences can be divided into an influence for behaviors that are similarly rated by the parents of 12% and an influence for behaviors that are differently rated by mothers of 18% (i.e., 30%-12%). For the cross-correlations of father ratings for boys, mother and father ratings for girls, and all ratings of the Externalizing scale, similar comparisons can be made.

Rater Models

As indicated by the lower χ^2 test statistic and the lower AIC in Table 3, the Psychometric model fitted the data better than the Rater Bias model both for the Internalizing and the Externalizing scales. This signified that although both parents partially assessed the same behaviors, there also was a component, which was unique to each rater. For sake of comparison we also performed a Cholesky or triangular decomposition (also called a Biometric model). This model can be viewed as a less psychologically informative rotation of the Psychometric model (Hewitt et al., 1992). Neither for the Internalizing scale nor for the

Table 3

Model Fitting Statistics for Psychometric and Rater Bias Models and Simplification of the Psychometric Model, for 10-year-old Twins' Internalizing and Externalizing Problems

| | χ^2 (95% CI) | df | p | AIC (95% CI) | $\Delta\chi^2$ | Δdf | p |
|------------------------------------|--------------------------|----|------|--------------------------|----------------|-------------|------|
| Internalizing | | | | | | | |
| Overall model: | | | | | | | |
| Psychometric model | 75.41 (55.58 – 103.147) | 47 | .005 | -18.59 (-38.43 – 9.15) | | | |
| Rater Bias model | 87.54 (65.26 – 117.66) | 49 | .001 | -10.46 (-32.74 – 19.66) | | | |
| Simplification overall model: | | | | | | | |
| <i>Factor estimates:</i> | | | | | | | |
| No common genetic effects | 132.93 (102.43 – 171.08) | 49 | .000 | 34.93 (4.43 – 73.08) | 57.52 | 2 | .000 |
| No unique genetic effects | 97.91 (73.74 – 129.83) | 51 | .000 | -4.09 (-28.26 – 27.83) | 22.50 | 4 | .000 |
| No common shared env. | 109.54 (82.992 – 143.90) | 49 | .000 | 11.54 (-15.00 – 45.90) | 34.13 | 2 | .000 |
| No unique shared env. | 150.94 (117.85 – 191.62) | 51 | .000 | 48.94 (15.85 – 89.62) | 75.54 | 4 | .000 |
| No common nonshared env. | 491.78 (424.36 – 566.66) | 49 | .000 | 393.37 (326.36 – 468.66) | 416.37 | 2 | .000 |
| <i>Sex differences:</i> | | | | | | | |
| No sex dif. common effects | 83.42 (62.14 – 112.59) | 50 | .000 | -16.58 (-37.86 – 12.59) | 8.02 | 3 | .046 |
| No sex dif. unique effects | 84.37 (63.22 – 113.43) | 53 | .000 | -21.63 (-42.78 – 7.43) | 8.96 | 6 | .176 |
| No sex dif. common + unique | 93.36 (70.63 – 124.11) | 56 | .000 | -18.64 (-41.37 – 12.11) | 17.96 | 9 | .036 |
| <i>Rater differences:</i> | | | | | | | |
| Unique rater effect: M-F identical | 130.18 (100.46 – 167.60) | 53 | .000 | 24.18 (-5.54 – 61.60) | 54.78 | 6 | .000 |
| Externalizing | | | | | | | |
| Overall model: | | | | | | | |
| Psychometric model | 55.68 (47.0 – 78.53) | 47 | .098 | -38.32 (-47.0 – 15.47) | | | |
| Rater Bias model | 113.99 (86.65 – 149.04) | 49 | .000 | 15.99 (-11.35 – 51.04) | | | |
| Simplification overall model: | | | | | | | |
| <i>Factor estimates:</i> | | | | | | | |
| No common genetic effects | 361.48 (304.79 – 425.66) | 49 | .000 | 263.50 (206.79 – 327.66) | 305.80 | 2 | .000 |
| No unique genetic effects | 128.46 (98.85 – 165.75) | 51 | .000 | 26.46 (-3.15 – 63.75) | 72.78 | 4 | .000 |
| No common shared env. | 79.86 (59.25 – 108.38) | 49 | .000 | -18.14 (-38.75 – 10.38) | 24.18 | 2 | .000 |
| No unique shared env. | 176.25 (139.60 – 220.50) | 51 | .000 | 74.25 (37.60 – 118.50) | 120.57 | 4 | .000 |
| No common nonshared env. | 513.55 (444.49-590.07) | 49 | .000 | 415.60 (346.49-492.07) | 457.87 | 2 | .000 |
| <i>Sex differences:</i> | | | | | | | |
| No sex dif. common effects | 85.34 (63.64 – 114.91) | 50 | .000 | -14.66 (-36.36 – 14.91) | 29.66 | 3 | .000 |
| No sex dif. unique effects | 66.35 (53.00 – 91.28) | 53 | .100 | -39.65 (-53.00 – 14.72) | 10.68 | 6 | .099 |
| No sex dif. common + unique | 99.53 (75.56 – 131.35) | 56 | .000 | -12.47 (-36.44 – 19.33) | 43.86 | 9 | .000 |
| <i>Rater differences:</i> | | | | | | | |
| Unique rater effect: M-F identical | 79.29 (59.29 – 107.27) | 53 | .010 | -26.71 (-46.71 – 1.26) | 23.61 | 6 | .001 |

Table 4

Genetic and Environmental Influences, Estimated Using Best Fitting Psychometric Model, for Internalizing and Externalizing Problems Rated for 10-year-old Twins

| | Internalizing | | Externalizing | | | |
|-----------|---------------|--------|---------------|-----|-----|-----|
| | Mother | Father | ♂ | ♀ | ♂ | ♀ |
| Genetic | | | | | | |
| Common | 26% | 30% | 54% | 41% | 56% | 43% |
| Unique | 10% | 8% | 11% | 14% | 3% | 3% |
| Shared | | | | | | |
| Common | 18% | 20% | 13% | 18% | 13% | 19% |
| Unique | 14% | 15% | 7% | 9% | 14% | 17% |
| Nonshared | | | | | | |
| Common | 16% | 17% | 8% | 10% | 8% | 11% |
| Unique | 16% | 10% | 7% | 8% | 6% | 7% |

Externalizing scale did this saturated model fit the data any better than the Psychometric model.

The Psychometric model was further examined for possible simplifications. None of the common and rater-specific genetic, shared and nonshared environmental factors could be dropped from the model. Between boys and girls, the estimates of the common and the rater-specific factors could be constrained to be equal for the Internalizing scale. For the Externalizing scale only the rater-specific effects could be set equal for boys and girls. Mother and father ratings could not be constrained to be equal for both scales. The fit results are given in Table 3.

The percentages of variance explained by the common and rater-specific genetic, shared, and nonshared environmental factors are given in Table 4. A major part of the variance was explained by common factors. For both the Internalizing and the Externalizing scales the largest part of the variance was explained by the common genetic factor. Common additive genetic effects explain around 30% of the variance in internalizing behavior in boys and girls. For Externalizing behavior, sex differences in the strength of the common genetic influence were found, explaining 55% of the variance in boys and 40% of the variance in girls. The common nonshared environmental factor explained 15% of the variance for the Internalizing scale and around 10% for the Externalizing scale. The common shared environmental factor explained around 18% of the variance for both the Internalizing scale in boys and girls and Externalizing scale in girls. For externalizing behavior in boys, only 13% of the variance is explained by common shared environmental factors. The rater-specific factors explained a relatively small part of the variance. For the Internalizing scale rater-specific genetic factors explained 9%, rater-specific shared environmental factors explained 15%, and rater-specific nonshared environmental factors explained around 15% of the variance. For the Externalizing scale, rater-specific factors also explained relatively small parts of the

variance: 12% genetic influence, 8% shared, and 8% nonshared environmental influences based on mother ratings, and 3% genetic influence, 15% shared, and 6% nonshared environmental influences based on father ratings.

Discussion

Models for Parental (Dis)Agreement

In a sample of 2956 Dutch 10-year-old twin pairs, we studied genetic and environmental influences on internalizing and externalizing problems, while taking the processes underlying agreement and disagreement between maternal and paternal ratings into account. The Psychometric model fitted the data better than the Rater Bias model for both scales. This implied that rater differences did not merely reflect measurement error, but were also the result of parents assessing different aspects of the child's behavior. These results are in accordance with previous studies (Hewitt et al., 1992; Van der Valk et al., 2001, 2003) suggesting that each parent provides additional information about the child's behavior.

For internalizing behavior a decrease in the relative importance of common effect (A, C and E on the behavior similarly assessed by both parents) versus rater-specific effects was observed over the years, representing a decrease in interparent correlation, as suggested by Achenbach and colleagues (1987). At age 3, Van der Valk et al. (2001) reported on common effects explaining 73% of the total variance in internalizing problem behavior, while in this study at age 10 only 64% of the total variance in internalizing problem behavior is explained by these common factors. The better fit of the Psychometric model suggests that individualization and specialization of the parent-child relation, instead of a decline in the quality of parent ratings, was the underlying cause of the decrease in parental agreement. For externalizing problem behavior, less change in interparent correlation was observed over the years (.67 on average at age 3 [Van der Valk et al., 2001]; .67 in this study).

The meta-analyses by Achenbach and colleagues (1987) reported more consistency in parental agreement for undercontrolled problems (externalizing behavior) versus overcontrolled problems (internalizing behavior); however this was not significant for mother/father pairs. A possible explanation, though, for the stability in parental agreement for externalizing behavior could be that these types of behavior are better observable for an external rater than internalizing problem behaviors and is in that manner less vulnerable to the suggested specialization or individualization of the parent-child relation.

Common Aspects of Parental Ratings

The common A, C, and E factors represent the genetic and environmental influences on the child's behavior similarly assessed by both parents. This part of the behavioral ratings is not affected by measurement error, rater bias or rater-specific views and represents a reliable measure of internalizing and externalizing problem behavior. Common additive genetic influences, common shared environmental influences, and common nonshared environmental influences explain about 30%, 20%, and 16% of the total variance of internalizing behavior, respectively. For internalizing behavior, no sex differences were found over the years. In comparison to previous studies by Van der Valk and colleagues (2001, 2003), the relative importance of the additive genetic effects decrease from age 3 to 7, but remain about the same from age 7 to age 10. An increase of shared environmental influences was found. At age 3 shared environmental influences were absent, while at age 7 and age 10 shared environmental influences were significant. An explanation for the presence of change between age 3 and 7 and the absence of changes between age 7 and 10 could be that the 3-7 year age interval includes children's transition to school. During this transition children must cope with many new demands like meeting academic challenges, learning school and teacher expectations, and adjusting to the daily routine of a school class (Barth & Parke, 1993; Cowan et al., 1994; Ladd & Price, 1987). An important aspect of this transition is the development of social relations with other children (Asher, 1990; Schneider, 1993). Although multiple pathways can be involved, poor relations with peers have shown to be a powerful predictor of behavior and emotional problems later in life.

Another possibility for the difference in the magnitude of shared environmental influences on internalizing behavior at age 3 versus age 7 and 10 is that the CBCL/2-3 (Achenbach, 1992) Internalizing scale taps somewhat different behaviors than their counterparts in the CBCL/4-18. In very young children the Internalizing scale could more strongly reflect temperamental factors, whereas in older children it might be more closely related to affective symptoms. For externalizing problems that involve more readily observable overt behaviors, the scales have clearer counterparts for preschool versus school-age children

(Koot et al., 1997) and thus we did not observe a similar age difference for this scale.

One explanation for a real increase in shared environmental influences on problem behavior is that if parents are only able to guide the child's behavior when he/she is able to understand other people's values and can direct its behavior accordingly, shared environmental influences are more likely to be found in older children. However, it may be important to realize that shared environment is not necessarily confined to the home environment. For instance, there are indications that these environmental effects are not merely shared by siblings but also by cousins (Van den Oord & Rowe, 1998, 1999). This suggests that shared environment reflects the wider community in which families are embedded as well (Bronfenbrenner, 1979; Parke & Kellam, 1994, p. 3). This point has also been stressed by Harris (1995) who argues that we should think about environmental effects in terms of group processes where peers play an important role. That is, phenomena such as within-group assimilation and between-group contrast, that increase the homogeneity of behaviors within groups and widen differences between social groups, could show as shared environment in behavior genetic analysis.

For externalizing behavior in boys, about 55% of the total variance is explained by common additive genetic factors, about 15% by common shared environmental factor, and about 10% by common nonshared environmental factors. For externalizing behavior, sex differences were found at age 10. For boys, an increase of additive genetic effects was found from age 7 to 10 (Van der Valk et al., 2003). A decrease in shared environmental influences was observed. For girls, the influences of genetic and environmental factors remained stable over the years. Pure nonshared environmental influences (undistorted by error or unreliability), represented by common nonshared environmental influences, were found for both internalizing and externalizing behavior. Thus idiosyncratic experiences seem to be of importance to explain individual differences in school-age children's problem behaviors.

Unique Parental Ratings

The rater-specific A, C, and E explain relatively small parts of the total variance in internalizing and externalizing problem behavior. For internalizing behavior, a possible specialization of the parent-child relationship over the years is represented by a relative increase of the rater-specific additive genetic factors. At age 3 (Van der Valk et al., 2001), the rater-specific additive genetic factors represent 16% of the total additive genetic effects, while at age 10 the rater-specific additive genetic effect explains 28% of total additive genetic variance based on mother ratings and 21% of the total additive genetic variance based on father ratings. While children grow older the mother-child and father-child relation may become more distinct because the child's behavior becomes

more diverse over the years. The diversity of behavior may create more situational-specific behavior, different for mothers and fathers

Rater bias was included in the estimate of the rater-specific shared environmental factor, accounting for at most 17% of the variance for both the Internalizing and Externalizing scales. Measurement errors and unreliability were estimated in the rater-specific nonshared environmental factor. However, neither for the Externalizing scale nor for the Internalizing scale did this factor account for more than 11% of the variance, except for internalizing behavior rated by the mother (16%).

Clinical Implications

Results of this study indicate that parental ratings are a valuable instrument for assessing behavioral and emotional problems in school-aged children. Using both mother and father ratings will give more reliable results by decreasing measurement error and rater bias. Further, results of this study indicate that parents assess a unique aspect of their child's behavior so that the combination of mother and father ratings will give a more complete picture of the child's behavior. Although parents have the advantage that they observe their children over longer periods of time and can witness both frequent and rare behaviors, they mainly interact with their children in the home environment. Two important implications for the clinician emerge from these findings. First, the relative importance of mother and father reports of child psychopathology has been debated over the years. The question of whether there is a right or wrong parent, or rather a single best informant, is addressed by this study. Our data support the practice of combining and contrasting both informants, as each adds unique information to the diagnostic formulation. Not only will this information be valuable in the tally of the number of symptoms accrued towards a diagnostic total, but perhaps more importantly, this information may provide relationship-specific measures of symptoms within mother-child and father-child settings. Such data can then be used in treatment planning. Parents can be educated about their agreement and disagreement, and further, about features of their child's suffering that may be unique to their viewpoint.

Second, our data indicate that the genetic contribution and parental agreement on measures of externalizing behavior remain high at age 10, while the genetic contribution and parental agreement on internalizing behavior diminishes between age 7 and 10. Such data are useful in understanding the difficult target that internalizing behaviors present the clinician and scientist alike. Where externalizing behavior is easily observable, stable, and highly influenced by genetic factors, internalizing behavior presents a more elusive clinical condition. Thus clinicians can educate themselves by understanding that parental ratings of internalizing behavior are more likely to be disparate, and thus, it is even more important to take the "both"

rather than the "single best informant" approach when assessing for internalizing syndromes. Further, research studies aimed at internalizing syndromes need be extraordinarily sensitive to the developmental and informant confounds present when studying internalizing behavior.

Additionally, a strength of the study is that the analyses were performed on a population-based sample. If one was to assume that psychopathology is caused by environmental hazards or pathogenic genes that are qualitatively distinct from those that cause variation in the normal range (Rutter et al., 1990), our results would have little clinical importance. There is, however, evidence that clearly suggests links between normal and abnormal behavior. First, several CBCL studies have shown correlations between behavior problem syndromes and DSM diagnoses (Costello et al., 1985; Edelbrock & Costello, 1988; Ferdinand et al., 1999; Kasius et al., 1997). This convergence indicates that behavior problem syndromes as studied in this article are relevant for psychiatric conditions. Second, several studies supported the view that the sources of normal variation may also affect psychopathology in children and adolescents. So latent class analyses have been used to identify subgroups of individuals with normal or pathological behavior (Eaves et al., 1993; Hudziak et al., 1998; Neuman et al., 1999). Results tend to suggest that these groups differ in degree rather than in kind. Furthermore, using methods from item response theory, Van den Oord et al. (2003) found that liability distributions for behavior and emotional problems show very little or no evidence of non-normality. This also seems to suggest that psychopathology may often be an extreme on the same continuum that describes variation in the normal range.

Limitations

The present study is based on parental ratings solely. Adding raters such as teachers who observe the child's behavior in other situations may contribute valuable information. For instance, comparison of the predictive power of parent and teacher information showed that teacher scores were a stronger predictor of poor outcomes than parents (Verhulst et al., 1994). Although teachers report fewer problems than parents about the same children, their reports are apparently informative with respect to later functioning. Further, self-reports might be valuable as well. Children may behave in a different manner when they are with their parents or their teacher. Parents and teachers can only rate those aspects of their children's behavior of which they are aware. Children, though, may be engaged in a variety of behaviors about which they do not tell their parents or teacher. Obviously, self-report becomes more important with increasing age.

The best selection of raters may depend on the type of problems that are studied. There is considerable evidence that parents are more likely to report symptoms of overactivity, inattention, and oppositional behavior

than their children (Edelbrock et al., 1986; Herjanic & Reich, 1982; Kashani et al., 1985; Loeber et al., 1991). On the other hand, children more frequently endorse emotional symptoms, including phobias and obsessional behavior (Herjanic & Reich, 1982) and depression (Angold et al., 1987; Kashani et al., 1985). Further, Loeber and colleagues (1989, 1991) have argued that parents and teachers are better informants of hyperactivity and oppositional behavior, while children and parents should be used to elicit conduct disorder symptomatology.

Psychopathology in parents seems to be correlated. Significant spousal correlations have been found for more internalizing behaviors such as depression and anxiety as well as externalizing behaviors such as antisocial behavior (Stallings et al., 1997; Krueger et al., 1998; Dufouil & Alperovitch, 2000; Mathews & Reus, 2001). These correlations could be a result either from assortative mating or contagion/interaction effects. For assortative mating, nonrandom mating occurs based on the psychopathology in both parents and in that case is a matter of selection. Contagion effects arise after mating and could be a result of the length of the relationship. Measuring parental temperament and psychopathology and estimating the contributions of these on measures of the child's internalizing and externalizing behavior may yield further valuable information.

Assortative mating is important for genetic research for two reasons. First, assortative mating increases genetic variance in a population (Falconer, 1989). In other words, positive assortative mating increases variance because the offspring differ more from the average than they would if mating were random. Even though spouse correlations are modest, assortative mating can greatly increase genetic variability in a population, because its effects accumulate generation after generation. Assortative mating is also important because it effects estimates of heritability. Positive assortative mating increases the resemblance between fraternal or dizygotic twins because it renders the parents of these twins more similar than they would be if there were no assortments. Identical or monozygotic twins, however, are already at the point of maximum genetic resemblance, and are thus unaffected by positive assortative mating (Fulker, 1988). This will result in an overestimation of shared environmental influences and an underestimation of additive genetic effects. A parent-offspring design would be necessary to investigate whether influences of shared environment are overestimated due to assortative mating.

Another effect of parental resemblance in psychopathology is that shared environmental effects of the part of the child's behavior assessed by both parents can be overestimated. Several studies suggest that depression in mothers may lead to their overestimating their children's symptomology (Fergusson & Horwood, 1987). In one study (Breslau et al., 1988),

mothers who were depressed rated their children as showing a greater number of symptoms of all psychiatric syndromes. Like mothers, fathers' reports of their children's behavioral problems are influenced by their own level of psychological symptoms (Phares et al., 1989; Jensen et al., 1988). The consequence of the facts that (a) parents tend to have similar levels of psychopathology, and (b) levels of parental psychopathology effect ratings of problem behavior in their children, is that the rater bias components of mothers and fathers become correlated. Because this shared rater bias component will effect MZ and DZ twin correlations in the same way, it will show as shared environmental effects on the common part of the parental ratings. The inclusion of measures of parental psychopathology or the use of different type of raters such as teachers will be helpful to account for these correlated rater bias effects.

In summary, besides parental agreement, rater-specific parental views on their children's behaviors seem to be significant at age 10. These results are in line with the findings in comparable samples of Dutch twins at ages 3 and 7 years. Additive genetic factors remain important as a source of individual differences in internalizing and externalizing problem behavior. Shared environmental influences, however, are also substantial. The changes in genetic and environmental effects occur mainly between age 3 and 7. No major changes are observed between age 7 and age 10. The significant influences of additive genetic factors indicate an innate vulnerability to childhood psychopathology. The influences of nonshared environmental influences suggest the importance of pure idiosyncratic experiences.

Acknowledgment

This research was financially supported by The Netherlands Organization for Scientific Research (575-25-012).

The Netherlands Organization for Scientific Research (NWO: R 56-467) and the Stichting Simonsfonds (SF053-iz) provided travel grants to facilitate collaboration with Dr E.J.C.G. van den Oord, at the Virginia Institute of Psychiatric and Behavioral Genetics, Department of Psychiatry, Medical College of Virginia/ Virginia Commonwealth University, Richmond, Virginia, USA.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.
- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

- Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Angold, A., Weissman, M. M., John, K., Merikangas, K. R., Prusoff, B. A., Wickramaratne, P., et al. (1987). Parent and child reports of depressive symptoms in children at low and high risk of depression. *Journal of Child Psychology and Psychiatry*, 28, 901-915.
- Asher, S. R. (1990). Recent advances in the study of peer rejection. In S. R. Asher & J. D. (Eds.), *Peer rejection in childhood* (pp. 3-14). Cambridge University Press.
- Barth, J. M., & Parke, R. D. (1993). Parent-child relationship influences on children's transition to school. *Merill-Palmer Quarterly*, 39, 173-195.
- Boomsma, D. I., Orlebeke, J. F., & Van Baal, G. C. M. (1992). The Dutch twin register: Growth data on weight and height. *Behavior Genetics*, 22, 247-251.
- Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research*, 1, 34-51.
- Breslau, N., Davis, G. C., & Prabucki, K. (1988). Depressed mothers as informants in family history research. *Psychiatry Research*, 24, 345-349.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Costello, E. J., Edelbrock, C. S., & Costello, A. J. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal Child Psychology*, 13, 579-595.
- Cowan, P. A., Cowan, C. P., Schultz, M. S., & Heming, G. (1994). Prebirth to preschool family factors in children's adaption to kindergarten. In R. D. Parke & S. G. Kellam (Eds.), *Exploring family relationships with other social contexts* (pp. 75-114). Hillsdale, NJ: Erlbaum.
- Dufouil, C., & Alperovitch, A. (2000). Couple similarities for cognitive functioning and psychological health. *Journal of Clinical Epidemiology*, 53(6), 589-593.
- Eaves, L. J. (1982). The utility of twins. In V. E. Anderson, W. A. Hauser, J. K. Penry, & C. F. Sing (Eds.), *Genetic basis of the epilepsies* (pp. 249-276). New York: Raven Press.
- Eaves, L. J., Silberg, J. L., Hewitt, J. K., Rutter, M., Meyer, J. M., Neale, M. C., et al. (1993). Analyzing twin resemblance in multisymptom data: Genetic application of a latent class model for symptoms of conduct disorder in juvenile boys. *Behavior Genetics*, 23, 5-19.
- Edelbrock, C., Costello, A. M., Dulcan, M. K., Conover, N. C., & Kala, R. (1986). Parent-child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry*, 27, 181-190.
- Edelbrock, C., & Costello, A. J. (1988). Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. *Journal of Abnormal Child Psychology*, 16, 219-231.
- Edelbrock, C., Rende, R., Plomin, R., & Thompson, L. (1995). A twin study of competence and problem behavior in childhood and early adolescence. *Journal of Child Psychology and Psychiatry*, 36, 775-785.
- Falconer, D. S. (1989). *Introduction to quantitative genetics*. Essex: Longman Scientific and Technical.
- Ferdinand, R. F., Stijnen, T., Verhulst, F. C., & Van der Reijden, M. (1999). Associations between behavioural and emotional problems in adolescence and maladjustment in young adulthood. *Journal of Adolescence*, 22, 123-136.
- Fergusson, D. M., & Horwood, L. J. (1987). The trait and method components of ratings of conduct disorder (Pt. 2). Factors related to the trait component of conduct disorder scores. *Journal of Child Psychology and Psychiatry*, 28, 261-272.
- Fulker, D. W. (1988). Genetic and cultural transmission in human behavior. In B. S. Weir, E. J. Eisen, M. M. Goodman, & G. Namkoong (Eds.), *Proceedings of the second international conference on quantitative genetics* (pp. 318-340). Sunderland, MA: Sinauer.
- Gjone, H., & Stevenson, J. (1997). The association between internalizing and externalizing behavior in childhood and early adolescence: Genetic and environmental influences? *Journal of Abnormal Child Psychology*, 25, 277-286.
- Harris, J. R. (1995). Where's the child's environment? A group socialization theory of development. *Psychological Review*, 102, 458-489.
- Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology*, 10, 307-324.
- Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves, L. J., & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics*, 22, 293-317.
- Hudziak, J. J., Heath, A. C., Madden, P. F., Reich, W., Bucholz, K. K., Slutske, W., et al. (1998). Latent class and factor analysis of DSM-IV ADHD: A twin study of female adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 848-857.
- Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., & Todd, R. D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39(4), 469-476.
- Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988). Child psychopathology rating scales and interrater agreement 1: Parents' gender and psychiatric symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 442-450.
- Kashini, J. H., Orvaschel, H., Burk, J. P. & Reid, J. C. (1985). Informant variance: The issue of parent-child disagreement. *Journal of the American Academy of Child Psychiatry*, 24, 437-441.

- Kasius, M. C., Ferdinand, R. F., & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 38, 625–632.
- Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry*, 143, 279–289.
- Koot, H. M., Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioural and emotional problems in young pre-schoolers: Cross-cultural testing of the validity of the Child Behavior Checklist 2/3. *Journal of Abnormal Child Psychology*, 25, 183–196.
- Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A. & Silva, P. A. (1998). Assortative mating for antisocial behavior: Development and methodological implications. *Behavior Genetics*, 28(3), 173–186.
- Ladd, G. W., & Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development*, 58, 1168–1189.
- Leve, L. D., Winebarger, A. A., Fagot, B. I., Reid, J. B., & Goldsmith, H. H. (1998). Environmental and genetic variance in children's observed and reported maladaptive behavior. *Child Development*, 69, 1286–1298.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*, 1, 317–337.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, 19, 75–95.
- Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, 38, 79–95.
- Mathews, C. A., & Reus, V. I. (2001). Assortative mating in the affective disorders: A systematic review and meta-analysis. *Comprehensive Psychology*, 42(4), 257–262.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.
- Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling* (5th ed.). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.
- Neuman, R. J., Todd, R. D., Heath, A. C., Reich, W., Hudziak, J. J., Bucholz, K. K., et al. (1999). Evaluation of ADHD typology in three contrasting samples: A latent class approach. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 25–33.
- Parke, R. D., & Kellam, S. G. (Eds.). (1994). *Exploring family relationships with other social contexts*. Hillsdale, NJ: Erlbaum.
- Phares, V., Compas, B. E., & Howell, D. C. (1989). Perspectives on child behavior problems: Comparisons of children's self-reports with parent and teacher reports. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 68–71.
- Rietveld, M. J. H., van der Valk, J. C., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research*, 3, 134–141.
- Rietveld, M. J. H., Hudziak, J. J., Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D. I. (in press). Heritability of attention problems in children: 1. Longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry*.
- Rowe, D. C., & Kandel, D. (1997). In the eye of the beholder? Parental ratings of externalizing and internalizing symptoms. *Journal of Abnormal Child Psychology*, 25, 265–275.
- Rutter, M., Macdonald, H., Le Couteur, A., Harrington, R., Bolton, P., & Bailey, A. (1990). Genetic factors in child psychiatric disorders (Pt. 2). Empirical findings. *Journal of Child Psychology and Psychiatry*, 31, 39–83.
- Schmitz, S., Fulker, D. W., & Mrazek, D. A. (1995). Problem behavior in early and middle childhood: An initial behavior genetic analysis. *Journal of Child Psychology and Psychiatry*, 36, 1443–1458.
- Schneider, B.H. (1993). *Children's social competence in context: The contribution of family, school and culture*. Oxford: Pergamon Press.
- Silberg, J. L., Erickson, M. T., Meyer, J. M., Eaves, L. J., Rutter, M. L., & Hewitt, J. K. (1994). The application of structural equation modeling to maternal ratings of twins' behavioral and emotional problems. *Journal of Consulting and Clinical Psychology*, 62, 510–521.
- Simonoff, E., Pickles, A., Hewitt, J., Silberg, J., Rutter, M., Loeber, R., et al. (1995). Multiple raters of disruptive child behavior: Using a genetic strategy to examine shared views and bias. *Behavior Genetics*, 25, 311–326.
- Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, 28, 825–837.
- Stallings, M. C., Cherny, S. S., Young, S. E., Miles, D. R., Hewitt, J. K., & Fulker, D. W. (1997). The familial aggregation of depressive symptoms, antisocial behavior, and alcohol abuse. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 74, 183–191.
- Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1996). A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins. *Journal of Abnormal Psychology*, 105, 349–357.

- Van den Oord, E. J. C. G., & Rowe, D. C. (1998). An examination of genotype-environment interactions for academic achievement in an US National Longitudinal Survey. *Intelligence, 25*, 205-228.
- Van den Oord, E. J. C. G., & Rowe, D. C. (1999). A cousin study of associations between family demographic characteristics and children's intellectual ability. *Intelligence, 27*, 251-266.
- Van den Oord, E. J. C. G., Simonoff, E., Eaves, L. J., Pickles, A., Silberg, J., & Maes, H. (2000). An evaluation of different approaches for behavior genetic analyses with psychiatric symptom scores. *Behavior Genetics, 30*(1), 1-18.
- Van den Oord, E. J. C. G., Pickles, A. P., & Waldman, I. D. (2003). Normal variation and abnormality: An empirical study of the liability distributions underlying depression and delinquency. *Journal of Child Psychology and Psychiatry, 44*(2), 180-192.
- Van der Valk, J. C., Verhulst, F. C., Stroet, T. M., & Boomsma, D. I. (1998a). Quantitative genetic analysis of internalizing and externalizing problems in a large sample of 3-year-old twins. *Twin Research, 1*, 25-33.
- Van der Valk, J. C., Verhulst, F. C., Neale, M. C., & Boomsma, D. I. (1998b). Longitudinal genetic analysis of problem behaviors in biologically related and unrelated adoptees. *Behavior Genetics, 28*, 365-380.
- Van der Valk, J. C., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology and Psychiatry, 42*(7), 921-931.
- Van der Valk, J. C., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2003). Using common and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems. *Behavior Genetics, 33*(4), 409-420.
- Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology, 22*(5), 531-546.
- Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18*. Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.
- Zahn-Waxler, C., Schmitz, S., Fulker, D., Robinson, J., & Emde, R. (1996). Behavior problems in 5-year-old monozygotic and dizygotic twins: Genetic and environmental influences, patterns of regulation, and internalization of control. *Development and Psychopathology, 8*, 103-122.