The Thought–Problem scale (TP) of the CBCL assesses symptoms such as hallucinations and strange thoughts/behaviors and has been associated with other behavioral disorders. This study uses parental reports to examine the etiology of variation in TP, about which relatively little is known, in 7-year-old twins. Parental ratings on TP were collected in 8,962 7-year-old twin pairs. Because the distribution of TP scores was highly skewed scores were categorized into 3 classes. The data were analyzed under a threshold liability model with genetic structural equation modeling. Ratings from both parents were simultaneously analyzed to determine the rater agreement phenotype (or common phenotype [TPc]) and the rater specific phenotype [TPs] that represents rater disagreement caused by rater bias, measurement error and/or a unique view of the parents on the child's behavior. Scores on the TP-scale varied as a function of rater (fathers rated fewer problems), sex (boys scored higher) and zygosity (DZ twins scored higher). The TPc explained 67% of the total variance in the parental ratings. Variation in TPc was influenced mainly by the children's genotype (76%). Variance in TPs also showed a contribution of genetic factors (maternal reports: 61%, paternal reports: 65%), indicating that TPs does not only represent rater bias. Shared environmental influences were only found in the TPs. No sex differences in genetic architecture were observed. These results indicate an important contribution of genetic factors to thought problems in children as young as 7 years.

Keywords: thought problems, parental rating, twins, heritability

The Thought–Problem syndrome (TP) of the Child Behavior Checklist (CBCL; Achenbach, 1991) is an empirically derived set of items that cover symptoms such as hallucinations, obsessive–compulsive symptoms and strange thoughts and behaviors. A strong statistical relation through which the scale was derived (Achenbach, 1991). The internal consistency of the TP-scale was shown to be sufficient in the CBCL (Kuo et al., 2004; Lin et al., 2006). Compared to the other empirically derived CBCL scales, the TP-scale has received relatively little attention. In a longitudinal study of Ferdinand et al. (2001), thought problems in childhood were associated with substance abuse (alcohol and tobacco) in young adulthood. Other studies showed TP in children to be associated with disorders, such as obsessive–compulsive disorder (OCD; Geller et al., 2004), multiple complex developmental disorder (MCDD; de Bruin et al., 2006) and fragile X syndrome (Hessel et al., 2006). The TP-syndrome also predicted DSM-III-R diagnoses of simple phobia, social phobia, separation anxiety disorder, mood disorders and psychotic disorders (Kasius et al., 1997). There have been a few studies on the heritability of the TP-scale in young twins that found evidence for significant heritability (Edelbrock et al., 1995; Kuo et al., 2004; Lin et al., 2006; Polderman et al., 2006; Schmitz et al., 1995). The aim of this study is to obtain more insight into the genetic and environmental contributions to variation in TP in a large group of 7-year-old twins. The rating of TP in young children might be complicated and we obtained ratings from both parents and modeled the extent to which both parents agreed on the presence or absence of TP.

Parental ratings, such as those collected with the CBCL, provide meaningful information about a child's behavioral and emotional problems. Parents observe their children for long periods and in natural situations, which makes them ideal informants regarding their children's behavior. Studies that collect information...
from both parents tend to show that parents agree for a substantial part in their assessment of behavioral problems in their children. An analysis by Achenbach et al. (1979) showed a correlation of 0.60 between the paternal and maternal ratings of the same child. Although correlations are high, they are not perfect. To explore the processes underlying disagreements in parental ratings, Hewitt et al. (1992) developed a series of models, in which the disagreement of the parents’ reports reflect rater bias and a unique accurately assessed part of the child's behavior.

Rater bias may be caused by the parents’ own characteristics (a projection bias), or by parents’ response biases (e.g., stereotyping, employing different normative standards or having certain response styles, i.e., judging certain types of behavior more or less severely). The unique views on the phenotype may arise firstly because parents observe their child in distinct situations and environments. For instance, the mother could have the task of bringing the child to school, while the father is the one who accompanies the child to soccer games. Second, the way the mother interacts with the child may differ from the way the father interacts with the child (Achenbach et al., 1979). Another cause for different views in parent reports could be that men are more sensitive to different kinds of input from their children than women (Baron-Cohen & Hammer, 1997; Connellan et al., 2000; Rosenthal et al., 1979).

A number of quantitative genetic studies have examined problem behavior in children with the CBCL, with the main focus on internalizing and externalizing behavior problems (Bartels et al., 2004a; Bartels et al., 2004b; Bartels et al., 2007; Edelbrock et al., 1995; Gjone & Stevenson, 1997; Hudziak et al., 2000; Leve et al., 1998; Schmitz & Mrazek, 2001; Silberg et al., 1994; Van den Oord et al., 1995; Van der Valk et al., 1998; Zahn-Waxler et al., 1996) and a number of these studies have modeled the agreement and disagreements among parental reports. For example, in a study by Rowe and Kandel (1997), parents rated the internalizing and externalizing behavior of their two oldest children (between 9 and 17 years old). ‘Individual view’ and ‘shared view’ models were used. The parents assessed similar aspects of the child’s behavior, and in addition the mother and father ratings contained a significant individual view component.

Twin studies provide the opportunity to analyze whether the variance of the unique view of the parental assessments can be explained by genetic factors. If genetic factors are found, this implies that the unique view of the parents partly represents real behavior of the child. This rules out the possibility of the rater disagreements containing only rater bias. Hewitt et al., (1992) studied internalizing behavior in prepubertal (8–11 years) and pubertal twins (12–16 years), and found evidence for such genetic effects. Dutch twin studies on internalizing and externalizing behaviors also found that the unique viewpoint of parents does not solely reflect rater bias (Bartels et al., 2003; Bartels et al., 2004a; Bartels et al., 2007; Van der Valk et al., 2001; Van der Valk et al., 2003).

When data are derived from questionnaires developed to measure the degree of dysfunctional behavior such as the CBCL, a large degree of skewness is often observed. These non-normal distributions can be explained by the fact that in symptom data a majority of the subjects displays few or no symptoms (Van den Oord et al., 2003). This is especially the case for the TP subscale, which is one of the subscales with the lowest mean scores in general population samples (Achenbach, 1991). Logarithmic and square root transformations are often not enough to correct for this non-normality. Categorizing the observations and analyzing the data with a threshold model has shown to be a successful way to decrease bias in parameter estimates (Derks et al., 2004). A disadvantage of categorizing the data may be that it reduces the statistical power of the analysis, and therefore large sample sizes are required.

This study estimates the genetic and environmental influences on variation in thought problems in a large group of 7-year-old Dutch twin pairs, while taking the agreement and disagreements between the parent reports into account. To overcome biased estimates due to skewness of the phenotype, the continuous TP scale was categorized into three classes (low, middle, high) and analyzed with threshold models.

**Methods**

**Subjects**

The participants were all registered with the Netherlands Twin Registry (NTR), which was established by the Department of Biological Psychology at the Vrije Universiteit in Amsterdam (Bartels et al., 2007; Boomsma et al., 2006). Parents of young twins receive questionnaires when their twins are 1, 2, 3, 5, 7, 10 and 12 years old. For this study, data of 7-year-old twin pairs from birth cohorts 1986–1997 were used. The questionnaires were mailed to the parents within 3 months of the twins’ 7th birthday. Reminders were sent after 2 to 3 months. The response rate was 62% (N = 8962). Reasons for families not participating at a particular wave of data collection vary; families may request to not take part in the research at a particular age due to various reasons. Also, families may move to new addresses without notifying the NTR staff. Bartels et al. (2007) showed that the dropout was largely random according to the definition of Little & Rubin (1987), making generalizations of the results more valid.

Zygosity was determined for 1492 same-sex twin-pairs by blood group (n = 389) or DNA polymorphisms (n = 1103). The zygosity of the other same-sex twins was determined using a discriminant analysis of questionnaire items answered by the parents. The
questionnaire led to correct classification of the zygos-
ity in about 93% of the cases (Rietveld et al., 2000).

The sample contained 1466 monozygotic male
(MZM), 1516 dizygotic male (DZM), 1675 monozy-
gotic female (MZF), 1445 dizygotic female (DZF),
and 2860 dizygotic opposite-sex (DOS) twin pairs.
Response rates for mothers were higher than for
fathers and the group could be further divided into
twin pairs for which both mother and father had
replied (1053 MZM, 1105 DZM, 1226 MZF, 999
DZF, 2002 DOS), pairs with only mother-reports (395
MZM, 392 DZM, 423 MZF, 424 DZF, 805 DOS) and
pairs with only father-reports (n = 113).

Measures
The Child Behavior Checklist (CBCL 4-18; Achenbach,
1991) was used to assess emotional and behavioral
problems including thought problems. The question-
aire consists of 113 items, and measures behavior
during the preceding 6 months. There are seven TP
items: ‘Can’t get his/her mind off certain thoughts’,
‘Repeats certain acts over and over’, ‘Strange behavior’,
‘Strange ideas’, ‘Hears sounds or voices that aren’t
there’, ‘Sees things that aren’t there’, ‘Stares blankly’.
The reliability of the CBCL has been confirmed by in
Dutch epidemiological samples. The 2-week test–retest
correlation for TP was 0.74 (Verhulst et al., 1996). The
TP scores were transformed from continuous to catego-
rical data with two thresholds to limit the number of
categories to three (low, middle and high levels of TP).
The thresholds were chosen in such a way that there
were no empty cells in the contingency tables of twin 1
versus twin 2 scores.

Genetic Analyses
Data from monozygotic (MZ) and dizygotic (DZ)
twins were used to decompose the variation in the lia-
bility of thought problems into a contribution of
additive genetic, shared environmental and non-shared
environmental components. The categorical trait
thought problems were modeled to have an unob-
served, underlying continuous distribution with two
thresholds that divide the distribution into three cate-
gories (low, middle, high levels of TP). Such underly-
in distributions have been termed the liability or vulnera-
bility (Falconer, 1989). The continuous variation in
liability may be genetic or environmental in origin.
The mean and variance of the liability distribution
were standardized with mean zero and unit variance.

The additive genetic influence (A) on the variation
in liability represents the sum of allelic influences at
each locus in the genome contributing to the pheno-
type. The environmental influences can be shared
and nonshared. Shared environment is common (C) to
both twins growing up in the same family. Nonshared
environmental influences (E) are unique to each twin
and do not lead to twin resemblance.

Model Fitting
With multiple raters (two parents in this study), the variance
of the liability distribution can be distinguished
into two parts that represent agreement (commonly
rated phenotype by both parents) and disagreement
(unique parental views). Both parts of the variance can
be influenced by genetic (A), shared (C) and nonshared
(E) factors and not necessarily to the same degree. We
use the psychometric model (Hewitt et al., 1992) as pre-
sented in Figure 1. All variables enclosed in circles are
latent (unobserved) that influence observed traits
(enclosed in rectangles). The influence of latent variables
on other latent variables or on observed traits is given
by factor loadings (a), (c), and (e). These factor loadings
come with (f) or (m) if the latent variable influences a
unique view of the father or the mother.

Mx (Neale et al., 2006) was used to obtain parameter estimates for thresholds and factor loadings. Genetic models were fitted to the raw data with maximum likelihood estimation procedures. In a satu-
rated model thresholds and polychoric twin
correlations were estimated separately for MZM,
DZM, MZF, DZF, DOSMF and DOSFM groups. This
model was also used to test whether the thresholds
could be constrained to be equal for mother and
father ratings, for MZ and DZ-twins, for boys and
girls and for the youngest and the oldest twin. Next,
the fit of the psychometric model was compared with
the fit of the saturated model.

The psychometric model was used to test for sex
differences in factor loadings. Next, the significant
contribution of the common and unique A and C vari-
ance components was tested. The significance of the
unique ‘A’ component can show whether the rater spe-
cific parts represent not only measurement error and
rater bias, but also reflects meaningful variation.
Significance of the estimates was established by com-
paring the full model with a simplified model. The
more parsimonious nested model is chosen over the

Figure 1
The psychometric-model.
full model when the analysis shows a low nonsignificant $\chi^2$ test statistic ($p > .05$). In addition to the $\chi^2$ test statistic, Akaike’s Information Criterion ($\text{AIC} = \chi^2 - 2 \times \text{degrees of freedom}$) was computed. The lower the AIC, the better the fit of the model to the observed data.

**Results**

**Description of the Data and Threshold Differences**

For descriptive purposes, Table 1 summarizes the means and variances for the untransformed measures of Thought Problems by sex for mother and father reports.

Table 2 shows the thresholds from the saturated model in Mx. The thresholds are higher for all father ratings, indicating a tendency for mothers to give higher ratings (i.e., lower thresholds) for both boys and girls ($\chi^2 (24) = 84.19, p < .01$, see Table 4). The thresholds also indicate more TP (i.e., lower thresholds) for the DZ-twins than for the MZ-twins ($\chi^2 (16) = 29.75, p = .02$). Finally, Table 2 shows higher thresholds for the liability for TP in girls than in boys (i.e., fewer thought problems for girls) ($\chi^2 (24) = 42.13, p = .01$). The differences between the thresholds of the youngest and oldest twin were not significant ($\chi^2 (24) = 32.35, p = .11$).

**Correlations**

From the saturated model polychoric correlations between twins, between raters and cross-twin cross-rater were obtained. Table 3 shows correlations between twins rated by the same parent in the first and second columns. The last four columns show the cross-correlations between twins each rated by a different parent. The interparent correlations were comparable for both first- and second-born twins. On average the interparent correlation is .66, which is in the same range as the parental agreement found in previous studies of about .60 (Achenbach, 1991).

**Model Fitting Results for the Psychometric Model**

The model fitting results of the simultaneous analysis of maternal and paternal ratings are summarized in Table 4. Sex differences in factor loadings were not detected ($\chi^2 (7) = 7.35, p = .60$). The common environmental effects on the common view of parents were not significant ($\chi^2 (1) = .972, p = .32$), but the contribution of C to the variance of the unique views of father and mother was significant ($\chi^2 (2) = 19.54, p < .01$). Dropping the additive genetic component on the common view ($\chi^2 (1) = 97.06, p < .01$) as well as on the unique views of father and mother was significant ($\chi^2 (2) = 18.09, p < .01$) of the parents also gave significant deteriorations of fit, indicating a significant contribution of A to the variance of the common and rater specific parts of the phenotype. Table 5 summarizes the estimates of genetic and environmental influences. The largest part of the variance...
in liability for thought problems could be explained by genetic factors.

The commonly assessed part of the phenotype explains 67% (51% [= A] + 16% [= E]; see Table 5) of the total variance in maternal and paternal ratings (total variance = sum of all maternal/paternal parameters) and was influenced mainly by the children’s genotype. Heritability of the commonly assessed TP phenotype was 76% (51% [= A] / 67% [= A + E]). The remaining variance of the commonly assessed phenotype was explained by non-shared environmental factors (16% [= E] / 67% [= A + E] = 24%).

The rater specific genetic influences seem somewhat higher for the father than the mother ratings (14% vs. 10%). These significant genetic influences on the unique views indicate that fathers and mothers assess reliable and rater specific information regarding TP in their children. The rater specific shared environmental influences explain about 13% of the variance both parent-ratings. This is the part of the variance that could represent rater bias. Higher maternal specific nonshared environmental influences are observed (10 % vs. 6 %), which could reflect reliable information about the nonshared environmental influences on the variance of the phenotype, but may also include measurement error.

**Discussion**

This study examined the influence of genetic and environmental factors on thought problems in 7-year-old twins, based on parental reports. The analyses modeled genetic and environmental influences on both the commonly agreed upon phenotype and on unique views of the parents. Parents agreed to a large extent on the occurrence of thought problems (TP) in their twin offspring. The commonly assessed phenotype explained 67% (51% [= A] + 16% [= E]) of the total variance in maternal and paternal ratings and is influenced mainly by genetic factors. Heritability of the commonly assessed TP phenotype was 76% (51% [= A] / 67% [= A + E]; see Table 5). There was no evidence that shared family environment contributed to variance of the common phenotype. In contrast, shared environmental influences contributed to the unique views of parents. The rater specific shared environmental influences may be due to rater bias, but explain only a small part of the total variance (13% [= C] / 33% [= A + C + E]; see Table 5). There was no evidence that shared family environment contributed to variance of the common phenotype. In contrast, shared environmental influences contributed to the unique views of parents. The rater specific shared environmental influences may be due to rater bias, but explain only a small part of the total variance (13% [= C] / 33% [= A + C + E]; see Table 5). There was no evidence that shared family environment contributed to variance of the common phenotype. In contrast, shared environmental influences contributed to the unique views of parents. The rater specific shared environmental influences may be due to rater bias, but explain only a small part of the total variance (13% [= C] / 33% [= A + C + E]; see Table 5). There was no evidence that shared family environment contributed to variance of the common phenotype.
observed significant small nonadditive genetic influences for boys. Shared environmental influences were not significant in any of the previous studies (Edelbrock et al., 1995; Schmitz et al., 1995). In the study by Polderman et al. (2006), TP was rated by teachers and familial influences were detected, but to explain the familial influences, it was not possible to distinguish between common environment and genetic factors.

The genetic influences on the variance of the raterspecific parts are higher for the father reports in the analyses. Previous studies of internalizing and externalizing disorders in young twins usually found these estimates to be larger in maternal ratings (Bartels et al., 2003; Bartels et al., 2004a; Boomsma et al., 2005; Van der Valk et al., 2001; Van der Valk et al., 2003). This may support the notion that there are sex differences in the perception of parents with respect to this phenotype in their children and that fathers may add important extra information regarding the thought problems phenotype. Sex differences in human perception have been suggested from a very early age onwards (Connelan et al., 2000).

We observed sex differences in TP with boys showing more TP than girls. A similar sex difference is found in OCD, which is associated with TP (Hanna, 1995; Geller et al., 1998; Zohar, 1999). We also observed differences between mono- and dizygotic twins, with MZ twins obtaining lower TP scores than DZ twins. This suggests the presence of negative social interactions in twin pairs, since under interaction the prevalence rates for a categorical variable between MZ and DZ twins are expected to differ if that trait has an underlying continuous distribution (Carey, 1992).

The distribution of the TP scale was highly skewed and we analyzed the data using a threshold model. The means of the untransformed (continuous) data shown in Table 1 are also indeed lower than most other CBCL-sub scales (Achenbach, 1991), which indicates that a majority of the subjects display few or no symptoms, which is expected in such a skewed distribution (Van den Oord et al., 2003). A simulation study by Derks et al. (2004) showed that analysis of L-shaped distributed data results in an underestimation of additive genetic and shared environmental influences and an overestimation of non-shared environmental effects. After conducting the Mx-analyses on the untransformed data (results not shown), these differences were also found when comparing those estimates with the estimates made through analysis of the categorical data. Categorical analysis also has disadvantages. First, the statistical power is reduced, and requires large samples. This study used a large sample of 8962 twin pairs. Another disadvantage of the categorical data analyses is that they are computationally more demanding.

The most important finding from these data is the substantial heritability that is observed in young children. This strengthens the notion that the TP-scale measures a true syndrome. Other ways to support this finding may include a linkage and genome wide association analyses. Further study is needed to investigate the differences between the parental reports. Further research is also needed to obtain insight into estimates of heritability in other age groups and to obtain information on the relation between thought problems and other disorders such as autism, OCD and psychotic disorders, to examine whether the thought problem phenotype could also have a predictive value for these disorders. It is important that such studies in children include both maternal and paternal reports.

Acknowledgments

We are grateful to the twin families for their participation. This research was supported by NWO grants 463-06-001; 575-25-006; 480-04-004; ‘Spinozapremie’ (NWO/SPI 56-464-14192); NIMH, RO1 MH58799-03; CNCR (Centre Neurogenetics Cognition Research); Dr Bartels is supported by NWO-VENI grant 451-04-034. We would like to thank David Sondervan for zygosity typing.

References


