Behav Genet (2014) 44:326–336
DOI 10.1007/s10519-014-9657-7

ORIGINAL RESEARCH

Assessing Genetic Influences on Behavior: Informant and Context Dependency as Illustrated by the Analysis of Attention Problems

Kees-Jan Kan · Catharina E. M. van Beijsterveldt · Meike Bartels · Dorret I. Boomsma

Received: 2 October 2013 / Accepted: 4 April 2014 / Published online: 6 May 2014

Abstract Assessment of genetic influences on behavior depends on context, informants, and study design: We show (analytically) that, conditional on study design, informant specific genetic variance is included in the genetic variance component or in the environmental variance component. To aid the explanation, we present an illustrative empirical analysis of data from the Netherlands Twin Register. Subjects included 1,571 monozygotic and 2,672 dizygotic 12-year-old twin pairs whose attention problems (AP) were rated by their parents, teachers, and themselves. Heritability estimates (h^2) of AP were about $\sim$0.75 for same informant ratings (mother, father, and same teacher ratings) and $\sim$0.54 for different informants' ratings (different parents', different teachers', and two twins' self-ratings). Awareness of assessment effects is relevant to research into psychiatric disorders. Differences in assessment can account for age effects, such as a drop in heritability of ADHD symptoms. In genome-wide association studies, effects of rating specific genetic influences will be undetectable.

Keywords (Missing) heritability · Informant effects · Context dependent behavior · Attention problems · ADHD

Edited by Deborah Finkel.

K.-J. Kan (✉) · C. E. M. van Beijsterveldt · M. Bartels · D. I. Boomsma
Department of Biological Psychology, VU University Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands
e-mail: k.j.kan@vu.nl

C. E. M. van Beijsterveldt
e-mail: t.van.beijsterveldt@vu.nl

M. Bartels
e-mail: m.bartels@vu.nl

D. I. Boomsma
e-mail: d.i.boomsma@vu.nl

K.-J. Kan
Department of Methodology, VU University Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

C. E. M. van Beijsterveldt · M. Bartels · D. I. Boomsma
EMGO+ Institute for Health and Care, VU University Medical Center, Amsterdam, The Netherlands

M. Bartels · D. I. Boomsma
Neuroscience Campus Amsterdam (NCA), Amsterdam, The Netherlands

A trait's heritability depends on many aspects. It is by definition *population dependent* and *time dependent* as heritability is defined as the relative contribution of genotypic variance to phenotypic variance in a particular population at a specific point in time (Falconer and Mackay 1996). Related to this, the heritability of a trait can be *gender dependent* or *age dependent*. Empirically gender dependency in behavioral and psychological traits is more often absent than present, but examples exist. Seasonal mood changes, for instance, are slightly more heritable in males than in females (Jang et al. 1997). Age dependency of heritability is common and is often illustrated by reference to the heritability of intelligence, which is well known to increase gradually from childhood to young adulthood (Haworth et al. 2009).

Perhaps less well-known, but increasingly acknowledged, is that symptoms of attention deficit/hyperactivity disorder (ADHD), and associated complex traits, such as attention problems (AP), demonstrate the opposite of intelligence in the sense that their heritability estimates are relatively high in childhood (Derks et al. 2009; Faraone et al. 2005), but relatively low in adulthood (Boomsma et al. 2010; Chang et al. 2013; Haberstick et al. 2008;

Kan et al. 2013a; Larsson et al. 2012; Merwood et al. 2013; Saviouk et al. 2011; van den Berg et al. 2006). Moreover, in the case of AP, the drop in estimated heritability is not gradual, but abrupt, which appears to be the result of a sudden increase in nonshared environmental variance (Kan et al. 2013a). In contrast to the increasing heritability of intelligence, which has been ascribed to a gradual increase in active genotype-environment covariance (Haworth et al. 2009; Kan et al. 2013b), the differences between heritability of ADHD symptoms and AP in adults versus children is still left unexplained (Merwood et al. 2013). It has been surmised, however, that these differences do not reflect true developmental changes, but rather difference in assessment, because they coincide with a shift from reliance on reports by others to reliance on self-reports (Kan et al. 2013a; Merwood et al. 2013; Chang et al. 2013). This suggests heritability is also *assessment* or *informant dependent*.

Indeed, a body of empirical research has established that heritability estimates of many behavioral problems that relate to psychiatric disorders vary across informant ratings (Merwood et al. 2013; Arseneault et al. 2003; Bartels et al. 2003a; Eaves et al. 1997; Lamb et al. 2012; Martin et al. 2002; Polderman et al. 2006; Sherman et al. 1997; Wood et al. 2008; Kendler et al. 2008). The results can be summarized as follows. While heritability estimates for mother and father ratings of their children's behavioral problems—attention problems and ADHD symptoms included—are about equal, they are (on average) larger than those for teacher ratings, which, in turn, are (on average) larger than those for self-ratings. In addition, heritability estimates for teacher ratings vary considerable. They are relatively high when two siblings or twins are rated by the same teacher and relatively low when rated by different teachers.

The heterogeneity among heritability estimates for parent, teacher and self-ratings has received several explanations (see Merwood et al. 2013, for a recent discussion). One explanation involves the role of measurement error variance and reliability (Hartman et al. 2007). In behavior genetics, variance due to measurement error is by definition subsumed under the environmental variance component (Plomin et al. 2012, p. 96). Increased measurement error variance, or, equivalently, decreased reliability, will result in increased environmental and observed phenotypic variance (see Fig. 1). As heritability is the proportion of genetic variance to phenotypic variance, relatively high measurement error variance (relatively low reliability) will thus result in relatively low heritability estimates. Hence, it can be hypothesized that self-reports are less reliable than teacher reports, and teacher reports less reliable than parent reports. Although such hypothesis would constitute a valid theoretical account, there is no convincing evidence for such a systematic pattern, neither from validation studies (e.g., Achenbach and Rescorla 2001), nor from empirical studies. In addition, differences in reliability *across* parent, teacher, and self-ratings do not provide an explanation for the large differences between heritability estimates *among* teacher ratings. So, even though differences in reliability will contribute to differences in heritability, reliability does not constitute the full explanation for the observed heterogeneity in heritability estimates.

Relevant to the previous explanation is the role of response bias ('rater bias'). Like measurement error, differences in response biases affect the degree of agreement between two ratings of the same variable. Examples of response bias include the situation in which informants consistently overrate or underrate occurrences of behavior or when the degree of similarity between twins is overestimated or underestimated. Such biases can result in (positively or negatively) correlated error terms between ratings from the same informant (Hewitt et al. 1992; Bartels et al. 2007b). If differences in response biases are present, the degree to which ratings from different informants are similar can thus be expected to differ from the degree to which ratings from the same informants are similar. Key is that depending on study design observed between twin pair correlations can deviate considerably from each other and from the actual between twin pair correlation. This is important knowledge, because in a twin design the correlation between monozygotic twin pairs determines the upper limit of a trait's heritability estimate (Falconer and Mackay 1996). The more raters disagree, the lower between twin pair correlation can get, hence the lower the upper limit of the heritability estimate. In this regard, rater bias is considered to be an appealing and important potential explanation of observed differences in heritability (Rice et al. 2002).

We note, however, that a lowering of the upper *limit* of the heritability does not (necessarily) imply that the *actual* heritability estimates are expected to be lower. Consider the measurement model in Fig. 1, for instance, in which error terms between same informant's ratings of phenotype P are correlated due to the presence of rater biases; error terms between different informants' ratings are uncorrelated. The observed twin similarity (hence the observed monozygotic twin correlation) between same informant's twin pair ratings (e.g., mother ratings, father ratings, or same teacher ratings) differs from the observed similarity between different informants' twin pair ratings (e.g., different teacher ratings and the self-ratings of the two twins), regardless whether differences in reliability are present or absent. This is because the effects of biases on the twin pair ratings are shared and nonshared, respectively. Yet, the contribution of genetic variance to observed phenotypic variance does not depend on the presence of these biases. In the model in Fig. 1, the estimated heritability based on

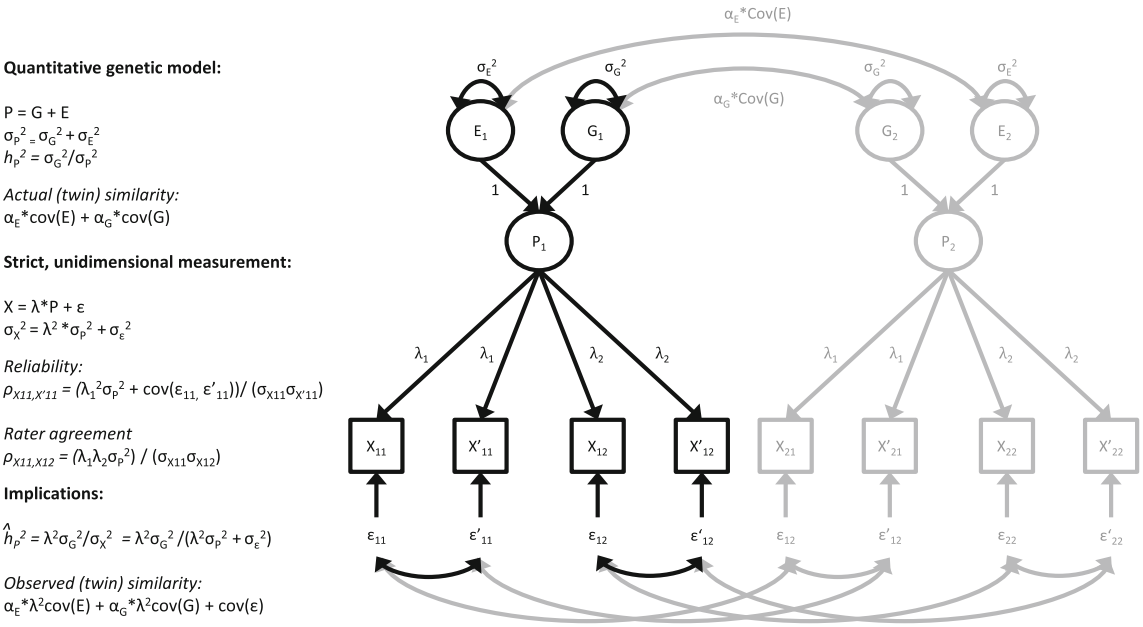

Fig. 1 Strict, unidimensional measurement model of phenotype P defined according to quantitative genetic theory (in *black*), and the extension of it in a genetic informative design (in *grey*). Variable P is regressed on latent genetic (G) and environmental (E) variables with variances σ_G and σ_E, such that the total phenotypic variance is $\sigma_P = \sigma_G + \sigma_E$ (when covariance between E and G and interaction terms are considered absent). Subjects resemble each other due to shared environmental effects and similarity in genetic makeup. The resemblance can be quantified as α_G*Cov(G) + α_E*Cov(E), where α_G represents the proportion shared segregating genes. In monozygotic twins α_G equals 1. In dizygotic twins the average α_G equals 0.5 (or lower in the presence of nonadditive genetic effects). Dummy α_E represents the proportion of environmental effects that cohabiting individuals share (= 0 in AE and ADE models and 1 in ACE models). Informants 1 and 2 (are imagined to) have provided both two ratings of the subjects' phenotype (X and repeated measure X′, as in a test–retest settings or when split-halves are used). Hence four repeated measures of the latent phenotype are available for each of the two subjects. Symbol λ represents a scaling parameter which may be assumed equal if raters provided ratings on the same rating scale. The symbol ε represents measurement error. Due to the presences of bias, error terms between ratings from the same rater can be correlated. In this model, cross informant twin correlations are expected to be lower than within informant twin correlations, but heritability estimates will not depend on whether ratings are provided by the same informant or by different informants

ratings from different informants will not differ from those based on same informant ratings, provided any differences in reliability are absent or taken into account.

We also stress that disagreement between two behavioral measures is not necessarily due to measurement error or response bias (Hewitt et al. 1992; Bartels et al. 2007b). Behaviors, hence phenotypic differences herein, tend to depend on context or situation. When different informants observe twins' behaviors in different situations, differences in the informants' responses may thus well reflect true behavioral differences, rather than (or on top of) differences in bias. Previous behavior genetic studies provide ample evidence of the presence of genetic and environmental influences that are informant specific (in addition to common genetic and environmental influences). This holds true not only for attention problems (Derks et al. 2006), symptom counts of ADHD (Burt et al. 2005; McLoughlin et al. 2011; Merwood et al. 2013), or related traits (activity; Wood et al. 2008), but also for symptom counts of oppositional defiant disorder (Burt et al. 2005), anxiety and depression (Kendler et al. 2008; Boomsma et al. 2005), and conduct disorder (Burt et al. 2005), and for ratings of internalizing and externalizing problem behavior, whether defined in general (Bartels et al. 2003a), or more specifically, e.g., as withdrawn behavior (Hoekstra et al. 2008) or aggressive behavior and rule-breaking behavior (Bartels et al. 2003b). That informant specific factors are genetically influenced suggests that rater disagreement is not limited to the (possible) effects of rater bias, and also that the model in Fig. 1 is too strict.

Studying the issue why heritability estimates of a trait vary across informant ratings is complicated, not in the least because of possible confounding effects: if both the manner of assessment and the influences of environmental or genetic effects change during the course of development, it is difficult to disentangle their contributions to observed changes. Yet, increasingly, efforts are being taken to do so. By analyzing longitudinal, genetic informative cross-informant data important contributions have been made to the understanding of the development of psychiatric disorders. From research on ADHD symptoms (Chang et al. 2013; Merwood et al. 2013), for instance, we know that (a) twins' parents and teachers and twins themselves rate a common phenotype that is highly heritable, (b) these

ratings also reflect informant specific genetic influences, (c) the estimated contributions of both the common and specific genetic influences differ across informants, (d) the estimated relative contribution of specific genetic influences on same teacher's ratings is larger than those on different teachers' ratings, and (e) agreement between mother and father ratings is stronger than agreement between parent and teacher ratings. *Why* the contributions of genetic effects differ across the informant ratings is still unclear, however. To date, the informant dependency of heritability has not been resolved. As a result the implications of this dependency have not been fully addressed in the literature.

The major and general aim of the present paper is to clarify the informant dependency of heritability and to consider its implications. An additional, more specific aim is to investigate whether the dependency can explain the aforementioned abrupt drop in heritability of attention problems in early adolescence that is accompanied by switch in informants. To these ends, we first present a series of genetic analyses of parent, teacher and self-ratings of 12-year-old twins' on the attention problems scales of the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach and Rescorla 2001). The results of these analyses serve to illustrate the informant dependency of heritability and to aid our explanation of this dependency. Next, we discuss the implications of the present and previous findings for future behavior and molecular genetic studies. Throughout the paper, it is important to realize that by analyzing cross-informant data from a homogeneous sample with respect to age, any differences in heritability cannot be attributed to any confounding age effects.

Method

Sample

Data were collected by the Netherlands Twin Register (NTR). The NTR's methods of recruitment and the NTR participants' background and response rates have been described in detail elsewhere (Bartels et al. 2007a; van Beijsterveldt et al. 2013).

The database contained ratings of twins' behavioral problems on items from the ASEBA at the time the twins (birth cohorts 1986–2001, $n = 17{,}757$) were 12 years old. Of 5.75 % of these twins data on blood or DNA polymorphisms were available, from which zygosity was obtained. If this information was absent zygosity was deduced from opposite-gender information or from the responses on validated questions (e.g. about physical resemblance). Of 0.42 % of the twin pairs zygosity was undetermined due to insufficient information or invalid responses; their data were excluded from further statistical analyses. Of 705 monozygotic male, 866 monozygotic female, 701 dizygotic male, 666 dizygotic female, and 1,305 dizygotic opposite gender twin pairs classroom information was available, next to ratings of attention problems. This group constituted the subject sample.

The subject sample consisted of 4,008 males and 4,304 females from 4,237 families, comprising 4,001 complete and 242 incomplete twin pairs. The number of twin pairs from whom (at least from one of the twins) at least two, three, and four AP ratings were available was 3,558, 2,758, and 484. The number of complete twin pairs of whom both twins received four ratings was 378. Of the twins pairs 893 MZ and 678 DZ pairs shared their classroom and 1,251 MZ and 1,421 DZ pairs did not.

Measures

The twins 'behavioral problems were rated by the twins' mothers, fathers, teachers, or by the twins themselves. Mothers and fathers filled out the Child Behavior Checklist (CBCL), teachers the Teacher Report Form (TRF), and the twins the Youth Self Report/11-8 (YSR) (Achenbach and Rescorla 2001).

Total scores (item sum scores or item mean scores) on the CBCL, TRF, and YSR AP scales were considered to constitute indicators of the same phenotype (Achenbach and Rescorla 2001, p. 28)—i.e., the individual's level of attention problems. We stress this does not (necessarily) imply that the phenotype was considered to be static or one-dimensional, for example. Firstly, as mentioned, there is ample evidence that attention problems are context dependent. Secondly, certain attention problems that are assessed by the AP scales can be considered more strongly related to diagnostic criteria of ADHD symptoms of inattention (e.g., day dreaming) and others to hyperactivity (e.g., having trouble to sit still) or impulsivity (e.g., acting without thinking). Thirdly, the items on the AP scales of the TRF (20 items), CBCL (11 items), and YSR (9 items) AP scales do not overlap perfectly (see the sample forms of the CBCLC, TRF and YSR at the ASEBA website; http://www.aseba.org/forms.html). An example of an overlapping item is: '[my child] can't concentrate, can't pay attention for long'(CBCL), '[the pupil] can't concentrate, can't pay attention for long'(TRF) and 'I have trouble concentrating or paying attention' (YSR). An example of a non-overlapping item is '[the pupil has] difficulty following directions', as this item is only included in the TRF.

Because the number of items on the AP scales differed across the instruments, internal consistencies (Cronbach's α's) differed considerably. These were 0.77 (for both mother CBCL reports and father CBCL reports), 0.90 (TRF) and 0.70 (YSR). Sum scores were also incomparable.

Table 1 Descriptive statistics of attention problem ratings on the Child Behavior Checklist (CBCL), Teacher Report Form (TRF) and Youth Self Report (YSR) of monozogotic (MZ) and dizygotic (DZ) twin pairs for whom classroom information was available

	Instruments (and informants)			
	CBCL		TRF	YSR
	(Mothers)	(Fathers)	(Teachers)	(Twins)
n_{MZ}	1,330	1,017	1,569	261
n_{DZ}	2,156	1,676	2,669	373
μ_{MZ}	0.22 (0.006)	0.21 (0.007)	0.22 (0.004)	0.48 (0.018)
μ_{DZ}	0.22 (0.021)	0.21 (0.005)	0.26 (0.005)	0.48 (0.013)
SD_{MZ}	0.25 (0.004)	0.24 (0.003)	0.27 (0.004)	0.33 (0.012)
SD_{DZ}	0.26 (0.003)	0.25 (0.005)	0.30 (0.003)	0.31 (0.008)
r_{MZ}	0.74 (0.013)	0.74 (0.014)	0.69 (0.014)	0.57 (0.043)
r_{DZ}	0.22 (0.021)	0.32 (0.022)	0.23 (0.022)	0.23 (0.052)

Standard errors within parentheses

n number, μ mean, *SD* standard deviation, *r* twin correlation

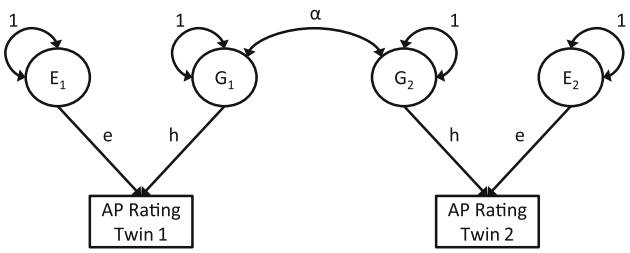

Fig. 2 Standardized behavior genetic model. The observed variable attention problems (AP) is regressed on latent genetic (G) and (nonshared) environmental (E) variables, with regression coefficients *h*, and *e*. Variance in the latent variables is set at 1, such that α represents the genetic correlation between twins

In view of the latter analyses of item mean AP scores were preferred, because these were comparable across the instruments, owing to the fact that all items on the TRF, CBCL, and YSR are scored on a three-point scale. Possible answer categories are 'Not true' (scored 0), 'Somewhat or sometimes true' (scored 1) and 'Very true or often true' (scored 2).

Statistical analyses

From the MZ and DZ correlations between the mean item AP scores (see Table 1), it can be inferred that according to all informants (mothers, fathers, teachers, and the twins themselves) MZ twins resembled each other in AP scores to a greater extent than DZ twins, which implies individual differences in attention problems reflect—to some extent—individual differences in genetic makeup. We note our aim was to clarify the informant dependency of heritability to explain previous empirical results, and not, for example, to find the best model fit. We therefore employed a standard univariate twin model (Plomin et al. 2012) to obtain heritability estimates of individual differences in attention problems, as is common practice in genetic research. That is, we decomposed the variance in item mean AP scores into genetic and environmental variance components (see Fig. 2, for a graphical representation of the model). Such decomposition is possible in a twin design since monozygotic twins share (nearly) 100 % of their genetic material whereas dizygotic twins share (on average) 50 % of segregating genes.

Because the MZ correlations were more than twice as high as the DZ correlations, which is generally interpreted as evidence for nonadditive genetic effects (but see the discussion), genetic variance was further decomposed into additive genetic variance and non-additive genetic variance (Plomin et al. 2012). As previous behavior genetic research indicated no sex differences in MZ and DZ correlations or in heritability estimates of AP (Vink et al. 2012; Rietveld et al. 2003), and to increase the power to detect differences in heritability across informants, data for boys and girls were collapsed. However, because previous research did report mean differences in AP between boys and girls (Rietveld et al. 2003; Kan et al. 2013a), we included gender as a binary predictor of AP. To allow for possible differences in heritability between twins who shared their classroom environment or teacher and twins who did not (Lamb et al. 2012), the model was fitted separately in the groups of twin pairs of whom it was known that they shared their classroom or teacher ('same teacher group') or that they did not ('different teachers group'). The same teacher group comprised 893 MZ and 1,251 DZ pairs; the different teachers group 678 MZ and 1,421 DZ pairs.

To check whether the use of twin pair ratings provided by different informants yields different heritability estimates than the use of twin pair ratings provided by the same informant, and to control for instrument effects, we first selected mother (CBCL) ratings of the one twin and father (CBCL) ratings of the other, and fitted the model on these' different parents' ratings', thereby allowing for mean differences between mother and father ratings. Next, we fitted the model on the mother (CBCL) ratings of both twins as well as on the father (CBCL) ratings of both twins. The results were compared with the results from model fits on the teacher (TRF) ratings and on (YSR) self-ratings (in both the same teacher group and the different teacher group). From the model fitting results we derived broad-sense heritability coefficients (i.e., h^2), which denote the relative contribution of estimated total genetic variance (i.e. the sum of additive and non-additive genetic variance) to phenotypic variance (i.e. the sum of environmental and total genetic variance).

All analyses were performed in the statistical program Mplus, version 6.11 (Muthén and Muthén 1998–2011).

Table 2 Results from the univariate genetic analyses of attention problems

	Informant(s)					
Same teacher group	Mother–father	Father–mother	Mother–mother	Father–father	Teacher 1–teacher 1	Twin 1–twin 2
Number of twin pairs [MZ, DZ]	[787, 1,040]	[787, 1,040]	[786, 1,035]	[604, 805]	[892, 1,250]	[152, 174]
Additive genetic variance	0.019 (0.009)	0.016 (0.008)	0.022 (0.007)	0.034 (0.003)	0.010 (0.008)	0.035 (0.030)
Nonadditive genetic variance	0.007 (0.010)	0.011 (0.008)	0.014 (0.007)	0.000 (0.000)	0.041 (0.009)	0.015 (0.031)
Total genetic variance	0.026 (0.003)	0.026 (0.002)	0.036 (0.002)	0.034 (0.003)	0.051 (0.003)	0.050 (0.006)
Environmental variance	0.023 (0.002)	0.022 (0.002)	0.014 (,001)	0.013 (0.002)	0.015 (0.001)	0.043 (0.005)
Phenotypic variance	0.050 (0.002)	0.048 (0.002)	0.050 (0.002)	0.047 (0.002)	0.066 (0.003)	0.093 (0.005)
Heritability in percentages	52.8 (4.31)	54.9 (3.89)	72.6 (2.58)	71.3 (3.44)	77.8 (2.04)	53.4 (5.44)
Different teacher group	Mother–father	Father–mother	Mother–mother	Father–father	Teacher 1–teacher 2	Twin 1–twin 2
Number of twin pairs [MZ, DZ]	[544, 1,101]	[546, 1,099]	[544, 1,081]	[413, 844]	[677, 1,250]	[109, 193]
Additive genetic variance	0.008 (0.012)	0.000 (0.000)	0.002 (0.010)	0.024 (0.011)	0.033 (0.014)	0.036 (0.037)
Nonadditive genetic variance	0.035 (0.014)	0.040 (0.003)	0.054 (0.011)	0.032 (0.012)	0.013 (0.016)	0.020 (0.036)
Total genetic variance	0.043 (0.004)	0.040 (0.004)	0.057 (0.003)	0.056 (0.004)	0.047 (0.005)	0.056 (0.008)
Environmental variance	0.032 (0.003)	0.033 (0.003)	0.020 (0.002)	0.015 (0.002)	0.043 (0.005)	0.045 (0.007)
Phenotypic variance	0.074 (0.003)	0.073 (0.003)	0.076 (0.003)	0.071 (0.003)	0.090 (0.003)	0.101 (0.007)
Heritability in percentages	57.4 (4.12)	55.2 (3.52)	74.4 (2.32)	78.7 (2.49)	52.0 (5.02)	55.3 (6.40)

Standard errors of parameter estimates within round brackets

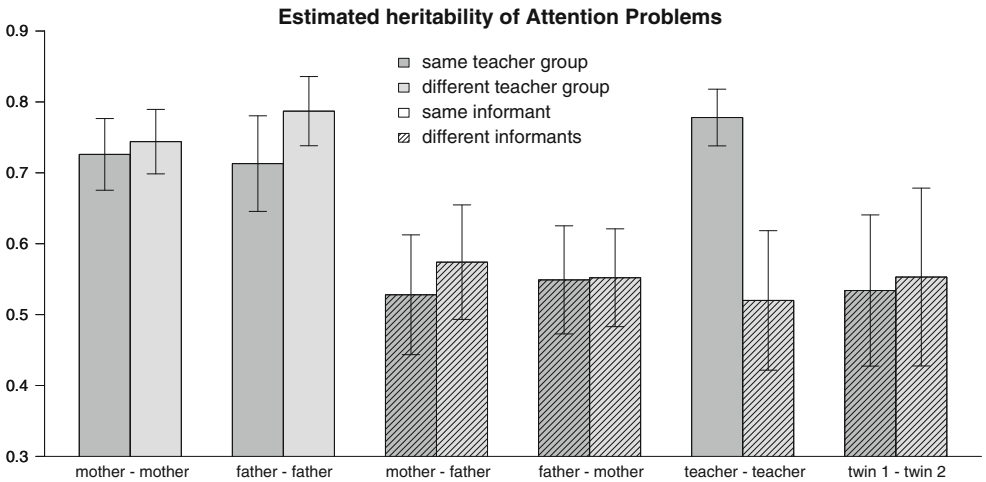

Fig. 3 Broad-sense heritabilities of parent ratings, teacher ratings, and self-ratings of AP in both the 'same teacher' group and the 'different teachers' group

Because the distribution of AP mean item scores was skewed, these analyses were carried out using robust raw data maximum likelihood estimation, which allows for the analysis of non-normally distributed, continuous outcome variables.

Results

Table 2 and Fig. 3 summarize the results of the genetic analyses. In both the same and different teacher group, $h^{2\prime}$s based on different parents' ratings was relatively low, whereas $h^{2\prime}$s based on same parent's ratings (mother ratings and father ratings) were relatively high. The former were comparable to $h^{2\prime}$s based on self-ratings and different teachers' ratings, while the latter were comparable to h^2 based on same teacher ratings.

That heritability estimates for different parents' and same parent's ratings differ, implies that using the same data one can arrive at different conclusions concerning the heritability of a trait. Before we provide an explanation of this finding, we note that certain effects can be ruled out as

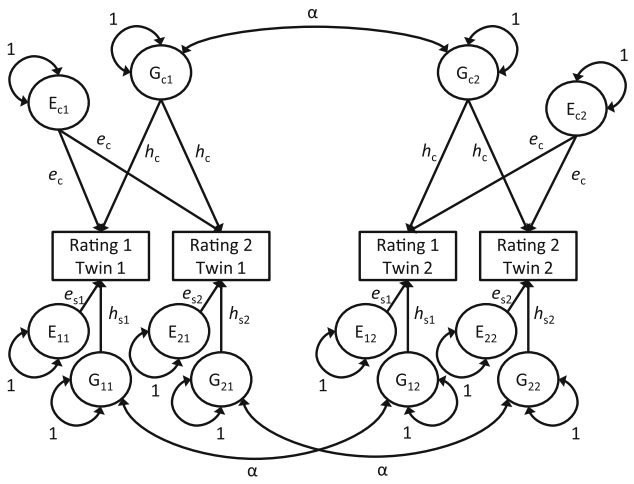

Fig. 4 Standardized multiple informant genetic model. Ratings are provided by Rater 1 and Rater 2. The variance in the ratings is explained by common (c) and informant (or context) specific (s) genetic (G) and nonshared environmental (E) factors. Symbols h and e denote their (standardized) regression coeffcients; α denotes genetic resemblance. Common genetic factors contribute to the correlation between ratings from the same informant and to correlations between ratings from different informants. The specific genetic factors contribute only to the correlation between ratings from the same informant

constituting an account for our observations. First, age effects were ruled out by design. Second, because reliability coefficients of a test do not change due to the analysis, of course, the difference in h^2 between same parent's and different parents' ratings cannot be attributed to differences in reliability. That low reliability accounts for the relatively low heritability of self-ratings, for example, can also be ruled out, because even though self-reports could be considered less reliable than parent or teacher ratings (based on the internal consistencies), h^2 for self-ratings was not lower than h^2 for different teachers' ratings. Lastly, because mothers and fathers rated the behaviors on the same items, instrument effects could also be ruled out (which would not be the case if parent versus teacher ratings would have been analyzed, for example).

It becomes clear that the differences in h^2 did not so much reflect who the informants were, but whether the ratings of a twin pair came from one and the same informant (mother, father, or same teacher) or from different informants (different parents, different teachers, or the two twins), even when informants rate the same kind of attention problems. In the first case, h^2 was around 75 %, in the latter case around 54 %.

Conclusion

How do we explain the 'same versus different informants' dependency of h^2? As a first step, we abandon the strict measurement model from Fig. 1 and imagine the presence of common and informant specific influences (which are not distinguishable in a univariate analysis). Consider the model in Fig. 4, which is a bivariate behavior genetic ('biometric') model. It denotes that observed phenotype variance is the result of variance in genetic factors (G), and (non-shared) environmental factors (E). Part of the genetic and environmental variance is common to the ratings, while another part is informant specific. The informant specific environmental part includes measurement error variance. Suppose this model reflects the true data generating mechanism. To aid the interpretation, also suppose that the model is standardized, such that all variables have a variance of 1 and all paths coefficients and covariances can be interpreted in terms of correlations.

Next, recall that in twin studies heritability is not directly obtained by measuring environmental and genetic variance, but derived from a comparison between the phenotypic similarity in MZ twins and the phenotypic similarity in DZ twins, hence by comparing differences in covariance structure. Also recall that in the absence of shared environmental factors, 'MZ twin correlations are direct estimates of heritability' (Falconer and Mackay 1996). At this point, we encounter ambiguity, because it is not spelled out which MZ correlation needs to be considered; the MZ correlations between ratings from the same informant (e.g., the ratings of twin 1 and twin 2 from informant 1, or, alternatively, the ratings of twin 1 and twin 2 from informant 2) or the MZ correlations between ratings from different informants (e.g., the rating of twin 1 from informant 1 and the rating of twin 2 from informant 2, or, alternatively, the rating of twin 1 from informant 2 and the rating of twin 2 from rater 1).

First consider the MZ correlations between same informant's ratings, e.g., the correlation between the ratings of twin 1 and twin 2 from informant 1. Using the rules of path tracing (Wright, 1934), it can be derived from the model in Fig. 4 that this correlation equals $h_c \cdot \alpha \cdot h_c + h_{s1} \cdot \alpha \cdot h_{s1}$, where α denotes genetic resemblance. In MZ twins $\alpha = 1$, such that the expression reduces to $h_c^2 + h_{s1}^2$. Similarly, it can be derived that the MZ correlation between the ratings of twin1 and twin 2 from informant 2 equals $h_c^2 + h_{s2}^2$. Hence, when analyzing mother ratings, father ratings, or same teacher ratings, heritability estimates reflect the sum of the relative contributions of the genetic variance that is shared between the ratings as well as the variance that is due to informant specific genetic influences.

Next, consider the MZ correlations between different informants' ratings, e.g., the correlation between the rating of twin 1 from informant 1 and the rating of twin 2 from informant 2. Using path tracing again, it can be derived that these equal $h_c \cdot \alpha \cdot h_c$, which reduces to h_c^2. In other words, when analyzing different teacher ratings or self-ratings

(or, as in the present paper, different parent ratings), heritability estimates merely reflect the contribution of the genetic variance that is shared between the ratings. Any variance due to informant specific genetic factors, whether additive or non-additive, will be attributed to the residual, nonshared environmental component.

With this knowledge in mind, examine the results in Table 2 again, from which it can be obtained that phenotypic variance in different parents' ratings is of about the same magnitude as the phenotypic variance in same parent's ratings. Yet, the estimated (total) genetic variance in the former is smaller than in the latter, while the estimated environmental variance is larger. These results reflect the fact that the informant specific part of the genetic variance gets assigned to the residual nonshared environmental variance component in case ratings from different informants are analyzed.[1] As a consequence, h^2 based on different parent's ratings is smaller than h^2 based on same parent's ratings.

Because different teachers ratings constitute different informants' ratings and same teacher ratings constitute same informant's ratings, h^2 for the former can expected to be relatively low, similar to h^2 for different parents' ratings compared to h^2 for same parent's ratings. Self-ratings also constitute different informants' ratings (twins did not rate their co-twins' problems) such that their heritability estimates can also be expected to be relatively low. All these expectations are in line with the results in Fig. 2.

Discussion

To clarify how heritability estimates are informant dependent, we analyzed parent, teacher and self-ratings of 12 year old twins' attention problems, the results of which fit with results from previous research (e.g., Merwood et al. 2013; Lamb et al. 2012). Heritability estimates of AP based on mother, father, (same and different) teacher ratings and self-ratings were similar to those reported in the literature.

By analyzing mother ratings of the one twin member and father ratings of the other twin member ('different parents' ratings'), we illustrated that h^2 for different parents' ratings is lower than h^2 for same parent ratings, and comparable to h^2 based on different teachers' and self-ratings. Such a result had not been reported previously, but provided important information, because from these results we were able to conclude that empirically h^2 of AP ratings on the same instrument depended mainly on whether twin pair ratings are provided by the same informant or by different informants. By means of path tracing, we demonstrated that our results are expected when informant specific genetic factors are present.

As mentioned in the introduction, the presence of such informant specific genetic factors have been established in numerous traits. The interpretation of the heritabilities of these traits, and the associated disorders themselves, thus requires caution. When they are based on different informants' ratings (self-ratings included), they will reflect the contribution of the genetic factors that are common to the informants' ratings; the contribution of informant specific genetic variance will be attributed to the residual, hence to nonshared (environmental) factors. When they are based on same informants' ratings heritability estimates include, additionally, the contribution of these informant specific genetic factors.

That informant specific factors are so widely present is consistent with the idea that many behaviors are context, situation, or environmental dependent (e.g., Lamb et al. 2012): Individual differences in the behavior in a specific context (e.g. classrooms, or home) can be highly genetically influenced, while specific situations can depend on genetically influenced behavior. It thus also fits with the idea of dynamical interactions between genes, behavior, and environment: When an individual exhibits certain behavior, which is genetically influenced, people (e.g., mothers, fathers, teachers) will respond to this behavior, and probably they do so in their own ways. As a result, the responses change the environmental circumstances, which in turn may alter the individual's behavior, the individual's gene-expression, and thus the genetic effects. In the end, nonshared environmental influences can thus give rise to nonshared genetic effects and to rater disagreement. The more environments differ, the less genetic variance tends to be shared. This provides an explanation why genetic overlap is larger and rater agreement between mother and father ratings is stronger than genetic overlap and rater agreement between parent and teacher ratings (e.g., Merwood et al. 2013): Both mothers and fathers typically observe and interact with their children at home, while teachers do so in classroom environments. Instrument effects cannot be ruled out however. On the other hand, one could also argue that instrument effects are explained by the context dependency, because the items of the CBCL, TRF, and YSR refer (in part) to different situations; different items will assess the effects of different dynamical genotype-environment interactions.

Dynamical interactions between genetic and environmental factors are important sources of changes in heritability (Molenaar et al. 1993; Turkheimer 2004; Kan et al. 2010) and of statistical genotype-environment effects (Carey 1986). They can lead to statistical genotype-

[1] Fitting a bivariate model on mother and father AP ratings revealed that the contribution of common genetic variance to phenotypic variance was indeed around 55 %, whereas the total estimated heritability coefficients were on average 75 %. For obvious reasons bivariate models could not be fitted on same and different teacher ratings or on self-ratings.

environment interaction, for instance, or to genotype-environment covariance. Statistical genotype-environment interaction refers to the fact that the relative contributions of environmental and genotypic variance to phenotypic variance vary across genotype (Plomin et al. 2012). Genotype-environment covariance refers to the situation in which the environmental factors that contribute to individual differences in a trait differ non-randomly across individuals as function of genotype (Plomin et al. 2012).

Both dynamical and statistical genotype-environment effects are receiving increasing attention from the scientific community. In this line of research, it is important to be aware of informant dependency. Take research into negative sibling interaction, for instance, which denotes a form of dynamical interaction where behavior of the one twin leads to behavior in the co-twin along the same dimension, but in the opposite direction. Negative sibling interaction leads, similarly to dominance, to DZ correlations that are more than twice as low as MZ correlations (Eaves et al. 1997; Rietveld et al. 2003). Comparable results will be observed when an informant effect called contrast effect is present (Simonoff et al. 1998). This effect denotes the situation in which the informant exaggerates phenotypic differences within pairs of relatives. Contrast effects, negative sibling interaction, and genetic dominance thus form alternative explanations to another. Increasingly, efforts are being made to disentangle these effects. In this line of research, researchers often use multiple informant models. Here, the need arises to distinguish between 'multiple same informants', 'multiple different informants' and 'a mix of same and different informants', because contrast effects can be assumed to be stronger in same informant's ratings than in different informants' ratings.

Awareness of the informant dependency of heritability is also relevant in other research contexts. The relevance to clinical research is obvious: Not only heritability estimates based on mother, father, teacher and self-ratings are informant dependent, but also those based on clinical diagnoses. After all, diagnoses can be established by the same diagnostician or by different diagnosticians (or by multiple diagnosticians, e.g., a team).

The relevance to developmental behavior research becomes clear when considering that certain phenotypes demonstrate relatively low heritability in adulthood and high heritability in childhood. Because informant dependency constitutes a source of systematical heterogeneity in heritability estimates, the dependency is an important additional explanation over age dependency of heritability. The shift from reliance on reports by others to reliance on self-reports provides a valid account of the drop in heritability of AP and ADHD symptoms in adolescence, as has been surmised (Kan et al. 2013a; Merwood et al. 2013; Chang et al. 2013), but had not yet been proven. The effect contributes, and perhaps fully explains, the difference between estimated heritability of AP and ADHD symptoms in adulthood versus childhood.

The relevance to molecular genetic investigations, such as genome-wide association studies (GWAS) and genome-wide complex trait analysis (GCTA), lies in premise that detectable genetic variants explain heritability. Empirically, there is often a discrepancy between, on the one hand, the proportion of phenotypic variance that is accounted for in GWAS (relatively low) and, on the other hand, in twin and family studies (relatively high) (Manolio et al. 2009) and in GCTA (in between) (Yang et al. 2011). Part of the genetic variance in complex traits thus seems unexplained, which is holds especially true for problem behaviors (Trzaskowski et al. 2013). In general, the phenomenon is referred to as the problem of 'missing heritability'. Explanations of missing heritability are usually sought in complex genetic mechanisms, such as epistasis, or in the contributions of rare genetic variants that go undetected in GWAS of common single-nucleotide polymorphisms (SNPs). Although these genetic mechanisms may contribute to the discrepancies, it must also be acknowledged that the discrepancies will often be smaller than assumed. After all, in GWAS, but also in adoption studies for example, one usually does not rely on ratings from one and the same informant (or diagnostician), but on ratings from many different informants (or diagnosticians), who have rated behaviors in many different contexts. In these research designs heritability can thus be expected to be relatively low, and should not to be compared with the relatively high heritability estimates derived from twin studies, especially when it concerns heritability estimates based on same informant ratings. In GWAS and GCTA, context, situation, hence rating specific genetic effects will be undetectable.

Acknowledgments We gratefully acknowledge the following Grants: Genetic influences on stability and change in psychopathology from childhood to young adulthood (ZonMW 912-10-020); Genetics of Mental Illness (European Research Council; ERC-230374) and support from the Neuroscience Campus Amsterdam (NCA). We thank Tina Glasner, Evelien de Zeeuw, Michelle van Fulpen, and Cyrina Brouwer for their assistance in the acquisition of the data and for administrative, technical, and material support.

Conflict of Interest Kees-Jan Kan, Catharina E. M. van Beijsterveldt, Meike Bartels, and Dorret Boomsma declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Informed consent was obtained from the parents of all participants for being included in the study.

References

Achenbach T, Rescorla LA (2001) Manual for the ASEBA school-age forms & profiles. University of Vermont, Research Center for Children, Youth, and Families, Burlington

Arseneault L, Moffit TE, Caspi A, Taylor A, Rijsdijk FV, Jaffee SR, Ablow JC, Measelle JR (2003) Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins self-reports. J Child Psychol Psychiatry 44(6):832–848

Bartels M, Hudziak JJ, Boomsma DI, Rietveld MJ, van Beijsterveldt CEM (2003a) A study of parent ratings of internalizing and externalizing problem behavior in 12-year-old twins. J Am Acad Child Adolesc Psychiatry 42(11):1351–1359

Bartels M, Hudziak JJ, van den Oord EJCG, van Beijsterveldt CEM, Rietveld MJH, Boomsma DI (2003b) Co-occurrence of aggressive behavior and rule-breaking behavior at age 12: multi-rater analyses. Behav Genet 33(5):607–621

Bartels M, van Beijsterveldt CEM, Derks EM, Stroet TM, Polderman TJ, Hudziak JJ, Boomsma DI (2007a) Young Netherlands twin register (y-ntr): a longitudinal multiple informant study of problem behavior. Twin Res Hum Genet 10(1):3–11

Bartels M, Boomsma DI, Hudziak JJ, van Beijsterveldt CEM, Oord EJCG (2007b) Twins and the study of rater (dis) agreement. Psychol Methods 12(4):451–466

Boomsma DI, van Beijsterveldt CEM, van Hudziak JJ (2005) Genetic and environmental influences on anxious/depression during childhood: a study from the Netherlands twin register. Genes Brain Behav 4(8):466–481

Boomsma DI, Saviouk V, Hottenga JJ, Distel MA, de Moor MHM, Vink JM, Geels LM, van Beek JH, Bartels M, de Geus EJ, Willemsen G (2010) Genetic epidemiology of attention deficit hyperactivity disorder (ADHD index) in adults. PLoS One 5(5):e10621

Burt SA, McGue MK, Krueger RF, Iacono WG (2005) Sources of covariation among the child-externalizing disorders: informant effects and the shared environment. Psychol Med 35(8):1133–1144

Carey G (1986) Sibling imitation and contrast effects. Behav Genet 16(3):319–341

Chang Z, Lichtenstein P, Asherson PJ, Larsson H (2013) Developmental twin study of attention problems high heritabilities throughout development. JAMA Psychiatry 70(3):311–318

Derks EM, Hudziak JJ, van Beijsterveldt CEM, Dolan CV, Boomsma DI (2006) Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. Behav Genet 36(6):833–844

Derks EM, Hudziak JJ, Boomsma DI (2009) Genetics of ADHD, hyperactivity, and attention problems. In: Kim YK (ed) Handbook of behavior genetics. Springer, New York

Eaves L, Silberg J, Meyer J, Maes H, Simonoff E, Pickles A, Rutter M, Neale MC, Reynolds C, Erickson M, Heath A, Loeber R, Truett K, Hewitt H (1997) Genetics and developmental psychopathology: 2 the main effects of genes and environment on behavioral problems in the Virginia twin study of adolescent behavioral development. J Child Psychol Psychiatry 38(8):965–980

Falconer D, Mackay T (1996) Introduction to quantitative genetics. Longman, New York

Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, Sklar P et al (2005) Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry 57(11):1313–1323

Haberstick B, Timberlake D, Hopfer C, Lessem J, Ehringer M, Hewitt J (2008) Genetic and environmental contributions to retrospectively reported DSM-IV childhood attention deficit hyperactivity disorder. Psychol Med 38(7):1057–1066

Hartman CA, Rhee SH, Willcutt EG, Pennington BF (2007) Modeling rater disagreement for ADHD: are parents or teachers biased? J Abnorm Child Psychol 35(4):536–542

Haworth C, Wright M, Luciano M, Martin N, De Geus EJ, Van Beijsterveldt CEM, Bartels M, Posthuma D, Boomsma D, Davis O et al (2009) The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry 15(11):1112–1120

Hewitt JK, Silberg JL, Neale MC, Eaves LJ, Erickson M (1992) The analysis of parental ratings of children's behavior using LISREL. Behav Genet 22(3):293–317

Hoekstra RA, Bartels M, Hudziak JJ, van Beijsterveldt CEM, Boomsma DI (2008) Genetic and environmental influences on the stability of withdrawn behavior in children: a longitudinal, multi-informant twin study. Behav Genet 38(5):447–461

Jang KL, Lam RW, Livesley WJ, Vernon PA (1997) Gender differences in the heritability of seasonal mood change. Psychiatry Res 70(3):145–154

Kan KJ, Ploeger A, Raijmakers MEJ, Dolan CV, van der Maas HLJ (2010) Nonlinear epigenetic variance: review and simulations. Dev Sci 13(1):11–27

Kan KJ, Dolan CV, Nivard MG, Middeldorp CM, van Beijsterveldt CEM, Willemsen G, Boomsma DI (2013a) Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry 52(1):12–25

Kan KJ, Wicherts JM, Dolan CV, van der Maas HLJ (2013b) On the nature and nurture of intelligence and specific cognitive abilities: the more heritable, the more culture-dependent. Psychol Sci 24(12):2420–2428

Kendler K, Gardner C, Lichtenstein P (2008) A developmental twin study of symptoms of anxiety and depression: evidence for genetic innovation and attenuation. Psychol Med 38(11): 1567–1575

Lamb DJ, Middeldorp CM, van Beijsterveldt CEM, Boomsma DI (2012) Gene-environment interaction in teacher-rated internalizing and externalizing problem behavior in 7-to 12-year-old twins. J Child Psychol Psychiatry 53(8):818–825

Larsson H, Asherson P, Chang Z, Ljung T, Friedrichs B, Larsson J, Lichtenstein P (2012) Genetic and environmental influences for adult attention-deficit hyperactivity disorder symptoms: a large swedish population-based study of twins. Psychol Med 1(1):1–11

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753

Martin N, Scourfield J, McGuffin P (2002) Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. Br J Psychiatry 180(3):260–265

McLoughlin G, Rijsdijk F, Asherson P, Kuntsi J (2011) Parents and teachers make different contributions to a shared perspective on hyperactive-impulsive and inattentive symptoms: a multivariate analysis of parent and teacher ratings on the symptom domains of ADHD. Behav Genet 41(5):668–679

Merwood A, Greven C, Price T, Kuntsi J, McLoughlin G, Larsson H, Asherson P (2013) Different heritabilities but shared aetiological influences for parent, teacher and self-ratings of ADHD symptoms: an adolescent twin study. Psychol Med 43(9):1973–1984

Molenaar PCM, Boomsma DI, Dolan CV (1993) A third source of developmental differences. Behav Genet 23(6):519–524

Muthén LK, Muthén BO (1998–2011) Mplus user's guide, 6th edn. Muthén & Muthén, Los Angeles

Plomin R, DeFries J, Knopik V, Neiderhiser J (2012) Behavioral genetics, 6th edn. Worth, New York

Polderman TJ, Posthuma D, De Sonneville LM, Verhulst FC, Boomsma DI (2006) Genetic analyses of teacher ratings of problem behavior in 5-year-old twins. Twin Res Hum Genet 9(1):122–130

Rice F, Harold G, Thapar A (2002) The genetic aetiology of childhood depression: a review. J Child Psychol Psychiatry 43(1):65–79

Rietveld M, Posthuma D, Dolan CV, Boomsma DI (2003) ADHD: sibling interaction or dominance: an evaluation of statistical power. Behav Genet 33(3):247–255

Saviouk V, Hottenga JJ, Slagboom EP, Distel MA, de Geus EJ, Willemsen G, Boomsma DI (2011) ADHD in Dutch adults: heritability and linkage study. Am J Med Genet B 156(3):352–362

Sherman DK, McGue MK, Iacono WG (1997) Twin concordance for attention deficit hyperactivity disorder: a comparison of teachers' and mothers' reports. Am J Psychiatry 154(4):532–535

Simonoff E, Pickles A, Hervas A, Silberg J, Rutter M, Eaves L (1998) Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. Psychol Med 28(4):825–837

Trzaskowski M, Dale PS, Plomin R (2013) No genetic influence for childhood behavior problems from DNA analysis. J Am Acad Child Adolesc Psychiatry 52(10):1048–1056

Turkheimer E (2004) Spinach and ice cream: why social science is so difficult. In: DiLalla L (ed) Behavior genetics principles: perspectives in development, personality, and psychopathology. Am Psychol Assoc, Washington, DC

van Beijsterveldt CEM, Groen-Blokhuis M, Hottenga JJ, Franic S, Hudziak JJ, Lamb D, Huppertz C, de Zeeuw E, Nivard MG, Schutte N et al (2013) The young Netherlands twin register (yntr): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet 16(1):252–267

van den Berg SM, Willemsen G, de Geus EJ, Boomsma DI (2006) Genetic etiology of stability of attention problems in young adulthood. Am J Med Genet B 141(1):55–60

Vink JM, Bartels M, van Beijsterveldt TC, van Dongen J, van Beek JH, Distel MA, Boomsma DI (2012) Sex differences in genetic architecture of complex phenotypes? PLoS One 7(12):e47371

Wood AC, Rijsdijk F, Saudino KJ, Asherson P, Kuntsi J (2008) High heritability for a composite index of children's activity level measures. Behav Genet 38(3):266–276

Wright S (1934) The method of path coefficients. Ann Math Stat 5(3):161–215

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1): 76–82