From Structural Equation Models to Next-Generation Sequencing: The Evolving Landscape of Modern Behavioral Genetics

Sanja Franić

Reading committee
Prof. Dr. Simon E. Fisher
Prof. Dr. Han L.J. van der Maas
Dr. Stephanie M. van den Berg
Dr. Jan B. Hoeksma
Prof. Dr. Albertine J. Oldehinkel
Dr. Marieke E. Timmerman

Paranymphs
Maria Groen-Blokhuis
Michel Nivard

ISBN: 978-94-6259-384-8
Print: Ipskamp Drukkers (http://www.ipskampdrukkers.nl/)
Layout: Sanja Franić
Cover design: Sanja Franić & Zvonimir Mandić (zvonimir.mandic@gmail.com)

VRIJE UNIVERSITEIT

From Structural Equation Models to Next-Generation Sequencing: The Evolving Landscape of Modern Behavioral Genetics

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Psychologie en Pedagogiek
op vrijdag 21 november 2014 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Sanja Franić

geboren te Karlovac, Kroatië

promotoren:	prof.dr. D.I. Boomsma
	prof.dr. C.V. Dolan
copromotor:	prof.dr. D. Borsboom

Contents

Chapter 1

Introduction

The past five years have seen vast and rapid changes in the landscape of behavior genetic research. At the time this dissertation was conceived, the field was dominated by studies estimating the proportions of individual differences in behavioral traits attributable to the effects of genes and the environment, via the application of structural equation modeling to covariance structures of phenotypes measured in family members. Beyond heritability estimation, structural equation modeling of genetically informative data enabled the study of a range of more complex etiological issues, including the study of the nature of developmental stability and change in behavioral phenotypes, the genetic and environmental contributions to inter-individual variation in age-related growth and decline, the dependency of genetic effects on environmental exposures, heterogeneity of genetic effects across the sexes, direction of phenotypic causality between traits, presence and magnitude of rater bias, and sibling imitation and contrast effects. The continuing developments in statistical methodology and computing resources over the past five years have facilitated an increased flexibility in the modeling of genetic covariance structures, both in terms of the models one may fit, and in terms of the number of variables that can be included into the analysis. This allowed for increasingly sophisticated analyses, including, for instance, the application of genetically informed item-level analyses in addressing questions regarding the ontology of latent psychological traits (e.g., general cognitive ability, depression), via the study of the mediatory role of the latent traits with respect to genetic and environmental effects (Chapters 3-5). The recent advent of the large-scale availability of measured genotype data has provided an additional impetus for the development of genetic covariance structure modeling, namely for 1) its use in refining the definition of the phenotype in genetic association analyses, and 2) the incorporation of measured genetic variables into structural equation modeling-based association analysis.

Most notably, the past five years have seen a sharp increase in the large-scale use of genomic microarrays and next-generation sequencing technologies, and a subsequent proliferation of gene-finding studies using measured genetic information to identify the genetic variants underlying the observed inter-individual variability. The interrogation of common single-nucleotide polymorphisms (SNPs) along the genome and their use in genome-wide association (GWA) studies were possibly the largest enterprise to this end, yielding over 2,000 associations for over 700 traits in the seven years since their inception (Visscher, Brown, McCarthy, & Yang, 2012). A complementary approach enabling a relatively rapid interrogation of the entire nucleotide sequence of a genome (including rare and structural variation, in addition to common point mutations) was made widely available through the rapid decline in the cost of DNA sequencing, and the ubiquitous move from Sanger sequencing (Sanger, Nicklen, & Coulson, 1977) to next-generation sequencing technologies. The development and the large-scale availability of these technologies, and of the methodology for the analysis of high-throughput data that they generate, have been remarkable. If one were to describe the largest point of progress in behavior genetics over the past five years, it would almost undoubtedly be in terms of the availability of measured genotype data.

The present dissertation reflects the progression of behavior genetic methodology over the past five years. It can be seen as a cross-section of the relevant applications: from

the basics of genetic covariance structure modeling, via its more advanced applications in psychometric dimensionality assessment, the study of the ontology of latent psychological traits (childhood internalizing problems, personality dimensions) and refinement of the phenotype definition in genetic association analyses, to GWA studies and the analysis of next-generation sequencing data. In particular, the dissertation focuses on applications of genetic covariance structure modeling that go beyond simple heritability estimation, dealing primarily with the ontological nature of latent variables employed in psychological research, and the related question of their suitability for use in genetic association studies.

The former topic (i.e., the nature of latent variables) has long generated controversy within psychology. Are latent constructs, as frequently postulated, psychological or biological entities that exist within individuals and manifest themselves in psychometric questionnaire item responses, behaviors, and symptoms (e.g., because one is depressed, one feels dejected and frequently ruminates), or purely statistical constructs that have no existence outside of the realm of abstraction, serving only to summarize clusters of item responses, behaviors, and symptoms (e.g., one is said to be depressed if they feel dejected and frequently ruminate; not because depression is an underlying cause of their symptoms, but because it is the term standardly used to describe them)? Apart from its obvious relevance to psychological theories that employ latent constructs and the psychometric practice that aims to measure them, the ontology of latent variables is highly relevant to behavior genetics, perhaps chiefly due to its implications for genetic association studies. For instance, if depression is nothing more than an index variable used to summarize a cluster of symptoms, efforts to identify genetic variants affecting the liability to depression will be characterized by a lower statistical power relative to a situation in which genes for its constituent symptoms are sought. A more relevant question in the context of genetic association studies is that of genetic unidimensionality: is a given cluster of symptoms affected by a single set of genes? For instance, depression symptoms related to dejected mood and those related to withdrawn behavior may be distinctly affected by environmental factors, but (co)vary as a function of a single set of genes (i.e., the distinction between the symptom clusters may be driven by differential environmental effects on a largely nonspecific genetic predisposition to developing depression). Genetically unidimensional constructs, regardless of the degree of their environmental etiological heterogeneity, generally represent more suitable targets for genetic association studies than genetically multidimensional ones. A related issue is that of genetic and environmental structure of a construct over time. In the presence of longitudinal data, how does one optimally define a target phenotype for a genetic association study? For instance, analyzing measures collected at a single age may be inefficient in terms of discarding other data, while using all measures simultaneously may dilute the genetic signal if different genetic factors affect the phenotype across development. If one opts to utilize all data, can one summarize the measures across time, or should one employ a multivariate approach? One of the relevant questions here is that of genetic unidimensionality over time – does a single set of genes affect the phenotype across the developmental period under study? The present dissertation is based on the idea that the answers to the above questions and, more generally, the treatment of the phenotype in genetic association studies, are consequential (both in terms of the statistical power to detect genetic effects, and in terms of the interpretability of the results obtained), and explores the aforementioned issues with respect to several behavioral phenotypes. The increasing availability of measured genotypic information has allowed me to take a step beyond the exploration of the phenotypic structure of psychological constructs, and complement this work with genetic association studies – initially using genomic microarray data to explore the role of common genetic polymorphisms in the etiology of intelligence,

and subsequently using DNA sequence data to test for effects of rare genetic variation. Although much work remains to be done, both in improving the treatment of the phenotype and in the gene finding domain, the dissertation hopefully opens some possibilities for the investigation of the phenotypic structure that go beyond a simple sum score approach.

The phenotypes of interest over the coming nine chapters are childhood internalizing psychopathology, adult personality, and intelligence. Research to date has proven challenging with respect to these phenotypes, in several regards: a) the ontology of the latent factors that feature in the theories of the respective phenotypes is unclear, b) relatedly, noncontroversial taxonomies have proven difficult to arrive at (as evidenced by the continuing debate over the factor structure of the phenotypes), and c) genetic association studies have thus far not identified genetic variants that would explain more than a minute fraction of their phenotypic variance. Rather than focusing on standard applications that study the etiology of these phenotypes as represented by latent factors or sum scores, this dissertation has inquired what the empirical support for the existence and causal relevance of such latent factors is (i.e., whether we should be conceptualizing these phenotypes as latent factors at all), and hypothesized that the search for genetic variants may be more successful if greater attention was paid to the issue of phenotype modeling.

The outline of this dissertation is as follows. Chapter 2 introduces the basics of genetic covariance structure modeling (GCSM), as applied in the classical twin design. Considering the intrinsic reliance of GCSM on the quantitative genetic theory-based predictions of genetic and environmental resemblance between individuals of differing degrees of genetic relatedness, in Appendix 1 I review how these predictions are derived. Chapters 3-5 discuss genetically informed item-level analyses and their application to the study of the ontology and the genetic and environmental etiology of personality dimensions and childhood internalizing problems (anxiety, depression, withdrawn behavior and somatic complaints). Chapter 6 reviews GCSM-based literature on childhood internalizing problems. Chapters 7-9 focus on the genetics of intelligence. Chapter 7 examines the genetic and environmental etiology of the temporal stability of verbal, nonverbal and general intelligence across ages 5-18; the results obtained in Chapter 7 are subsequently used to inform the definition of the phenotype in the association studies reported in Chapters 8 and 9. Working on the assumption that that the genetic variation affecting normal-range intelligence may be concentrated in the same areas of the genome as that underlying intellectual disability, Chapters 8 and 9 test for an enrichment of a number of intellectual disability genes for polymorphisms associated with normal-range intelligence. Chapter 8 examines 43 intellectual disability genes using a common-variant approach, and Chapter 9 extends this work to 168 candidate genes and examines the possible effects of rare genetic variation.

Chapter 2

Structural Equation Modeling in Genetics

Abstract

The present chapter introduces structural equation modeling, as applied in human quantitative genetics. After introducing the basic method of exploiting familial relationships to infer the effects of unmeasured genetic and environmental factors, the chapter reviews the implementation of models from the structural equation modeling literature into genetically informative designs, and structural equation models developed specifically within genetics. The former include simplex and latent growth curve models; the latter include common and independent genetic factor models, genotype-environment interaction models, sex-limitation models, and direction of causation models. The chapter concludes with a discussion of the incorporation of measured genetic variables into structural equation modeling-based association analysis.

Based on: Franić, S., Dolan, C.V., Borsboom, D., & Boomsma, D.I. (2012). Structural Equation Modeling in Genetics. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 617-635). New York: Guilford Press.

The aim of the present chapter is to discuss structural equation modeling (SEM[1]) as applied in human quantitative genetics. Taking the seminal paper by Martin and Eaves (Martin & Eaves, 1977) as a starting point, the genetic analysis of covariance structures spans a period of over 30 years (see Hottenga & Boomsma, 2008, for a brief history). Martin and Eaves (1977) is the first published account of genetic covariance structure modeling (GCSM) using Maximum Likelihood (ML) estimation in SEM. Although Martin and Eaves used their own programs to fit multivariate twin models, it was soon realized that the LISREL program (Jöreskog & Sörbom, 2006) could be used to fit genetic models (Boomsma & Molenaar, 1986; Cantor, 1983; Fulker, Baker, & Bock, 1983). The adoption of the LISREL program cemented the view of quantitative genetic modeling as a class of structural equation modeling of data observed in family members. In addition, it encouraged the applications of multivariate models developed in SEM (e.g., the common factor, simplex, and growth curve models), and it inspired geneticists to develop their own models. Finally, the incorporation of SEM in genetic modeling resulted in the development of Mx, a SEM program with a flexible matrix syntax, which is well suited to the data structures and modeling requirements of GCSM (Boker et al., 2010; M. C. Neale, 2000).

The present chapter aims to introduce GCSM, as applied in the classical twin design. We first present the basic method of exploiting familial relationships to infer the effects of unmeasured genetic and environmental factors. We then emphasize that any SEM can be incorporated in GCSM of twin data to study the structures of the genetic and environmental covariances matrices. Next, we discuss several models developed specifically in GCSM. These include models which require data collected in twins, pedigrees or adoption designs for identification. Finally, we briefly discuss the recent incorporation of measured genetic variables in GCSM-based association analyses.

Genetic covariance structure modeling

A principal aim of GCSM (Boomsma, Martin, & Neale, 1989; Eaves, Last, Young, & Martin, 1978; Martin & Eaves, 1977; M. C. Neale & Cardon, 1992) is to estimate the contributions of genetic and environmental variables to individual differences in one or more measured variables (i.e., phenotypes). If the genetic and environmental variables are unobserved (latent), their effects are inferred from resemblance among family members in a SEM. However, measured environmental and (or) genetic variables may also be modeled directly (e.g., Cherny, 2008; van den Oord, 2000).

To infer the contributions of unmeasured genetic and environmental variables to the phenotypic variance, quantitative geneticists employ a number of designs, which include individuals in known genetic and environmental relations (Falconer & Mackay, 1996, Mather & Jinks, 1971). Samples of such individuals are called *genetically informative*, because, given various assumptions, genetic and environmental effects are identified in the associated phenotypic covariance structures. The classical twin design, which involves the analysis of phenotypes measured in monozygotic (MZ) and dizygotic (DZ) twins living together, is the best known of such designs (Boomsma, Busjahn, & Peltonen, 2002), but others the such as the adoption design also achieve identification in GCSM.

In GCSM, different classes of genetic and environmental variables are distinguished. A polygenic factor represents the total effects of multiple, possibly very many, genes. A gene refers to a unit of heredity, that resides on a stretch of DNA and codes for a protein or for an RNA chain. Genes are situated at a given chromosomal region, referred to as a locus. If the

[1] We use this abbreviation to refer to modeling and model(s).

gene influences a complex (or a quantitative) trait, the location is referred to as a QTL: a quantitative trait locus. To contribute to phenotypic variation, a gene has to be polymorphic, i.e., different forms of the gene (i.e., different alleles) must exist. The combination of alleles at a locus determines the effect of the gene (Evans, Gillespie, & Martin, 2002; Slagboom & Meulenbelt, 2002). We distinguish between additive polygenic variable(s) (A), which represent the combined additive effects of alleles within and across loci, and genetic dominance variable(s) (D), which represent intra-locus allelic interaction effects. One can also consider interactions between loci (inter-locus nonlinear effects, i.e., epistasis), although in practice such effects are hard to resolve in the non-experimental designs typically used in GCSM. With respect to environmental effects, environmental effects that are shared by family members (shared environment; C), and individual-specific environmental effects (unshared environment; E) are distinguished. In the classical twin design, the latter contribute to the phenotypic differences between the twins, and the former contribute to resemblance among the twins. Note that environmental influences are defined in terms of their *effect*. For instance, twins may be exposed to the shared event of parental divorce, but the effects of divorce on the individual twin pair members may differ. Thus, a shared event can have a specific (unshared) consequence, which will contribute to what is interpreted as specific or unshared environmental effects.

The identification of model parameters in GCSM is achieved by incorporating in the model the information on the degree of genetic and environmental relatedness among different types of relatives (Evans et al., 2002; Falconer & Mackay, 1996; Mather & Jinks, 1971). In the classical twin design the sample consists of MZ and DZ twin pairs. DZ twins share an average of 50% of their polymorphic (also termed segregating) genes and MZ twins share 100% of their genetic material, as they arise from a single fertilized egg. This information is used in model specification as follows: the A factors correlate 1 in MZ twins and .5 in DZ twins, while the D factors correlate 1 in MZ twins and .25 in DZ twins (Falconer & Mackay, 1996). Shared environmental factors C correlate unity across twins, regardless of zygosity, and unshared environmental factors E correlate zero[2].

All designs in GCSM include specific assumptions and limitations. For instance, in the classical twin design, a model including effects of A, C, D, and E is not identified. Researchers must therefore limit their comparisons to submodels including three of the four sources of individual differences, i.e., an ACE or ADE model (or submodels thereof). The DE model is biologically implausible (Falconer & Mackay, 1996). The twin design involves many further assumptions, some of which are mentioned below. For an exhaustive treatment we refer the reader to the literature (e.g., G. Carey, 2009; Plomin, Defries, McClearn, & McGuffin, 2008).

GCSM based on the twin design

GCSM based on the classical twin design can be used to analyze univariate and multivariate data. In the univariate case, the phenotypic measure is regressed on the genetic and environmental variables. For instance, the univariate ACE model can be expressed as:

$$P_{ij} = t + a*A_{ij} + c*C_{ij} + e*E_{ij},$$

[2] Carey (2009) suggested an alternative model in which the C and E latent variables are replaced by a variable T representing total environmental effects, which may correlate among family members to account for shared environmental effects.

where P_{ij} is the continuous phenotypic measure observed in the j-th member (j=1,2) of the i-th twin pair. The genetic (A) and environmental variables (C and E) are unobserved, and as such are subject to standard identifying scaling constraints: the variances are fixed to unity, the means are fixed to zero. The parameter t represents the intercept, i.e., given the scaling constraints, the mean of the phenotype. We assume the phenotypic means of the twin pair members are equal (a testable assumption of the twin model). The parameters a, c, and e are regression coefficients that represent the effects of the A, C, and E factors on the phenotype.

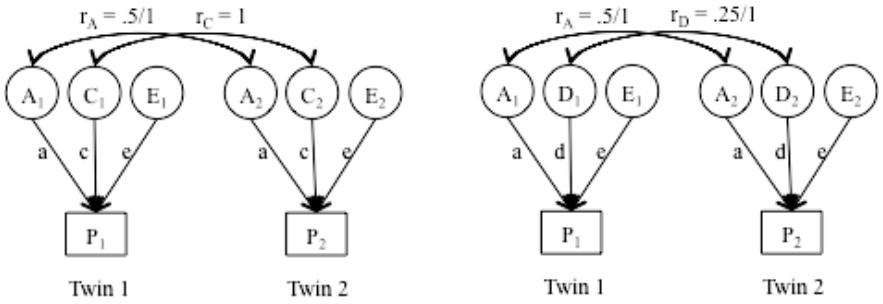

Figure 1. An ACE (left) and an ADE (right) univariate genetic factor model.

Figure 1 depicts two examples of a univariate model for twin data. Assuming the variables have been centered, we can drop the intercept t from the path diagrams. The path diagrams graphically convey some of the assumptions associated with the twin model. For instance, barring the correlations as depicted, the A, C (D), and E variables are uncorrelated within and between twin pair members. The zero correlations between A and D, and between E and C, follow from their definitions. However, certain correlations (e.g., between A and E, or A and C) are fixed to zero by assumptions (not by any substantive theory). Any violation of such assumptions will bias estimates in the model (e.g., Purcell, 2002). Note also that absence of any interaction among the latent variables is assumed[3]. Expressing the ACE model for the mean-centered observations in matrix notation, we have:

$$\mathbf{P}_i = \mathbf{\Lambda}\eta_i,$$

where i represents twin pair, $\mathbf{P}_i^t = [P_{i1}\ P_{i2}]$,

$$\mathbf{\Lambda} = \begin{bmatrix} a & c & e & 0 & 0 & 0 \\ 0 & 0 & 0 & a & c & e \end{bmatrix},$$

and $\eta_i^t = [A_1\ C_1\ E_1\ A_2\ C_2\ E_2]_i$. The expected covariance matrix is $\mathbf{\Sigma} = E[P_i\ P_i^t] = E[\mathbf{\Lambda}\eta_i\eta_i^t\mathbf{\Lambda}^t] = \mathbf{\Lambda}E[\eta_i\eta_i^t]\mathbf{\Lambda}^t = \mathbf{\Lambda\Psi\Lambda}^t$, where the correlation matrix of the latent variables is denoted $\mathbf{\Psi}$. The correlation matrix $\mathbf{\Psi}$ contains the expected correlations among $\eta_i^t = [A_1\ C_1\ E_1\ A_2\ C_2\ E_2]_i$:

[3] Modeling genotype-environment correlation and genotype-environment interaction will be discussed in subsequent sections.

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 & 0 & \varrho_k & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \varrho_k & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where ϱ_k is the correlation between the twins' additive polygenic factors, i.e., unity in MZ twins and .5 in DZ twins (the k subscript denotes zygosity). As $\boldsymbol{\Psi}$ differs over zygosity, we require a separate model for MZ and DZ twins, i.e., $\boldsymbol{\Sigma}_{MZ} = \boldsymbol{\Lambda}\boldsymbol{\Psi}_{MZ}\boldsymbol{\Lambda}^t$ and $\boldsymbol{\Sigma}_{DZ} = \boldsymbol{\Lambda}\boldsymbol{\Psi}_{DZ}\boldsymbol{\Lambda}^t$. The actual structures of the 2x2 phenotypic covariance matrices are:

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma^2_{k11} & \sigma^2_{k12} \\ \sigma^2_{k21} & \sigma^2_{k22} \end{bmatrix} = \begin{bmatrix} a^2+c^2+e^2 & \varrho_{Ak}a^2+c^2 \\ \varrho_{Ak}a^2+c^2 & a^2+c^2+e^2 \end{bmatrix},$$

where k denotes zygosity, and the additive polygenic correlation ϱ_{Ak} is 1 in MZ and .5 in DZ twins. The standardized decomposition of variance is a^2/σ^2, c^2/σ^2, and e^2/σ^2, where σ^2 equals the total phenotypic variance (note $\sigma^2 = \sigma^2_{11} = \sigma^2_{22}$). The component is a^2/σ^2 is commonly denoted the narrow-sense heritability. In the ADE model $(a^2+d^2)/\sigma^2$, the proportion of total genetic effects, is called the broad-sense heritability.

Application of the univariate twin model has provided important insights into the structure of individual differences in a variety of psychological phenotypes such as personality, cognitive abilities, and psychopathology. For instance, it is now clear that C plays a minor role in determining individual differences on personality dimensions. Furthermore, the role of C in general intelligence is considerable in young children, but with increasing age, the role of C wanes, while that of A waxes. By young adulthood, the heritability of general intelligence is as high as .7, while shared environmental influences are no longer discernible (e.g., Bartels, Rietveld, Van Baal, & Boomsma, 2002; Boomsma et al., 2002).

As demonstrated originally by Martin and Eaves (1977), a powerful feature of GCSM lies in the possibility to analyze multivariate phenotypes. Two examples of a multivariate ACE twin model are depicted in Figure 2. First consider the model on the right. While we have dropped the C factors from the model to avoid clutter in the figure, we include C in the following representation of the model:

$$\mathbf{P}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i,$$

where i represents twin pair, $\mathbf{P}_i^t = [P_{i11}\ P_{i21}\ P_{i31}\ P_{i12}\ P_{i22}\ P_{i32}]_i$,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_A & \boldsymbol{\Lambda}_C & \boldsymbol{\Lambda}_E & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{\Lambda}_A & \boldsymbol{\Lambda}_C & \boldsymbol{\Lambda}_E \end{bmatrix},$$

and $\eta_i^t = [A_{11}\ A_{21}\ A_{31}\ C_{11}\ C_{21}\ C_{31}\ E_{11}\ E_{21}\ E_{31}\ A_{12}\ A_{22}\ A_{32}\ C_{12}\ C_{22}\ C_{32}\ E_{12}\ E_{22}\ E_{32}]_i$. The 3x3 matrix $\boldsymbol{\Lambda}_A$ contains the regression coefficients in the regression of the phenotypes on the additive genetic factors ($\boldsymbol{\Lambda}_C$ and $\boldsymbol{\Lambda}_E$ are defined analogously):

$$\boldsymbol{\Lambda}_A = \begin{bmatrix} a_{11} & & \\ a_{21} & a_{22} & \\ a_{32} & a_{32} & a_{33} \end{bmatrix}.$$

The implied phenotypic covariance matrices can be expressed as $\mathbf{\Sigma}_{mz} = \mathbf{\Lambda\Psi}_{mz}\mathbf{\Lambda}^t$ and $\mathbf{\Sigma}_{dz} = \mathbf{\Lambda\Psi}_{dz}\mathbf{\Lambda}^t$, where

$$\mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{\Sigma}_{k11} & \mathbf{\Sigma}_{k12} \\ \mathbf{\Sigma}_{k21} & \mathbf{\Sigma}_{k22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t+\mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t+\mathbf{\Lambda}_E\mathbf{\Lambda}_E{}^t & \varrho_{Ak}\mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t+\mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t \\ \varrho_{Ak}\mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t+\mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t & \mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t+\mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t+\mathbf{\Lambda}_E\mathbf{\Lambda}_E{}^t \end{bmatrix},$$

and k, as above, denotes zygosity. Given p phenotypes, $\mathbf{\Sigma}_{k11}$ ($\mathbf{\Sigma}_{k22}$) is the expected pxp phenotypic covariance matrix of twin 1 (twin 2), and $\mathbf{\Sigma}_{k12}$ is the expected pxp twin 1 - twin 2 phenotypic cross-covariance matrix.

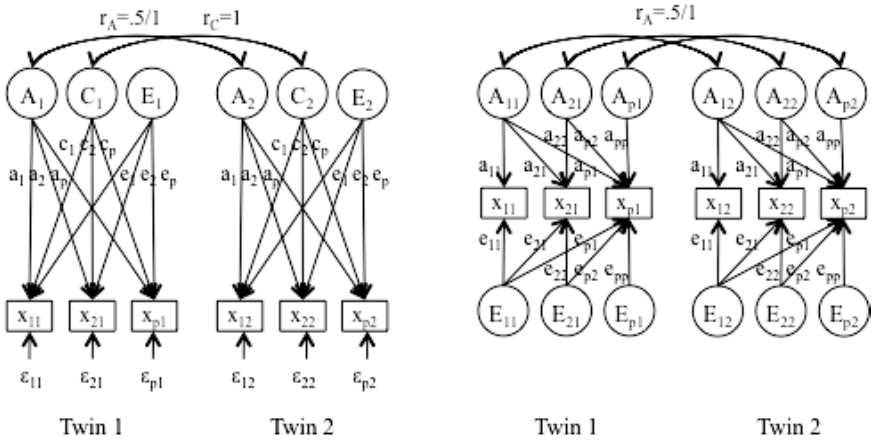

Figure 2. Multivariate genetic factor models with single (left) and multiple (right) genetic and environmental factors.

It is important to note that in the right panel of Figure 2, the phenotypic covariance matrix $\mathbf{\Sigma}_{11}$ ($\mathbf{\Sigma}_{22}$) is decomposed into covariance matrices $\mathbf{\Sigma}_A=\mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t$, $\mathbf{\Sigma}_C=\mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t$, and $\mathbf{\Sigma}_E=\mathbf{\Lambda}_E\mathbf{\Lambda}_E{}^t$, where $\mathbf{\Lambda}_A$ (shown above), $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ are lower triangular matrices. This triangular decomposition has the advantage that the sum of the underlying covariance matrices ($\mathbf{\Sigma}_A+\mathbf{\Sigma}_C+\mathbf{\Sigma}_E$) yields a covariance matrix that is almost certainly positive definite (M. C. Neale & Cardon, 1992). Beyond this restriction, the underlying covariance matrices are not modeled. However, the covariance matrices $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ may themselves be subjected to covariance structure modeling. That is, we can specify any model for each of the covariance matrices underlying the phenotypic covariance matrix. For example, see the left panel of Figure 2, where we have introduced common A, C, and E factors, and residuals that represent the effects of error and phenotype-specific environment. The model for the unshared environmental effects is now a standard factor model (Lawley & Maxwell, 1971), i.e., $\mathbf{\Sigma}_E = \mathbf{\Lambda}_E\mathbf{\Psi}_E\mathbf{\Lambda}_E{}^t+\mathbf{\Theta}_E$. The path diagram is overly simple (e.g., genetic and shared environmental residuals may be added), but it illustrates the principle of modeling the genetic and environmental covariance matrices.

Examples of GCSM based on the twin design

By multivariate GCSM, we obtain the decomposition of covariances among the phenotypes, and thus insight into the cause of phenotypic covariation. For instance, phenotypic measures of depression and anxiety covary quite considerably (Angold, Costello, & Erkanli, 1999; Brady & Kendall, 1992). Multivariate GCSM has been used to estimate the

contributions of genetic and environmental factors to the phenotypic covariance (e.g., Hettema, Prescott, & Kendler, 2004; Kendler, Heath, Martin, & Eaves, 1987). Interestingly, from these analyses it appears that the distinction between anxiety and depression is not a function of genetic differences, since the additive genetic factors that underlie anxiety and depression are hardly separable; rather, the distinction between these disorders appears to be driven by the unique environmental covariance structure. GCSM has also been used to study the genetic and environmental contributions to the intercorrelations among cognitive ability tests (e.g., subtests of the WAIS or the WISC). For instance, the phenotypic covariance structure of the WAIS can be represented by a hierarchical factor model with three or four first-order factors, and a second-order general factor. Rijsdijk, Vernon, & Boomsma (Rijsdijk, Vernon, & Boomsma, 2002) found that the underlying additive genetic influences resembled the hierarchical phenotypic structure, while the structure of the underlying unshared environmental influences resembled a single factor model.

Both growth curve and simplex models have been applied in twin studies to study the roles of genetic and environmental factors in development. Applied phenotypically, growth curve models are used to study individual differences in growth curves by regressing repeated measures on the (appropriately coded) time index. Often, a polynomial regression model is used, which may include higher order components to accommodate non-linearity (see M. C. Neale & McArdle, 2000, for other nonlinear models). A simple linear model may be conveyed as $X_{it} = I_i + t*S_i + \varepsilon_{it}$, where X_{it} is the phenotypic measure of subject i at occasion t (t=0,1,2,...), I is the random intercept and S is the random slope (Figure 3). The phenotypic mean at occasion t is $E(X_t)=E(I)+ t*E(S)$. In a growth curve model, regression coefficients (S and I), are random over subjects, which allows for individual differences in the form of the growth curve. The covariance matrix of interest is therefore:

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma^2_I & \sigma_{IS} \\ \sigma_{IS} & \sigma^2_S \end{bmatrix}.$$

Using GCSM, this covariance matrix can be decomposed into genetic and environmental components (e.g., $\boldsymbol{\Psi} = \boldsymbol{\Psi}_A + \boldsymbol{\Psi}_C + \boldsymbol{\Psi}_E$), which provides a window on the role of genetic and environmental factors in growth or decline (e.g., McArdle, 1986). A notable area of application of growth curve modeling is that of age-related changes in cognitive abilities (e.g., McArdle, Prescott, Hamagami, & Horn, 1998; Reynolds, Finkel, Gatz, & Pedersen, 2002; Reynolds et al., 2005), especially with regard to cognitive decline. Multiple studies of aging have demonstrated, for instance, that additive genetic factors account for most of the variance in intercept (or level) in adults of age 50 or more, whereas the rate of change (decline) is primarily affected by unshared environmental factors (e.g., Reynolds et al., 2002). The model has also been applied in other research areas, such as personality, psychopathology (e.g., Burt, McGue, Carter, & Iacono, 2007; Kendler et al., 2007) and health research (e.g., Hjelmborg et al., 2008).

An alternative approach to the analysis of repeated measures is provided by the simplex model. The simplex model is used to assess stability over time, by regressing the data at occasion t (t=1,…,T) on data at the preceding occasion (t-1) (Boomsma & Molenaar, 1987; Eaves, Long, & Heath, 1986; Hewitt, Eaves, Neale, & Meyer, 1988). The simplex model is depicted in Figure 3. To ease presentation we limit the model to additive genetic (A) and unshared environmental influences (E). In this model, the phenotypic variable X measured at time point t, X_t, is related to the additive genetic and unshared environmental factors A_t and E_t (t=1,…,T). Simplex models, or first order autoregressions, are specified to account for the stability and change at the level of the A_t and E_t. For instance, for the unshared environmental part, the autoregression is $E_t=\beta_{Et,t-1}*E_{t-1}+\zeta_{Et}$, and the implied decomposition of

variance is $\beta_{Et,t-1}{}^2\sigma^2_{Et-1} + \sigma^2_{\zeta Et}$. The simplex model has been applied extensively in GCSM. For a study of repeatedly measured full scale IQ at age 5, 7, 10, and 12, see Bartels, et al. (Bartels et al., 2002). Hoekstra, Bartels, and Boomsma (R. A. Hoekstra, Bartels, & Boomsma, 2007) applied the model to repeatedly measured verbal and nonverbal IQ tests administered at 5 occasions from ages 5 to 18 (see also Bishop et al., 2003; Eaves et al., 1986; Petrill, Lipton, Hewitt, & Plomin, 2004; Rietveld, Dolan, Van Baal, & Boomsma, 2003). Generally, these studies found that the observed temporal stability in cognitive performance was due to a single common genetic factor, and a common shared environmental factor. The latter declined in effect over the years, such that it was all but absent in early adulthood. In addition, age-specific additive genetic factors emerged at different ages (i.e., $\sigma^2_{\zeta At} \neq 0$), partly accounting for the lack of complete temporal stability. The genetic simplex model has also been applied in other domains, such as personality (e.g., Gillespie, Evans, Wright, & Martin, 2004; Pedersen & Reynolds, 1998) and psychopathology (e.g., Boomsma, Van Beijsterveldt, & Hudziak, 2005a; Gillespie, Kirk, et al., 2004).

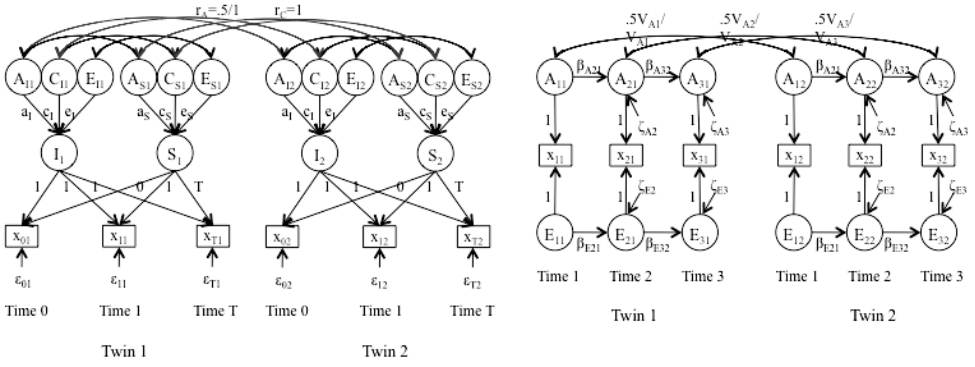

Figure 3. A linear growth curve (left) and a simplex AE (right) genetic covariance structure model.

Structural equation models developed within genetics

The examples of GCSM discussed above essentially involve the simultaneous estimation and modeling of the covariance matrices $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ (or $\mathbf{\Sigma}_D$), and $\mathbf{\Sigma}_E$. The fact that these matrices may be subjected to any identified SEM resulted in the full scale adoption of SEM in GCSM. However, the twin design itself and its various extensions (e.g., the use of parental ratings of the twins) posed modeling challenges and provided unique modeling possibilities. We now discuss several models that were developed in GCSM of twin data. These models include 1) the common and independent pathway factor models, 2) moderation models, 3) sex interaction models, and 4) direction of causality models.

Factor models: common pathway and independent pathway models

With regard to the relationship between the genetic and environmental factors, on the one hand, and the observed phenotypes, on the other, two kinds of factor models may be distinguished: the common pathway model and the independent pathway model (Kendler et al., 1987; McArdle & Goldsmith, 1990). The common pathway model is depicted in the left

panel of Figure 4. In this model, the influences of A, C (or D), and E on the phenotypes are mediated by a latent phenotype, represented by the common factors P_1 and P_2 in Figure 4. In this model, the factors P_1 and P_2 generally have substantive interpretations (e.g., neuroticism or verbal intelligence). The latent phenotypes mediate the genetic and environmental effects, as the path from the A, C, and E factors to the observed phenotypes runs via the latent phenotype. In the common pathway model, the observed variables may be interpreted as indicators of the latent phenotype (Mellenbergh, 1994).

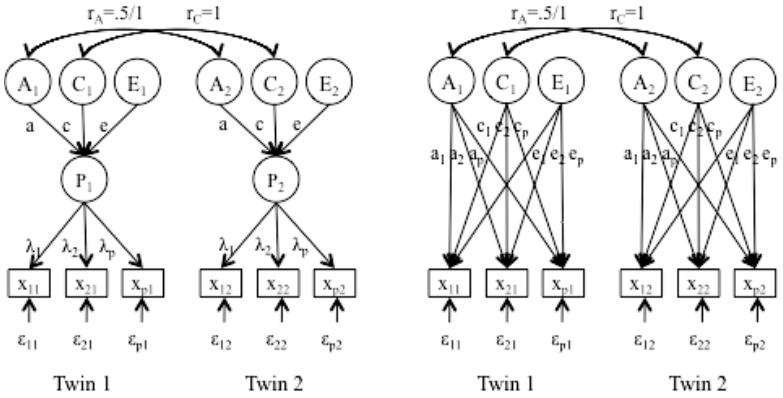

Figure 4. A common pathway (left) and an independent pathway (right) genetic factor model.

In the independent pathway model (Kendler et al., 1987), or the biometric factors model (McArdle & Goldsmith, 1990), the common factors A, C, and E influence the phenotypes directly; these is no mediating phenotypic common factor. A simple instance of this model is shown in the right panel of Figure 4. We can convey the common pathway model as

$$\boldsymbol{\Sigma}_{k11} = \boldsymbol{\Sigma}_{k22} = \boldsymbol{\Lambda\Psi\Lambda}^t + \boldsymbol{\Theta}_{cp} = \boldsymbol{\Lambda}(\boldsymbol{\Psi}_A + \boldsymbol{\Psi}_C + \boldsymbol{\Psi}_E)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp}$$
$$\boldsymbol{\Sigma}_{k11} = \boldsymbol{\Sigma}_{k22} = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}_A\boldsymbol{\Phi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Phi}_C\boldsymbol{\Gamma}_C{}^t + \boldsymbol{\Gamma}_E\boldsymbol{\Phi}_E\boldsymbol{\Gamma}_E{}^t)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp}$$
$$\boldsymbol{\Sigma}_{k21} = \boldsymbol{\Lambda}(\varrho_{Ak}\boldsymbol{\Psi}_A + \boldsymbol{\Psi}_C)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp21} = \boldsymbol{\Lambda}(\varrho_{Ak}\boldsymbol{\Gamma}_A\boldsymbol{\Phi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Phi}_C\boldsymbol{\Gamma}_C{}^t)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp21},$$

and the independent pathway model as

$$\boldsymbol{\Sigma}_{k11} = \boldsymbol{\Sigma}_{22} = \boldsymbol{\Lambda}_A\boldsymbol{\Phi}_A\boldsymbol{\Lambda}_A{}^t + \boldsymbol{\Lambda}_C\boldsymbol{\Phi}_C\boldsymbol{\Lambda}_C{}^t + \boldsymbol{\Lambda}_E\boldsymbol{\Phi}_E\boldsymbol{\Lambda}_E{}^t + \boldsymbol{\Theta}_{ip}$$
$$\boldsymbol{\Sigma}_{k21} = \varrho_{Ak}\boldsymbol{\Lambda}_A\boldsymbol{\Phi}_A\boldsymbol{\Lambda}_A{}^t + \boldsymbol{\Lambda}_C\boldsymbol{\Phi}_C\boldsymbol{\Lambda}_C{}^t + \boldsymbol{\Theta}_{ip21}.$$

Here, $\boldsymbol{\Phi}_A$, $\boldsymbol{\Phi}_C$, and $\boldsymbol{\Phi}_E$ are the covariance matrices of the A, C, and E factors, respectively. In the common pathway model, the covariance matrix of the psychometric factor, $\boldsymbol{\Psi}$, equals $\boldsymbol{\Psi}_A + \boldsymbol{\Psi}_C + \boldsymbol{\Psi}_E$, i.e., $\boldsymbol{\Gamma}_A\boldsymbol{\Phi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Phi}_C\boldsymbol{\Gamma}_C{}^t + \boldsymbol{\Gamma}_E\boldsymbol{\Phi}_E\boldsymbol{\Gamma}_E{}^t$, where $\boldsymbol{\Gamma}_A$, $\boldsymbol{\Gamma}_C$, and $\boldsymbol{\Gamma}_E$ are the vectors of factor loadings. The $\boldsymbol{\Lambda}$ (in the common pathway model) and $\boldsymbol{\Lambda}_A$, $\boldsymbol{\Lambda}_C$, and $\boldsymbol{\Lambda}_E$ (in the independent pathway model) vectors contain the factor loadings of the indicators on the psychometric factor, and on the biometric (A, C, and E) factors, respectively. Note that in both models the matrices $\boldsymbol{\Theta}$ (denoted $\boldsymbol{\Theta}_{cp}$ and $\boldsymbol{\Theta}_{ip}$, as they may vary over the models) contain the residuals of the indicators in the model. By considering two phenotypes (x_{j1}, j=1,2) in the common pathway model:

$$X_1 = \lambda_1 (aA + cC + eE) + \varepsilon_{11} = \lambda_1 aA_1 + \lambda_1 cC_1 + \lambda_1 eE_1 + \varepsilon_1,$$

$$X_2 = \lambda_2 (aA + cC + eE) + \varepsilon_{12} = \lambda_2 aA_1 + \lambda_2 cC_1 + \lambda_2 eE_1 + \varepsilon_2,$$

and the independent pathway model:

$$X_1 = a_1 A_1 + c_1 C_1 + e_1 E_1 + \varepsilon_1,$$
$$X_2 = a_2 A_1 + c_2 C_1 + e_2 E_1 + \varepsilon_2,$$

we note that the common pathway model is nested under the independent pathway model, i.e., that we may derive the common factor model from the independent pathway model by imposing appropriate proportionality constraints on the factor loadings. Specifically, the introduction of the constraints $a_1/a_2 = c_1/c_2 = e_1/e_2$ renders the common and the independent pathway equations above equivalent (see also Yung, Thissen, & McLeod, 1999). Hence, restrictions of the common pathway models can be tested using a likelihood ratio test. Such comparisons are particularly useful in addressing methodological issues pertaining to the conceptual status of latent variables (Franić, Dolan, Borsboom, Hudziak, et al., 2013). Specifically, if the independent pathway fits better than the corresponding common pathway model, we may conclude that the genetic and environmental influences on the indicators in the model are not fully mediated by the phenotypic latent variable (i.e., the psychometric factor). If the measured phenotypes are taken as indicators of the phenotypic latent variable, this calls into question the substantive meaning of the phenotypic latent variable. Ideally, if the common factor obtained in a phenotypic analysis represents a substantive unitary construct, and the phenotypes are indicators of this construct, one would expect the genetic and environmental influences on the indicators of the construct to be mediated by the construct.

The independent pathway model may be applied in a purely exploratory manner to determine the (possibly different) dimensionalities of the covariance matrices $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ (or $\mathbf{\Sigma}_D$), and $\mathbf{\Sigma}_E$. For example, Kendler, et al. (Kendler et al., 1987) concluded that the dimensionality of anxiety and depression symptoms differs with respect to genetic and environmental factors; while genetic factors appear to represent a unidimensional structure affecting the overall level of symptoms, environmental influences distinctly affect symptoms of anxiety and symptoms of depression, giving rise to the observed phenotypic clustering of the two disorders.

Genotype-environment interaction

The possibility of genotype by environment interaction (G x E) is widely recognized in human genetics and, if present, may have a biasing effect on estimates obtained in the standard twin model. We speak of G x E if an environmental variable moderates the genetic effects in the sense that the magnitude of the genetic variance varies over the levels of the moderator. Similarly, a genetic variable (a given genotype) may moderate environmental effects.

The phenotypic variance in the presence of G x E may be expressed as $\sigma^2_P = \sigma^2_G + \sigma^2_E + \sigma^2_{GxE}$, where σ^2_{GxE} represents variance due to the interaction. In the twin model, the effect of the interaction depends on the exact nature of the interaction (Purcell, 2002). In the ACE twin model, the variance due to A x C interaction cannot be distinguished from the A variance. Thus A x C interaction will result in overestimation of the A variance. On the other hand, variance due to A x E interaction cannot be distinguished from E variance.

Several methods have been proposed to detect interaction in the twin model. Jinks and Fulker proposed the regression of MZ twin pair differences on MZ twin pair sums

(Eaves, 1984; Jinks & Fulker, 1970). In this method, the MZ pair difference features as a measure of environmental variability, and the MZ pair means as measure of the polygenic effects. In the absence of G x E, the environment variability should not depend on the genotypic level, i.e., the regression coefficient should be zero (see van der Sluis, Dolan, Neale, Boomsma, & Posthuma, 2006, for a related method).

Modeling G x E is relatively easier if one has measured the variable which moderates the genetic effects. Given a measured moderator, G x E can be modeled by fitting the twin model conditional on the moderator (e.g., in a multigroup model, with groups corresponding to the levels of the moderator). One can then test for homogeneity in A, C, and E variance components over the levels the moderator. For instance, Boomsma, de Geus, van Baal, and Koopmans (Boomsma, de Geus, van Baal, & Koopmans, 1999) found that the heritability of disinhibition (a personality trait related to sensation seeking), as estimated in the twin design, depended on whether the twins had a religious upbringing or not. In the latter case the heritability was about .45 (typical for personality traits; Boomsma et al., 2002), but in the former, it was less than .10.

Purcell (Purcell, 2002) proposed a general method to accommodate a measured moderator in the twin model, where the moderator can be any variable (not necessarily environmental; Kendler & Baker, 2007; Plomin et al., 2008; Vinkhuyzen, van der Sluis, de Geus, Boomsma, & Posthuma, 2010; Vinkhuyzen, van der Sluis, & Posthuma, 2010). This method can also accommodate genetic and environmental effects on the moderator itself, and the possible correlation between the moderator and the trait. For instance, parenting style may moderate the heritability of neuroticism in children, but it is quite possible that the parenting style of the parents and the neuroticism of the children are correlated, either directly (say, common genetic influences), or indirectly (say, a highly neurotic child elicits a given parenting style).

Purcell's approach to modeling G x E is depicted in Figure 5. We limit the depiction to an AE model to ease presentation. In this model, M1 (M2) is the moderator measured in twin 1 (twin 2), and T1 (T2) is the phenotype of interest measured in twin 1 (twin 2). The models for the moderator M_i and the trait T_i are

$$M_i = e_m{}^*E_{ci} + a_mA_{ci}$$
$$T_i = (e_c + \beta_{ec}{}^*M_i)^*E_{ci} + (a_c + \beta_{ac}{}^*M_i)^*A_{ci} + (e_u + \beta_{eu}{}^*M_i)^*E_{ui} + (a_u + \beta_{au}{}^*M_i)^*A_{ui}.$$

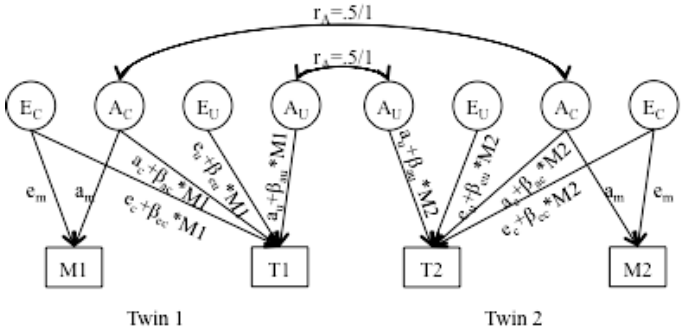

Figure 5. A G x E model (Purcell, 2002). The model accommodates moderation of the possible covariation between a measured moderator (M) and a phenotype (T), and the moderation of the residual genetic and environmental effects on T. In addition, the model includes the decomposition of the phenotypic variance of the moderator itself.

This model accommodates moderation of the covariance between the moderator and the trait [$(e_c+\beta_{ec}*M_i)*E_{ci}$ + $(a_c+\beta_{ac}*M_i)*A_{ci}$], and moderation of the residual [$(e_u + \beta_{eu}*M_i)*E_{ui}$ + $(a_u + \beta_{au}*M_i)*A_{ui}$], and includes the decomposition of the phenotypic variance of the moderator itself. Tests of moderation can be carried out by means of a likelihood ratio test.

This bivariate moderation model describes the relations between T and M in such detail that computational problems (e.g., sensitivity to starting values, converging problems) may arise, especially if the covariance between trait T and moderator M is small. In addition, Rathouz et al. (Rathouz, Van Hulle, Rodgers, Waldman, & Lahey, 2008) have shown that this model sometimes produces spurious moderation effects. An alternative approach is to regress the trait directly on the moderators, without decomposing the variance of the moderator. The moderation of the regression of the trait on its genetic and environmental factors is retained (van der Sluis, Posthuma, & Dolan, 2012).

The popularity of the G x E model is evident given its frequent use in twin studies on moderation in the context of, for instance, cognitive abilities (e.g., Bartels, van Beijsterveldt, & Boomsma, 2009), personality (e.g., Brendgen et al., 2009), health (e.g., Johnson & Krueger, 2005), or brain morphology (e.g., Lenroot et al., 2009). This method of handling moderation, i.e., modeling moderation directly on the path parameters of the model, is also used in SEM outside the field of GCSM. Bauer and Hussong (Bauer & Hussong, 2009) applied it to test whether the parameters in the one factor model depend on a continuous moderator, or in the present case, a differentiation variable. Molenaar et al. (Molenaar, Dolan, Wicherts, & van der Maas, 2010) and Tucker-Drob (Tucker-Drob, 2009) used this method to investigate ability differentiation (Spearman, 1927) in the higher-order common factor model.

Although G x E is frequently discussed in conjunction with genotype-environment correlation (rGE), G x E and rGE represent very different mechanisms. rGE refers to a non-random distribution of genotypes over the environments. rGE may arise, for instance, from genetic control of exposure to environmental events (Kendler & Eaves, 1986). Examples of research on rGE include, e.g., a study by Kendler & Karkowski-Shuman (Kendler & Karkowski-Shuman, 1997), in which rGE was shown to explain the association between life events and depression. However, not all studies supported this finding (e.g., Middeldorp, Cath, Beem, Willemsen, & Boomsma, 2008).

Sex interaction in the twin model

An important possible moderator of genetic and environmental effects in the twin model is sex. The classical twin design can be broken down by sex, i.e., we can distinguish between same-sex pairs (MZ males, DZ males, MZ females, DZ females) and opposite-sex pairs (DZOS). This extended design, specifically the presence of DZ opposite-sex twins, provides the information to study both qualitative and quantitative sex differences in genetic and environmental effects. Figure 6 depicts a partial path diagram of a general sex limitation model (Eaves et al., 1978; M. C. Neale & Cardon, 1992), conveying both quantitative and qualitative sex differences. In the former case, the genetic factors are the same, but sex modulates their effects. In the latter case, different genetic factors (different genes) are expressed in men and women. To model quantitative effects, the genetic and environmental correlations (ϱ_A and ϱ_C) in opposite-sex DZ twin pairs are constrained to equal those in same-sex DZ twin pairs (.5 and 1, respectively), while the genetic and environmental factor loadings (a, c, and e) may differ across the sexes. This covariance structure implies that the A_f (C_f) and A_m (C_m) factors represent sets of genes (environmental influences) common to both sexes, but not necessarily of the same magnitude of effect in males and females. In addition, a sex-specific additive genetic factor (A'_m), uncorrelated with other additive

genetic factors in the model, is specified. This factor represents genetic effects unique to the phenotype of one, in this example male, sex. Note that we may also choose to model a sex-specific C factor, but cannot model both A'_m and C'_m. The model is fitted in a multi-group analysis, in which the parameters pertaining to men are equated across male groups (e.g. MZM, DZM and males from opposite-sex pairs) and the same is done for parameters pertaining to women. As a result, the expectations of variances are equal within, but not necessarily between, the sexes.

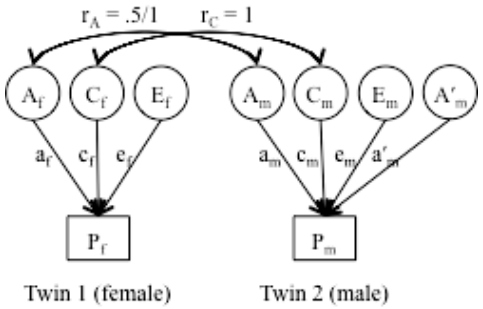

Figure 6. A general sex-limitation ACE model.

Testing for the presence of quantitative and qualitative sex differences may be performed by likelihood ratio tests based on the loglikelihood of the general sex-limitation model, and that of its various subset models.

The sex interaction model has been used extensively in various domains of genetics research, such as psychopathology (e.g., Boomsma, van Beijsterveldt, & Hudziak, 2005b; Eley & Stevenson, 1999; Rice, Harold, & Thapar, 2002a), intelligence (e.g., Bartels et al., 2002), personality (e.g., Eaves, Heath, Neale, Hewitt, & Martin, 1998; Rettew et al., 2006), health and well-being (e.g., Mosing, Pedersen, Martin, & Wright, 2010; Roysamb, Harris, Magnus, Vitterso, & Tambs, 2002; Schousboe et al., 2003), physiological traits (e.g., Weiss, Pan, Abney, & Ober, 2006), and substance abuse research (e.g., Prescott, Aggen, & Kendler, 1999, 2000). These studies generally indicate absence of any substantial sex-related differences. However, there are exceptions. For instance, Eaves, et al. (Eaves et al., 1998) found that the relative contribution of non-additive genetic effects to neuroticism is larger in males. Rettew, et al. (Rettew et al., 2006) showed that different genes may produce variation in neuroticism in male and female adolescents.

Direction of causation model

So far we have considered models in which the measured phenotypes are dependent variables and the genetic and environmental variables are independent variables. Twin models in which the measured phenotypes may be related directly, i.e., models in which the phenotypes are not strictly dependent variables, have also been developed. These models include longitudinal models (Eaves et al., 1986), the sibling interaction model (G. Carey, 1986), and the direction of causality (DOC) model (Heath et al., 1993). The DOC model is interesting from a SEM point of view as this twin model allows one to test hypotheses concerning the direction of causality among two (or more) phenotypes. For instance, a correlation between psychopathology and recall of early childhood environment may be due to a causal influence of the childhood environment on psychopathology (A → B), or, for

instance, to a biasing effect of current psychopathology on recall of childhood environment (B → A) (Heath et al., 1993). An instance of a bivariate genetic model with a reciprocal causal relationship between two indicator variables is depicted in Figure 7. Note that in this example, the model for trait x is an ADE model, while the model for trait y is a CE model.

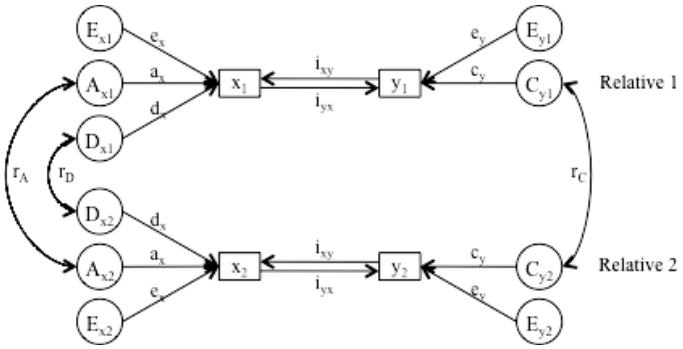

Figure 7. A bivariate genetic covariance structure model with a reciprocal causal relationship between the two indicator variables (Heath et al., 1993).

The expectations for the cross-relative cross-trait covariance (CRCTC, i.e., the covariance between trait x (y) in relative 1 and trait y (x) in relative 2) derived under this model may be employed to test hypotheses about the direction of causation between the two indicator variables. Specifically, consider the case in which trait x exerts a causal influence on trait y (x → y). Given that the variances of the latent factors are scaled at 1, the expected covariance structure is:

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22} = \begin{bmatrix} a_x^2 + d_x^2 + e_x^2 & i_{yx}(a_x^2 + d_x^2 + e_x^2) \\ i_{yx}(a_x^2 + d_x^2 + e_x^2) & c_y^2 + e_y^2 + i_{yx}^2(a_x^2 + d_x^2 + e_x^2) \end{bmatrix},$$

$$\boldsymbol{\Sigma}_{k21} = \boldsymbol{\Sigma}_{k12}{}^t = \begin{bmatrix} r_A a_x^2 + r_D d_x^2 & i_{yx}(r_A a_x^2 + r_D d_x^2) \\ i_{yx}(r_A a_x^2 + r_D d_x^2) & r_C c_y^2 + i_{yx}^2(r_A a_x^2 + r_D d_x^2) \end{bmatrix},$$

where $i_{yx}(r_A a_x^2 + r_D d_x^2)$ is the expectation for the CRCTC. Conversely, if y → x, the expected CRCTC can be shown to be $i_{xy} r_C c_y^2$. Given that the CRCTC depends on r_A and r_D if x → y, and on r_C if y → x, a comparison of CRCTCs in groups of different degrees of genetic and environmental relatedness is informative about the direction of causality. For instance, if x → y, CRCTC is positive in biological relatives, but its magnitude will depend on the degree of genetic relatedness. Alternatively, if y → x, CRCTC will be positive and independent of the degree of genetic relatedness in individuals reared in the same family, and zero in individuals reared in separate families. Family data will, however, only be informative about direction of causality if the phenotypes have different modes of inheritance, i.e., if the effects of A, C (D), and E differ across the two phenotypes.

In the DOC model depicted in Figure 7, the latent genetic and environmental factors affecting each of the two traits are uncorrelated; thus the only mechanism that generates the correlation between traits x and y is the unidirectional causal effect of x on y, or of y on x (or bidirectional causal effects, which may be resolved using models with multiple indicators; Heath, et al., 1993). The standard bivariate twin model, in contrast, models the phenotypic correlation between x and y as a function of the underlying genetic and environmental

correlations. Given that the DOC model is nested under the general bivariate model (Heath et al., 1993), the fit of both uni- and bidirectional causal models can be compared to that of the general bivariate model by means of a likelihood ratio test.

Thomsen, et al. (Thomsen et al., 2009) applied the this model to data on asthma and severe respiratory syncytial virus (RSV) infection, and found that the high positive association between these phenotypes is not due to RSV causing asthma, but to both phenotypes reflecting a common genetic predisposition. In the area of intelligence research, Luciano, et al. (Luciano et al., 2005) showed that the well-established correlation between inspection time (a measure of perceptual discrimination speed; IT) and general cognitive ability is due neither to the efficiency of IT increasing general cognitive ability, nor to general ability affecting IT. Instead, both processes seem to be indicators of common genetic factors. De Moor et al. (de Moor, Boomsma, Stubbe, Willemsen, & de Geus, 2008) used bivariate genetic modeling, analyses of longitudinal data and intra-pair differences in identical twins to show that the association between exercise and symptoms of anxiety and depression is not due to causal effects of (lack of) exercise.

In all models considered above the genetic (or polygenic) factors featured as latent variables, which represent the action of many polymorphic genes. In the final section of this chapter, we briefly discuss the incorporation of measured genes in genetic association analysis.

From genetic latent variables to measured genetic markers: association analysis

Developments in high-throughput genotyping technologies have enabled geneticists to measure vast amounts of genetic information directly (Slagboom & Meulenbelt, 2002). Consequently, twin and family registries now include measured genotypic material in the form of genetic markers, alongside phenotypic data. Accordingly, the aim of studies has shifted towards localizing and identifying individual genes that contribute to variation in complex phenotypes (Cherny, 2008; Guo & Adkins, 2008; Vink & Boomsma, 2002). Such genes are generally expected to have a relatively small effect (<1% of the phenotypic variance), and are referred to as quantitative trait loci (QTLs). In this section we outline how information on genetic markers has been incorporated in SEM.

A genetic marker is a DNA sequence with a known chromosomal position. This sequence is measurable and is polymorphic, i.e., displays inter-individual variation (i.e., there are at least 2 alleles). The marker may be (part of) a functional gene of interest (i.e., a candidate gene), but that is not necessarily the case. Markers are used in two ways to identify the QTLs that contribute to individual differences in the trait of interest: linkage and association analysis. Both methods hinge on the phenomenon of *co-segregation*, i.e., the fact DNA variants located closely together on the same chromosome are not inherited independently. This means that a marker can serve as an indicator of functional genes in its chromosomal vicinity. Linkage analysis is carried out in pedigrees, in which long stretches of DNA are shared among family members. Genetic association tests are usually performed in samples of unrelated subjects, either patients (cases) and controls, or subjects who differ phenotypically on a quantitative scale. Although both linkage and association analysis have been incorporated in SEM (e.g., Hottenga & Boomsma, 2008), here we discuss association analysis, as this technique currently dominates gene hunting enterprises and is characterized by a higher statistical power to identify genes of relatively small effect (see e.g. T.A. Manolio & Collins, 2009).

Distant loci (or loci on different chromosomes) are subject to recombination as a consequence of chromosomal crossing over during meiosis (the formation of gametes).

However, when large numbers of markers are genotyped (e.g. typically 0.5 to 1.5 million variants), the chromosomal locations of a QTL and a marker may be so close that configurations of the alleles (i.e., haplotypes) are almost always transmitted from parents to offspring as single units. This means that at the population level a marker allele (say M1 of a marker locus with alleles M1 and M2) almost always forms a haplotype (M1-Q1) with a given allele of at a QTL location (say Q1 at a QTL locus with alleles Q1 and Q2). The observed marker allele can therefore serve as indicator of the QTL allele. The closer the marker is to the QTL on the chromosome (the more tightly they are linked), the more reliable it is as an indicator, as the process of crossing over is less likely. High resolution microarrays typically assess genetic variants called SNPs (single nucleotide polymorphisms) that have two alleles and whose minor alleles frequencies (MAF) are not extremely low. The reliability of a SNP as an indicator can be expressed in terms of the degree of linkage disequilibrium, i.e., a measure of the extent to which a SNP allele at locus M is predictive of the presence of an allele at locus Q in the population (Wray & Visscher, 2008).

In the case of a continuous phenotype, association analysis involves the regression of the phenotype on the number of M1 (or M2) alleles observed in each subject. If explained variance is statistically significant, the marker itself (say, if it is situated within a candidate gene), or a QTL in the vicinity of the SNP is associated with the phenotype. Note that this analysis can be carried out in samples of unrelated individuals, if the sample is genotyped across the genome. Given multivariate phenotypes, one can consider a multivariate test. Ferreira and Purcell (Ferreira & Purcell, 2009) considered the power of MANOVA given varying number of phenotypes (5, 10, and 20), of which a varying number were affected by the QTL. They found that the multivariate test was more powerful than univariate tests, with 1) increasing correlations among the phenotypes and 2) increasing number of phenotypes affected (i.e., by the QTL) increasing the power. They noted a sharp loss of power of the multivariate test when all phenotypes were affected by the QTL.

Given multivariate data, one can also consider embedding the test of association in a proper SEM. Medland and Neale (Medland & Neale, 2010) considered single factor models with 3 or 5 indicators, in unrelated cases and in sib pairs. They varied the locus of the effect of the QTL in the factor model such that it was part of the common factor, thus conveying its effect via the factor loadings on all variables, common to all phenotypes, but not conveyed via the factor, or common to some phenotypes, but not conveyed via the factor. The main conclusion is that their combined multivariate approach (where the QTL effect is conveyed via the common factor, or the QTL affects the phenotypes directly) was almost universally as powerful as, or depending on specific circumstances more powerful than, the univariate tests using weighted or unweighted sumscores. Van der Sluis et al. (van der Sluis, Verhage, Posthuma, & Dolan, 2010) discussed the power to detect the effects of genetic variants in the context of uni- and multidimensional common factor models, and contrasted the power in these designs to the sum score operationalisation in the case that the sum score is not a sufficient statistic (i.e., the sum score entails a loss of information). They showed that the sum score is as powerful as the factor analytic design only under very specific circumstances. In addition, they discussed how violations of measurement invariance across multiple samples or with respect to the genetic variant itself, affect the power to detect genetic variants. Violations with respect to the genetic variant itself proved very disadvantageous in both the sum score model as well as in incorrectly specified factor models. Van der Sluis et al. considered association tests in factor models fitted to ordinary samples (not genetically informative). Medland and Neale also considered sib data (e.g. DZ twins), as these allow one to test for stratification (i.e., spurious association due to population heterogeneity; Fulker, Cherny, Sham, & Hewitt, 1999). Minica, et al. (Minică,

Boomsma, Van Der Sluis, & Dolan, 2010) considered the incorporation of the marker information in a variety of structural equation models, including the simplex model and the multiple factor model. They considered both twin samples and ordinary samples. Overall, their results were consistent with Medland and Neale (Medland & Neale, 2010), and van der Sluis et al. (van der Sluis et al., 2010).

Other statistical issues in genetic association studies include the multiple testing problem (when e.g. more than a million SNPs are tested), meta-analyses to combine association tests, and the fact that a statistically significant association is not yet proof of a causal relation between the gene and the phenotype. In addition, rare genetic variants are difficult to hunt down in association tests, and attention has consequently shifted to large sequencing projects.

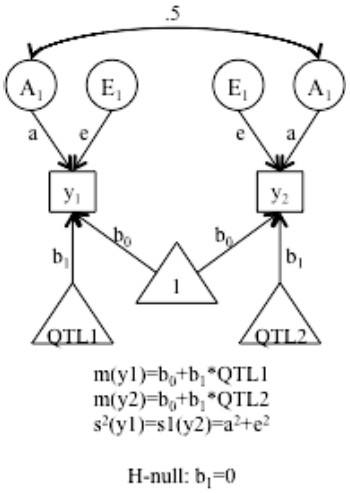

Figure 8. A genetic association model.

Discussion

The aim of the present chapter was to discuss SEM as applied in human quantitative genetics. We have concentrated on the classical twin design, as this provides a good basis for understanding genetic covariance structure modeling. Given its various assumptions, the twin design applied to multivariate phenotypic data provides the information to estimate the genetic and environmental covariance matrices underlying the phenotypic covariance matrix. As we discussed, any structural equation model that can be fitted to the phenotypic covariance matrix can be fitted to the genetic and environmental covariance matrices. An important fact is that the genetic covariance structure does not necessarily resemble the environmental covariance structure. The phenotypic structure may resemble either, or indeed neither. For instance, Rijsdijk et al. (Rijsdijk et al., 2002) in their analysis of the WAIS showed that the hierarchical phenotypic structure resembled closely the additive genetic structure, whereas the unshared environmental structure was a single factor model. One model in which the relationships between the genetic, environmental, and phenotypic covariance structures are compared explicitly is the common pathway model (e.g., Franić, Dolan, Borsboom, Hudziak, et al., 2013). This model implies that all three have the same

structure (i.e., are isomorphic). This follows from the fact that the phenotypic latent variables mediate the effect of the genetic and environmental factors on the indicators. Interestingly, the isomorphism implied by the common pathway model appears to be the exception rather than the rule. Often, therefore, the phenotypic structure is a function of two qualitatively different underlying structures. The fact that the phenotypic structure is in effect a weighted average of distinct underlying genetic and environmental structures underlines the theoretical importance of GCSM.

In the sections that followed, we reviewed a selection of models developed within genetics and suited specifically to address research questions arising in this particular field (such as, for instance, the possible moderation of genetic and environmental effects on the phenotype, or the direction of causation between observed variables). The utility of these models is evident given their widespread use in genetics research. However, the classical twin design represents only the most basic design employed in the field, and many extensions to this design are currently in use. These more elaborate designs usually involve adding one or more relatives to the study, which greatly enhances the resolution to detect more subtle effects, and increases the assemblage of research questions that may be addressed (for an overview see e.g. Boomsma et al., 2002; Truett et al., 1994). For instance, the inclusion of parents of twins allows one to study the effects of social homogamy and cultural transmission (Eaves, Fulker, & Heath, 1989; Rao, Morton, & Yee, 1974) or differential gene expression as a function of age (see e.g. Snieder, van Doornen, & Boomsma, 1997). Including spouses of twins allows one to study assortative mating (Eaves, 1979; Heath & Eaves, 1985). The inclusion of siblings of twins makes it possible to examine social interactions and special twin effects, such as prenatal hormone transition or shared prenatal environment (see e.g. Stoel, De Geus, & Boomsma, 2006), and including offspring of MZ twins (who are genetically half-sibs but socially cousins) allows for the study of maternal effects and imprinting (see, e.g., Nance, Kramer, Corey, Winter, & Eaves, 1983). In addition, adding relatives to the twin model results in an increase in power to detect and distinguish between different sources of variation (see also Dolan, Boomsma, & Neale, 1999; Posthuma & Boomsma, 2000).

In the present chapter we focused on continuous outcome measures; however, the possibility to model discrete data using the liability-threshold model (Falconer & Mackay, 1996) has also been developed. The liability-threshold model assumes the discrete phenotype to be a manifestation of an underlying continuous liability distribution, with one or more thresholds imposing a discontinuity on the visible expression. Estimation of thresholds and polychoric correlations allows one to specify models with respect to the underlying liability distributions rather than the observed discrete indicators. This model bears a close resemblance to the ordinal factor model as commonly applied in SEM.

As mentioned, the present chapter addressed only a selection of models employed in GCSM. More recent developments include, for instance, genetics applications of mixture modeling (B. O. Muthén, Asparouhov, & Rebollo, 2006), or extensions of GCSM to include linear feedback or recursiveness between multivariate phenotypes (Gianola & Sorensen, 2004). Finally, we note that GCSM may be performed using any standard SEM software, e.g., Mx (M. C. Neale, 2000), the OpenMx package in R (Boker et al., 2010; R Core Team, 2013), MPlus (L. K. Muthén & Muthén, 2007), or LISREL (Jöreskog & Sörbom, 2006). Extensive libraries containing Mx specifications of most of the models discussed in this chapter are available at http://www.psy.vu.nl/mxbib/ (Posthuma & Boomsma, 2005) and http://www.vcu.edu/mx/examples.html (M. C. Neale, 2007).

Chapter 3

Can Genetics help Psychometrics? Improving Dimensionality Assessment through Genetic Factor Modeling

Abstract

In the present chapter, we discuss the role that quantitative genetic methodology may play in assessing and understanding the dimensionality of psychological (psychometric) instruments. Specifically, we study the relationship between the observed covariance structures, on the one hand, and the underlying genetic and environmental influences giving rise to such structures, on the other. We note that this relationship may be such that it hampers obtaining a clear estimate of dimensionality using standard tools for dimensionality assessment alone. One situation in which dimensionality assessment may be impeded is that in which genetic and environmental influences, of which the observed covariance structure is a function, differ from each other in structure and dimensionality. We demonstrate that in such situations settling dimensionality issues may be problematic using standard factor analytic techniques, and propose using quantitative genetic modeling to uncover the (possibly different) dimensionalities of the underlying genetic and environmental structures. We illustrate using simulations and an empirical example on childhood internalizing problems.

Appendices can be obtained at http://sanjafranic.com/dissertation.

Based on: Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2013) Can Genetics Help Psychometrics? Improving Dimensionality Assessment Through Genetic Factor Modeling. *Psychological methods, 18(3)*, 406-433.

It could be argued that all psychometric modeling starts and ends with the assessment of dimensionality, i.e., with the determination of the number of latent psychological attributes that are measured through a set of indicators (e.g., questionnaire items, subtest scores etc). Psychometrics starts with dimensionality assessment because *some* idea of how many attributes one intends to measure, however implicit, guides the test construction and item selection process, as well as the psychometric models one subsequently entertains as viable candidate models for the data. Ideally, it also ends with dimensionality assessment in that, when the fog clears and validity issues begin to be settled, a picture emerges of *which* psychological attributes are measured by the test items; clearly, this question cannot be answered without simultaneously resolving the dimensionality issue.

The importance of dimensionality assessment, however, extends beyond purely psychometric issues pertaining to test construction, as dimensionality assessment impacts the research questions that psychologists pose and, as a result, the answers they obtain. For instance, via identification of item clusters, dimensionality assessment steers the allocation of items to subscales. This not only determines which subtest scores are analyzed in empirical data analysis, but also significantly influences the interpretation of latent variables hypothesized in psychological research. This interpretation may in turn result in revisions of theory concerning the nature of the psychological construct under consideration. In this way, procedures aimed at determining dimensionality play a central role in psychology; not just in the development of psychological tests, but also in the revision of interpretations of psychological constructs, and thus in the development of psychological theory (Cronbach & Meehl, 1955; Gorsuch, 1983; Haig, 2005a, 2005b; Mulaik, 1987; Rummel, 1970).

The most widely used, and in this sense most important, way of investigating dimensionality is through the statistical method of Exploratory Factor Analysis (EFA) and related models (e.g., Principal Component Analysis; PCA; Lawley & Maxwell, 1971). The influence of this method pervades many different areas in psychology. For instance, EFA has played an important role in the development of the Five Factor Model of personality (Costa & McCrae, 1985; Goldberg, 1990), the theory of child psychopathology associated with the Child Behavior Checklist (CBCL; Achenbach, 1966, 1991), and the Cattell-Horn-Carroll model of the structure of cognitive abilities (Carroll, 2003; Cattell, 1941; Horn, 1965). Many other examples could be listed, as EFA is one of the most widely used statistical techniques in the psychological science (Fabrigar, Wegener, MacCallum, & Strahan, 1999). In the past decades, confirmatory methods - e.g., Item Response Theory (IRT) modeling and Confirmatory Factor Analysis (CFA) - have been added to the repertoire for dimensionality assessment, and a good deal of work has gone into the development of heuristics to facilitate the process (Fabrigar et al., 1999; Henson & Roberts, 2006; Zwick & Velicer, 1982, 1986).

Notwithstanding the availability of these statistical tools, the evaluation of dimensionality remains difficult. For instance, in the area of cognitive abilities research, there is currently a lack of consensus on whether the *g* factor (general intelligence) can be equated with some of the more specific common factors, such as working memory or fluid reasoning (e.g., Ackerman, Beier, & Boyle, 2005; Matzke, Dolan, & Molenaar, 2010). Given the lack of sufficiently elaborate theory, research relies heavily on the intercorrelations among common factors as a source of information concerning dimensionality, conditional on the specified factor structure. Similar issues arise in psychopathology research, where some of the most prominent debates concern the origin of covariation between symptoms of two or more disorders (e.g., Angold et al., 1999; Cramer, Waldorp, van der Maas, & Borsboom, 2010; Lilienfeld, Waldman, & Israel, 1994). For instance, the co-occurrence of symptoms of anxiety and depression is typically subject to many different explanations, ranging from those that view the disorders as different points on the same continuum, to

those conceptualizing them as empirically and conceptually distinct phenomena (Clark, 1989).

It would thus appear that EFA and related methods, which work purely on the observed covariation between the items[4], do not always have sufficient resolution to firmly clinch dimensionality issues. However, it is not entirely clear *why* dimensionality assessment is so difficult. In the light of work done in the field of quantitative genetics (e.g., Boomsma & Molenaar, 1986; Martin & Eaves, 1977), we propose that one of the possible reasons underlying this difficulty is that item covariation, upon which EFA and related methods work, may be the result of genetic and environmental influences which differ from each other in dimensionality and structure. In the current chapter, we study the relationship between the item covariance structures, on the one hand, and the underlying genetic and environmental covariance influences giving rise to such structures, on the other. This relationship, as we will show, may be such that it hampers obtaining a clear phenotypic dimensionality (i.e., dimensionality assessed on the basis of observed item covariation only). Incorporating genetic information in item analysis may yield a deeper understanding of the number of latent variables measured through the test scores. This provides important insights and research opportunities in the context of dimensionality assessment.

The structure of this chapter is as follows. We first introduce genetic factor modeling as applied in the classical twin design, and note that the genetic and environmental influences underlying the observed item covariation do not necessarily resemble each other in structure. This fact, in turn, may have implications for dimensionality assessment. We illustrate using 1) a simulation study, and 2) an empirical example on childhood internalizing psychopathology. Before addressing these issues, however, it is necessary to cover the basics of the genetic factor model as applied in the classical twin design.

Genetic covariance structure modeling and the twin design

Genetic covariance structure modeling (GCSM; Martin & Eaves, 1977) is the application of structural equation modeling (Bollen, 1989; Kline, 2005) to data collected in genetically informative samples, such as siblings or adoptees (Boomsma et al., 2002; Franić, Dolan, Borsboom, & Boomsma, 2012; M. C. Neale & Cardon, 1992). The fact that the samples are genetically informative (i.e., they consist of relatives whose average degree of genetic resemblance is known based on quantitative genetic theory; Falconer & Mackay, 1996) makes it possible to assess the relative contributions that genetic and environmental factors make to individual differences in observed traits (i.e., phenotypes). This is done by modeling genetic and environmental effects as contributions of latent variables to individual differences in observed traits, and estimating these contributions as regression coefficients in the linear regression of the observed traits on the latent genetic and environmental variables. The genetic and environmental latent variables themselves represent the effects of many unidentified influences: the genetic factors represent the effects of an unknown number of genes (polygenes), while the environmental factors correspond to effects of a potentially large number of unmeasured environmental influences. Measured genotypic and environmental information may also be included in the analyses (Cherny, 2008; Medland & Neale, 2010), but we do not consider this possibility in the present chapter.

Identification in GCSM is achieved by using the information on the average degree of genetic resemblance between relatives in specifying the model. For instance, in the classical twin design the sample consists of monozygotic (MZ) and dizygotic (DZ) twin pairs. DZ

[4] Or on the estimated covariation between latent distributions assumed to underlie discrete items.

twins share on average 50% of their segregating genes, while MZ twins share nearly their entire genome (Falconer & Mackay, 1996; J. P. van Dongen, Draisma, Martin, & Boomsma, 2012). The observed (i.e., phenotypic) covariance structure is typically modeled as a function of latent factors representing three sources of individual differences: additive genetic (A), shared environmental (C) and individual-specific environmental (E) sources[5]. Additive genetic influences are modeled by one or more A factors, which represent the total additive effects of genes relevant to the phenotypes. Based on quantitative genetic theory (Falconer & Mackay, 1996), the A factors are known to correlate 1 across MZ twins and .5 across DZ twins. Environmental influences affecting a phenotype in family members in an identical way, thereby increasing their similarity beyond what is expected based on genetic resemblance alone, are modeled by one or more C factors. Therefore, by definition, the C factors correlate unity across twins, regardless of zygosity. All environmental influences causing the observed trait to differ in two family members are modeled by one or more E factors. These influences include environmental events to which each family member is uniquely exposed (e.g., two members of a twin pair engaging in different extracurricular activities), as well as events to which multiple family members are exposed, but are affected by in a different way (e.g., both twins may be exposed to parental divorce, but the divorce may affect the trait of interest in each of the twins differently). Thus, by definition, the E factors correlate zero across twins.

The twin design relies on several further assumptions, which include the equal environment assumption (i.e., it is assumed that MZ and DZ twins are equally correlated in their exposure to environmental factors of etiological relevance to the trait under study), equality of variance in MZ and DZ twin pairs, absence of genotype-environment interaction (i.e., of dependency of genetic effects on the environment and vice versa), of genotype-environment correlation (i.e., of non-random placement of genotypes in the range of available environments), of rater bias, and of recruitment bias (e.g., Dolan, 1992; Lykken, McGue, & Tellegen, 1987; Martin & Wilson, 1982; M. C. Neale, Eaves, Kendler, & Hewitt, 1989; Stoolmiller, 1999). The presence of these phenomena does not hamper the approach, but requires them to be modeled explicitly (see e.g., Derks, Dolan, & Boomsma, 2006). For other assumptions of the twin model, see e.g. Derks, et al. (Derks et al., 2006), Falconer & Mackay (Falconer & Mackay, 1996), Lykken, McGue, Bouchard, & Tellegen (Lykken, McGue, Bouchard, & Tellegen, 1990), Martin, Boomsma, & Machin (Martin, Boomsma, & Machin, 1997), Plomin, Defries, McClearn, & McGuffin (Plomin et al., 2008), Purcell (Purcell, 2002).

Figure 1 depicts two examples of the particular multivariate twin model relevant to the present chapter. Within a given model two identical parts are specified, one for each twin. These parts relate the observed phenotypic variables to the latent common variables. For each twin, the covariation in item scores is specified to be a function of the twins' A, C, and E factors. The A, C, and E factors are correlated 1, 1, and 0 in the MZ twins, and .5, 1, and 0 in the DZ twins, respectively. Note that the correlations between the A, C, and E factors within an individual are assumed to be zero, as are the cross-correlations between twin 1 and twin 2. Subsequently, the data are analyzed in a multi-group analysis of MZ and DZ covariance matrices. The expected covariance structure in a multivariate twin model is thus:

[5] In addition, the trait may be influenced by non-additive genetic effects (D). Unlike additive genetic effects, which result from additive action of genes, non-additive genetic effects represent interactive effects of genes on the trait of interest. These will not be modeled in the present chapter, as the classical twin design does not allow for simultaneous estimation of A, D, C, and E effects. In the empirical example, we performed a series of univariate analyses with the results showing most of the items in our dataset to conform better to an ACE than to an ADE model.

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E & r_A\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C \\ r_A\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C & \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E \end{bmatrix}$$

where, given p phenotypes (i.e., observed traits, indicators) per individual, $\boldsymbol{\Sigma}_{11}$ ($\boldsymbol{\Sigma}_{22}$) is the p × p covariance matrix of twin 1 (twin 2), $\boldsymbol{\Sigma}_{12}$ is the twin 1 - twin 2 p × p covariance matrix, and $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_C$ and $\boldsymbol{\Sigma}_E$ are the additive genetic, shared environmental, and unique environmental p × p covariance matrices, respectively. The coefficient r_A is the additive genetic twin correlation (1 for MZ twins, .5 for DZ twins). The $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_C$ and $\boldsymbol{\Sigma}_E$ matrices may be subject to further modeling, as depicted in Figure 1. Although the $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_C$ and $\boldsymbol{\Sigma}_E$ covariance matrices may be subjected to any kind of a covariance structure model (see Boomsma & Molenaar, 1987; Eaves et al., 1986; M. C. Neale & Cardon, 1992), we focus on the type of model depicted in Figure 1.

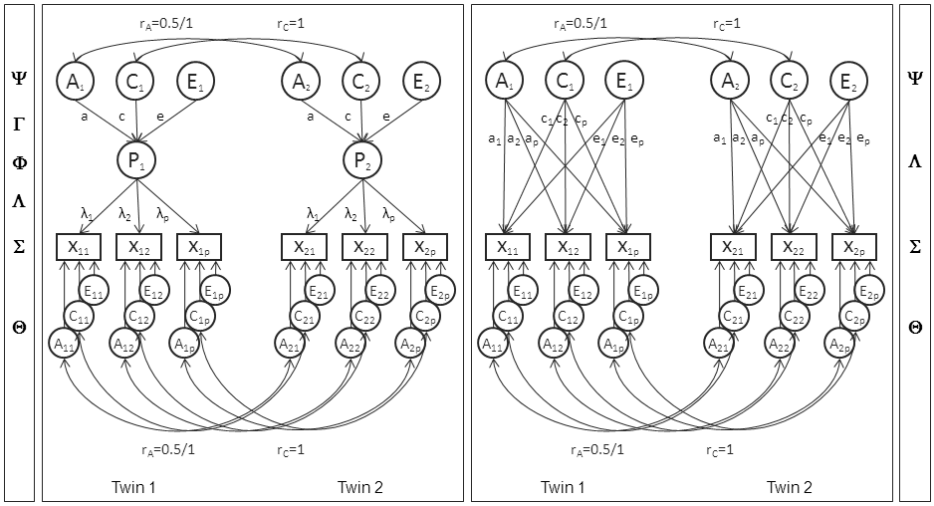

Figure 1. A common pathway (left) and an independent pathway (right) genetic factor model. Matrix names on the sides correspond to notation in the text. Note: as indicated by the notation, the a, c, e, and λ parameters (as well as the residual factor loadings) are subject to equality constraints over Twin 1 and Twin 2.

The first model in Figure 1 is a *common pathway model* (Kendler et al., 1987), also known as the psychometric factor model (McArdle & Goldsmith, 1990). In common pathway models, all of the A, C, and E influences on item covariation are mediated by a latent variable, henceforth referred to as the *psychometric factor* (factors P_1 and P_2 in Figure 1). P_1 and P_2 may be viewed as phenotypic latent factors (i.e., latent factors obtained in factor analysis as usually applied in psychological research), e.g. 'neuroticism' or 'g'. In common pathway models, the psychometric factor acts as a mediator of the genetic and environmental effects.

The second model is the *independent pathway model* (Kendler et al., 1987), also known as the biometric factor model (McArdle & Goldsmith, 1990). An example of this model is depicted in the right panel of Figure 1. In independent pathway models, there is no phenotypic latent variable that mediates genetic and environmental effects on the item responses. Rather, the A, C, and E factors influence item responses directly. In terms of the phenotypic covariance matrix of item responses, we can convey the common and the independent pathway model as follows:

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp} = \boldsymbol{\Lambda}(\boldsymbol{\Phi}_A + \boldsymbol{\Phi}_C + \boldsymbol{\Phi}_E)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp}$$
$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22} = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}_A\boldsymbol{\Psi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Psi}_C\boldsymbol{\Gamma}_C{}^t + \boldsymbol{\Gamma}_E\boldsymbol{\Psi}_E\boldsymbol{\Gamma}_E{}^t)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp}$$
$$\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Lambda}(r_A\boldsymbol{\Phi}_A + \boldsymbol{\Phi}_C)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp21} = \boldsymbol{\Lambda}(r_A\boldsymbol{\Gamma}_A\boldsymbol{\Psi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Psi}_C\boldsymbol{\Gamma}_C{}^t)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_{cp21},$$

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22} = \boldsymbol{\Lambda}_A\boldsymbol{\Psi}_A\boldsymbol{\Lambda}_A{}^t + \boldsymbol{\Lambda}_C\boldsymbol{\Psi}_C\boldsymbol{\Lambda}_C{}^t + \boldsymbol{\Lambda}_E\boldsymbol{\Psi}_E\boldsymbol{\Lambda}_E{}^t + \boldsymbol{\Theta}_{ip}$$
$$\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12} = r_A\boldsymbol{\Lambda}_A\boldsymbol{\Psi}_A\boldsymbol{\Lambda}_A{}^t + \boldsymbol{\Lambda}_C\boldsymbol{\Psi}_C\boldsymbol{\Lambda}_C{}^t + \boldsymbol{\Theta}_{ip21},$$

respectively. Here, $\boldsymbol{\Lambda}$ (in the common pathway model) and $\boldsymbol{\Lambda}_A$, $\boldsymbol{\Lambda}_C$, and $\boldsymbol{\Lambda}_E$ (in the independent pathway model) matrices contain the loadings of the indicators on the psychometric factor and on the biometric (A, C, and E) factors, respectively, and $\boldsymbol{\Psi}_A$, $\boldsymbol{\Psi}_C$, and $\boldsymbol{\Psi}_E$ are the covariance matrices of the A, C, and E factors. In the common pathway model, the covariance matrix of the psychometric factor, $\boldsymbol{\Phi}$, equals $\boldsymbol{\Phi}_A + \boldsymbol{\Phi}_C + \boldsymbol{\Phi}_E$, i.e. $\boldsymbol{\Gamma}_A\boldsymbol{\Psi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Psi}_C\boldsymbol{\Gamma}_C{}^t + \boldsymbol{\Gamma}_E\boldsymbol{\Psi}_E\boldsymbol{\Gamma}_E{}^t$, where $\boldsymbol{\Gamma}_A$, $\boldsymbol{\Gamma}_C$, and $\boldsymbol{\Gamma}_E$ are the vectors of factor loadings $\boldsymbol{\Gamma}_A = [a]$, $\boldsymbol{\Gamma}_C = [c]$, $\boldsymbol{\Gamma}_E = [e]$. Note that in both models the diagonal matrices $\boldsymbol{\Theta}$ (denoted $\boldsymbol{\Theta}_{cp}$ and $\boldsymbol{\Theta}_{ip}$, as they may vary over the models) contain the residuals of the items in the model, and $\boldsymbol{\Theta}_{cp21}$ and $\boldsymbol{\Theta}_{ip21}$ matrices contain the twin 1 - twin 2 covariance among the residuals. The residual matrices may be subjected to their own decomposition, i.e., $\boldsymbol{\Theta} = \boldsymbol{\Theta}_A + \boldsymbol{\Theta}_C + \boldsymbol{\Theta}_E$ and $\boldsymbol{\Theta}_{21} = r_A\boldsymbol{\Theta}_A + \boldsymbol{\Theta}_C$ (M. C. Neale & Cardon, 1992), as depicted in Figure 1. It is immediately clear from Figure 1 that the common pathway model differs from the independent pathway model in the presence of the psychometric factors P1 and P2. As we explain next, this difference can have important implications for dimensionality assessment.

Phenotypic latent variable model and the common pathway model

In the present chapter, we distinguish between genetic factor models (introduced above), and phenotypic factor models. By 'phenotypic factor model' we refer to the factor model as usually formulated and applied in psychological research. The term 'phenotypic' is used because the model is applied only to the observed (i.e., phenotypic) covariation; no genetic information is used. The 8-factor cross-informant model of the CBCL (Achenbach, 1966) and the 5-factor model of personality (McCrae & Costa, 2003) are examples of a phenotypic factor model.

The common pathway model bears a number of similarities to the phenotypic factor model. Notably, both the phenotypic factor model and the common pathway model are based on the assumption that all covariation in item responses is attributable to one or more latent variables. In phenotypic factor modeling, this is formulated as the requirement of measurement invariance: influences of all external variables affecting covariation in item responses run only via the latent variable (Mellenbergh, 1989; Meredith, 1993). Likewise, in common pathway modeling one assumes that all of the A, C, and E influences on item covariation run only via the psychometric factor. That is, there are no direct effects of A, C, and E on the items[6].

The assumption of full mediation of external influences by a latent variable has strong implications. For instance, different external variables affecting a set of item responses via the same latent variable exert the same magnitude of influence relative to each other on all the items that depend on that latent variable. For instance, if an A and a C

[6] As such, the common pathway model may be interpreted as a MIMIC model (Jöreskog & Goldberger, 1975), as the causal influences of A, C, and E factors on the observed responses are mediated by the phenotypic factor. However, in this case the multiple causes are latent rather than observed variables.

variable affect a set of items via the same psychometric factor, the magnitude of influence exerted by the variable A on any individual item will be a scalar multiple of the magnitude of influence exerted by the variable C on that item, and this scalar multiple (k) will be a constant across all the items depending on this psychometric factor. This can be seen from the regression equations describing the common pathway model, e.g. (in terms of the symbols used in Figure 1):

$$x_{11} = \lambda_1(aA_1 + cC_1 + eE_1) + \varepsilon_{11} = \lambda_1 aA_1 + \lambda_1 cC_1 + \lambda_1 eE_1 + \varepsilon_{11},$$
$$x_{12} = \lambda_2(aA_1 + cC_1 + eE_1) + \varepsilon_{12} = \lambda_2 aA_1 + \lambda_2 cC_1 + \lambda_2 eE_1 + \varepsilon_{12},$$

etc. (note that $\varepsilon_{11}=A_{11}+C_{11}+E_{11}$ in Figure 1, etc). In contrast, the independent pathway model imposes no proportionality constraints on the factor loadings, e.g.:

$$x_{11} = a_1A_1 + c_1C_1 + e_1E_1 + \varepsilon_{11},$$
$$x_{12} = a_2A_1 + c_2C_1 + e_2E_1 + \varepsilon_{12},$$

etc. Specifically, letting k denote a positive constant, we note that the introduction of the constraints $a_1/a_2 = c_1/c_2 = e_1/e_2 = k$ renders the common and the independent pathway equations above equivalent (Yung et al., 1999). Thus, the common pathway model makes explicit an assumption of the phenotypic latent variable model concerning the sources of item covariation – all influences on item covariation run via the phenotypic latent variable. This means, barring cases of model equivalence, that *a latent variable model cannot hold unless the corresponding common pathway model holds*. Because any given latent variable hypothesis implies a corresponding common pathway model, a refutation of that common pathway model constitutes evidence against the latent variable hypothesis.

For this reason, one may test the latent variable hypothesis by comparing the fit of a common pathway model to that of a corresponding independent pathway model. Specifically, if a model in which all of the A, C, and E factors exert direct influence on the phenotype fits the data statistically better than a model in which these influences are mediated by a phenotypic latent variable, this would provide evidence against the hypothesis that the effects on the observed item covariation are completely mediated by the phenotypic latent variable. In that case the latent factors employed in the phenotypic factor model are no more than an amalgamation of the direct influences of the A, C, and E factors on the observed item responses. This would have implications for the substantive interpretation of such factors as well-defined, causal entities that produce the observed item covariation (e.g., Borsboom, Mellenbergh, & van Heerden, 2003; Haig, 2005a, 2005b)[7]. If, on the other hand, an independent pathway model does not fit the data better than the corresponding common pathway model, this would provide support for the structure employed in the common pathway model, and substantiation for the corresponding phenotypic latent variable model. Comparison of an independent pathway model and a common pathway model may be conducted using a likelihood ratio test, because, as shown above, a common pathway model can be derived from an independent pathway model by imposing appropriate proportionality constraints on the factor loadings (i.e., the models are nested).

[7] This would, however, not diminish the usefulness of phenotypic latent variables as a means of summarizing data or their utility as predictors. In addition, the specific reasons for rejecting the common pathway model may be local (due only to a subset of observed variables), and thus the violation may be accommodated by the addition of parameters or by the removal of offending variables.

The logic underlying the present approach is essentially the same as that involved in measurement invariance research and MIMIC modeling: the latent variable is required to screen off the effects of genetic and environmental factors (in Pearl, 2000, terminology, the latent variable *d*-separates genes and environment from the item responses). However, what makes the genetic case special is that the A, C, and E factors (a) plausibly determine the variance of the latent variable completely, and (b) can be highly structured by applying standard genetic theory to genetically informative data. This allows for unique possibilities to investigate hypotheses on the origins of structures seen in the correlations among psychometric items. To demonstrate that the proposed methodology works under realistic conditions, we next provide a simulation example.

Simulation study

To illustrate the relationship between the observed association structures and the underlying genetic and environmental structures, we simulated several datasets. In each dataset, a different pattern of genetic and environmental effects gives rise to the observations. These patterns depart progressively from the ideal situation of a common pathway model. As we will show, such departures lead to psychometrically indeterminate covariation structures, in the sense that standard psychometric research practices would not (and in fact could not) converge on correct assessments of the underlying dimensionality. However, we also show that attending to genetic information, present in the widely available twin datasets, allows one to resolve the psychometric puzzle accurately (i.e., to better understand the dimensionality of the dataset).

In total, four datasets were simulated. In the first dataset (Dataset 0) the data are consistent with a common pathway model. In the three subsequent datasets (Datasets 1-3), the assumption of the common pathway model concerning the proportionality of the genetic and environmental effects on the items is violated to an increasing extent. This was achieved by manipulating the dimensionalities of the latent A, C, and E structures (i.e., the order of the covariance matrices $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$). Figure 2 outlines the general structure of this simulation.

Each of the four datasets comprises 12 continuous normally distributed variables per individual (24 variables per twin pair), for 1000 MZ and 1000 DZ twin pairs. We used exact data simulation (i.e., the simulated data fitted the generating model exactly; e.g., van der Sluis, Dolan, Neale, & Posthuma, 2008). We limit the current presentation to a single set of parameter values (given in Table 1)[8], which we do not vary over the four simulations. The manipulation involves only 1) varying the dimensions of the $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ covariance matrices (and the dimensions of the corresponding $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices), and 2) varying the patterns of factor loadings within the $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices. However, all simulations were performed with five different sets of parameter values, and our conclusions were found to be invariant.[9] The simulation script provided in Appendix 2B may be used to verify the generality of our inferences. In the following text, we will first review the four generating models. Subsequently, we present the results of dimensionality assessment for the four datasets.

[8] The table does not detail the parameters of the ACE model for the residuals given our focus on dimensionality assessment; these are given in Appendix 2A and the simulation script (Appendix 2B).
[9] Details on these five sets of parameter values may be obtained from the first author.

Models

The baseline model (**Model 0**, depicted in the first panel of Figure 2) is a common pathway model. The expected phenotypic covariance matrix ($\mathbf{\Sigma}_{CP}$) under this model is:

$$\begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}(\mathbf{\Phi}_A + \mathbf{\Phi}_C + \mathbf{\Phi}_E)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp} & \mathbf{\Lambda}(r_A\mathbf{\Phi}_A + \mathbf{\Phi}_C)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp21} \\ \mathbf{\Lambda}(r_A\mathbf{\Phi}_A + \mathbf{\Phi}_C)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp21} & \mathbf{\Lambda}(\mathbf{\Phi}_A + \mathbf{\Phi}_C + \mathbf{\Phi}_E)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp} \end{bmatrix},$$

where $\mathbf{\Sigma}_{11}$ ($\mathbf{\Sigma}_{22}$) is the 12 × 12 phenotypic covariance matrix of twin 1 (twin 2), $\mathbf{\Sigma}_{12}$ is the 12 × 12 twin 1 - twin 2 phenotypic covariance matrix, $\mathbf{\Lambda}$ is a vector containing the loadings of the indicators on the psychometric factor, $\mathbf{\Phi}_A$, $\mathbf{\Phi}_C$, and $\mathbf{\Phi}_E$ are the A, C, and E variance components of the psychometric factor, respectively, coefficient r_A is the additive genetic twin correlation (1 for MZ twins, .5 for DZ twins), $\mathbf{\Theta}_{cp}$ is a diagonal matrix containing the residuals of the items, and $\mathbf{\Theta}_{cp21}$ and $\mathbf{\Theta}_{ip21}$ are matrices containing the twin 1 - twin 2 covariance among the residuals. In the present case, the variance of each of the items in $\mathbf{\Sigma}_{11}$ ($\mathbf{\Sigma}_{22}$) is 1, and the correlations between the indicators range from .12 to .62.

The model above may also be expressed in terms of parameters of an independent pathway model, as presented in Table 1. In this independent pathway model, the expected covariance structure ($\mathbf{\Sigma}_{IP}$) is:

$$\begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_A\mathbf{\Psi}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Psi}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Lambda}_E\mathbf{\Psi}_E\mathbf{\Lambda}_E{}^t + \mathbf{\Theta}_{ip} & r_A\mathbf{\Lambda}_A\mathbf{\Psi}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Psi}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Theta}_{ip21} \\ r_A\mathbf{\Lambda}_A\mathbf{\Psi}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Psi}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Theta}_{ip21} & \mathbf{\Lambda}_A\mathbf{\Psi}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Psi}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Lambda}_E\mathbf{\Psi}_E\mathbf{\Lambda}_E{}^t + \mathbf{\Theta}_{ip} \end{bmatrix}.$$

Here, $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ vectors contain the loadings of the indicators on the A, C, and E factors, respectively, and the residual matrices $\mathbf{\Theta}_{ip}$, $\mathbf{\Theta}_{ip21mz}$, and $\mathbf{\Theta}_{ip21dz}$ are equal to those in the common pathway model. In the case of the present model (Model 0), $\mathbf{\Sigma}_{CP} = \mathbf{\Sigma}_{IP}$. Note that the independent pathway factor loading parameters above are fully consistent with a common pathway model, i.e., the elements of $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices satisfy the proportionality constraint $a_i/a_{i+1} = c_i/c_{i+1} = e_i/e_{i+1} = k$, where i=1, ...11, and k is a constant). Taking these parameter values as a point of departure, we specify the three subsequent models.

In **Model 1**, the additive genetic influences on the items are represented by two orthogonal A factors per twin. Note that this model (depicted in the second panel of Figure 2) may alternatively be represented as a common pathway model with two phenotypic factors per twin, each factor being a function of its own A, C, and E factor (where the two A factors are uncorrelated, and the two C factors, as well as the two E factors, correlate unity). In this sense, the model does not represent a severe violation of the common pathway structure.

In **Model 2** (depicted in the third panel of Figure 2), the structure employed in Model 1 is further altered, by increasing the dimensionality of the E structure. This model represents a more severe violation of the common pathway structure, as the items here no longer cluster identically with regard to A and E influences (i.e., the patterns of factor loadings in the $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_E$ matrices differ from each other). For instance, sets of items that form a unidimensional structure with respect to additive genetic influences, are two-dimensional with respect to unique environmental influences. In **Model 3** (fourth panel of Figure 2), the common pathway structure is further violated by increasing the dimensionality of the latent C structure. Here, the clustering of the items is markedly different with regard to the A, C, and E influences; thus the phenotypic dimensionality is a function of A, C, and E influences that severely violate the common pathway structure.

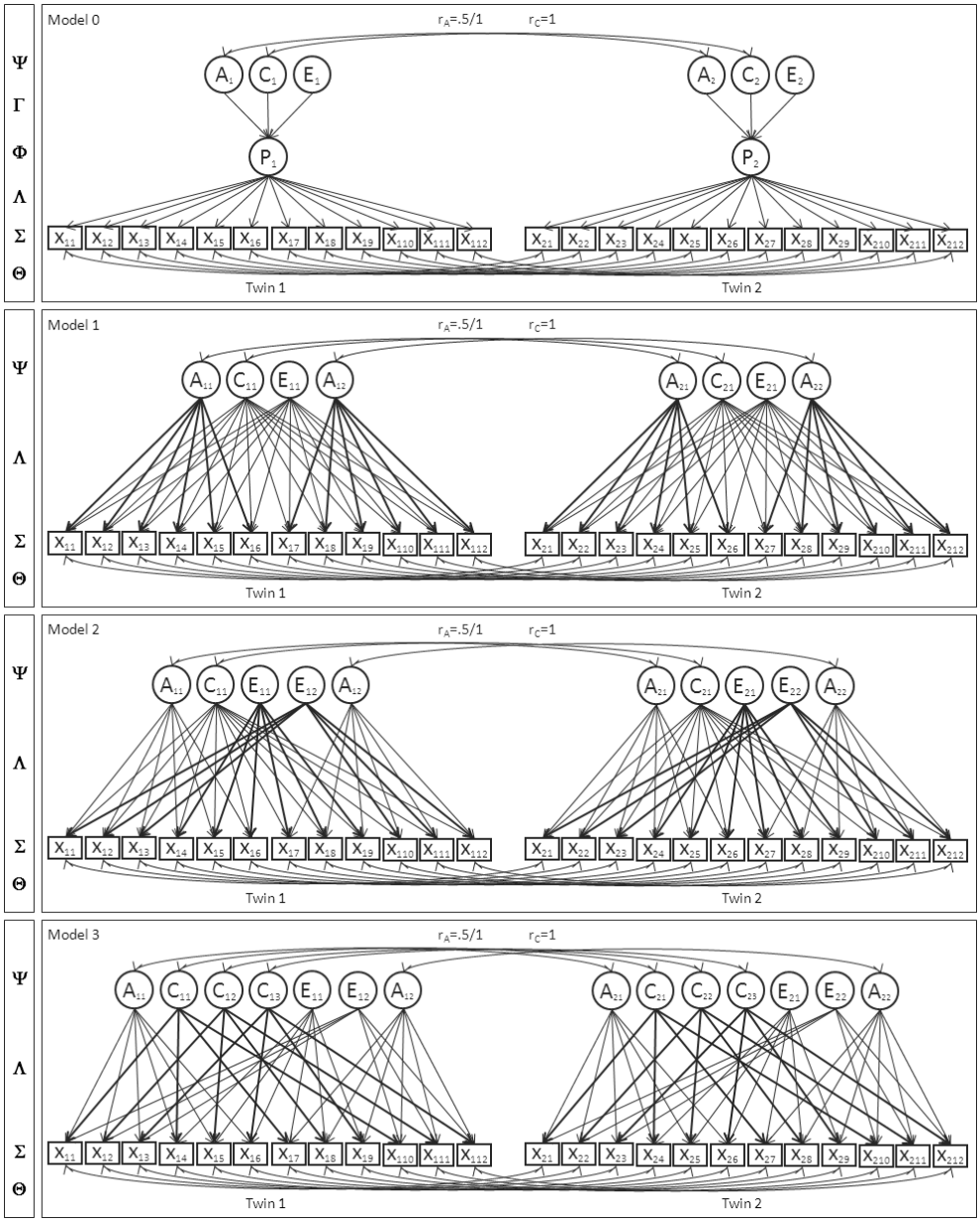

Figure 2. Path diagrams of Models 0-3. Matrix names on the left correspond to the notation in the text.

Table 1
Parameters of Models 0-3.

Model 0 in terms of common pathway paratemers:

$\boldsymbol{\Lambda}^t$ = [√.1, √.15, √.2, √.25, √.3, √.35, √.4, √.45, √.5, √.55, √.6, √.65]
$\boldsymbol{\Gamma}_A$ = [√.5], $\boldsymbol{\Gamma}_C$ = [√.3], $\boldsymbol{\Gamma}_E$ = [√.2]
$\boldsymbol{\Psi}_A = \boldsymbol{\Psi}_C = \boldsymbol{\Psi}_E$ = [1]
$\boldsymbol{\Theta}_{cp} = \mathbf{I}$ - diag($\boldsymbol{\Lambda}(\boldsymbol{\Phi}_A + \boldsymbol{\Phi}_C + \boldsymbol{\Phi}_E)\boldsymbol{\Lambda}^t$) = $\mathbf{I}$ - diag($\boldsymbol{\Lambda}(\boldsymbol{\Gamma}_A\boldsymbol{\Psi}_A\boldsymbol{\Gamma}_A{}^t + \boldsymbol{\Gamma}_C\boldsymbol{\Psi}_C\boldsymbol{\Gamma}_C{}^t + \boldsymbol{\Gamma}_E\boldsymbol{\Psi}_E\boldsymbol{\Gamma}_E{}^t)\boldsymbol{\Lambda}^t$)
= diag(.9,.85,.8,.75,.7,.65,.6,.55,.5,.45,.4,.35)
$\boldsymbol{\Theta}_{cp21mz}$ = .8*$\mathbf{I}$ - diag($\boldsymbol{\Lambda}(\boldsymbol{\Phi}_A+\boldsymbol{\Phi}_C)\boldsymbol{\Lambda}^t$)
= diag(.72,.68,.64,.6,.56,.52,.48,.44,.4,.36,.32,.28)
$\boldsymbol{\Theta}_{cp21dz}$ = .55*$\mathbf{I}$ - diag($\boldsymbol{\Lambda}(.5\boldsymbol{\Phi}_A+\boldsymbol{\Phi}_C)\boldsymbol{\Lambda}^t$)
= diag(.495, .4675, .44, .4125, .385, .3575, .33, .3025, .275, .2475, .22, .1925)

Model 0 in terms of independent pathway paratemers:

$\boldsymbol{\Lambda}_A{}^t = \boldsymbol{\Gamma}_A\boldsymbol{\Lambda}^t$ = [√.05, √.075, √.1, √.125, √.15, √.175, √.2, √.225, √.25, √.275, √.3, √.325]
$\boldsymbol{\Lambda}_C{}^t = \boldsymbol{\Gamma}_C\boldsymbol{\Lambda}^t$ = [√.03, √.045, √.06, √.075, √.08, √.105, √.120, √.135, √.150, √.165, √.180, √.195]
$\boldsymbol{\Lambda}_E{}^t = \boldsymbol{\Gamma}_E\boldsymbol{\Lambda}^t$ = [√.02, √.03, √.04, √.05, √.06, √.07, √.08, √.09, √.10, √.11, √.12, √.13]
$\boldsymbol{\Psi}_A = \boldsymbol{\Psi}_C = \boldsymbol{\Psi}_E$ = [1], $\boldsymbol{\Theta}_{ip} = \boldsymbol{\Theta}_{cp}$, $\boldsymbol{\Theta}_{ip21mz} = \boldsymbol{\Theta}_{cp21mz}$, $\boldsymbol{\Theta}_{ip21dz} = \boldsymbol{\Theta}_{cp21dz}$

Model 1:

$\boldsymbol{\Psi}_A$ = diag(1, 1)

$\boldsymbol{\Lambda}_A{}^t$ = √.05 √.075 √.1 √.125 √.15 √.175 √.2 √.225 √.25 √.275 √.3 √.325

Model 2:

$\boldsymbol{\Psi}_E$ = diag(1, 1)

$\boldsymbol{\Lambda}_E{}^t$ = √.02 √.03 √.04 √.05 √.06 √.07 √.08 √.09 √.10 √.11 √.12 √.13

Model 3:

$\boldsymbol{\Psi}_C$ = diag (1, 1, 1)

$\boldsymbol{\Lambda}_C{}^t$ = √.03 √.045 √.06 √.075 √.08 √.105 √.120 √.135 √.150 √.165 √.180 √.195

$\boldsymbol{\Lambda}$ = vector containing the loadings of the indicators on the psychometric factor
$\boldsymbol{\Gamma}_A$, $\boldsymbol{\Gamma}_C$, $\boldsymbol{\Gamma}_E$ = vectors of factor loadings of the psychometric factor on the A, C, and E factors
$\boldsymbol{\Psi}_A$, $\boldsymbol{\Psi}_C$, $\boldsymbol{\Psi}_E$ = covariance matrices of the A, C, and E factors
$\boldsymbol{\Theta}_{cp} = \boldsymbol{\Theta}_{ip}$ = 12 x 12 diagonal matrix containing the residual item variances
$\boldsymbol{\Theta}_{cp21mz} = \boldsymbol{\Theta}_{ip21mz}$ = 12 x 12 diagonal matrix of twin 1 - twin 2 covariances among MZ twins
$\boldsymbol{\Theta}_{cp21dz} = \boldsymbol{\Theta}_{ip21dz}$ = 12 x 12 diagonal matrix of twin 1 - twin 2 covariances among DZ twins
$\boldsymbol{\Lambda}_A$, $\boldsymbol{\Lambda}_C$, $\boldsymbol{\Lambda}_E$ = matrices containing direct factor loadings of the items on the A, C, and E factors

Note. The models are conveyed only in terms of parameters that differ from the preceding model. For instance, the parameter matrices not listed under Model 1 ($\boldsymbol{\Lambda}_C$, $\boldsymbol{\Lambda}_E$, $\boldsymbol{\Psi}_C$, $\boldsymbol{\Psi}_E$, $\boldsymbol{\Theta}_{ip}$ and $\boldsymbol{\Theta}_{ip21}$) equal those in Model 0. In addition, the factor loading parameters are conveyed in terms of square roots, as this gives straightforward information on the proportion of variance explained (e.g. a factor loading of $\sqrt{.1}$ indicates $\sqrt{.1}^2=.1$ explained variance).

Analyses

The analyses of the datasets consisted of two parts. In the first part, the aim was to examine the effect that the violations of the common pathway structure had on the phenotypic dimensionality estimates. To this end, the dimensionality of the datasets was assessed using EFA. The phenotypic latent factors obtained in the EFA were subsequently used as a basis for specifying confirmatory genetic factor common pathway models. As in standard genetic research, here we decomposed the variation in the latent factors obtained in the phenotypic EFA into genetic and environmental components. In the second part, the aim was to obtain a clearer indication of the data generating mechanism by disposing of the hypotheses concerning the number of latent variables in the model, and applying independent pathway modeling in a purely exploratory manner, to uncover the (possibly different) structures of the A, C, and E influences. Specifically, we used EFA to determine the possibly different dimensionalities of the covariance matrices $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_C$, and $\boldsymbol{\Sigma}_E$, in terms of the latent covariance matrices $\boldsymbol{\Psi}_A$, $\boldsymbol{\Psi}_C$, and $\boldsymbol{\Psi}_E$. Here the dimensionality of the observed covariance matrix is a function of the A, C, and E covariance structures, which may differ in dimensionality, and in no way satisfy the common pathway model. The advantage of this is that it provides an insight into the dimensionality of the phenotypic structure that does not assume, but does not exclude, the common pathway model. The analyses were performed using Mplus (L. K. Muthén & Muthén, 2007), Mx (M. C. Neale, 2000), and R (R Core Team, 2013).[10] In evaluating model fit, we used the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA).

Results

Given that Model 0 has a unidimensional structure and was used only as a baseline model from which parameter values were derived, we limit the presentation to the results obtained in analyses of Datasets 1-3.

Dataset 1. Seeing as Model 1 can be viewed as a 2- factor common pathway model in which the two C factors, as well as the two E factors, correlate unity, one can simply accommodate the violation of the 1-factor common pathway structure by fitting a 2-factor model. The phenotypic EFA results, a summary of which is provided in Table 2, reflect this: a 2-factor EFA solution (detailed in Table 2) provides a perfect fit to the data, as do a 2-factor common pathway model ($\chi^2=0$, df=581, p=1, RMSEA=0, CFI=1, TLI=1) and a 2 A, 2 C, 2 E independent pathway model ($\chi^2=0$, df=508, p=1, RMSEA=0, CFI=1, TLI=1) based on this 2-factor EFA solution. Note that perfect fit is associated with c^2 values of zero because we used exact data simulation. The parameter estimates obtained in genetic factor modeling indicate that C and E are unidimensional (the correlations between the two C factors in both the common and the independent pathway model are 1, as are the correlations between the two E factors), while A may be represented by two orthogonal factors. The structure depicted in

[10] All scripts may be obtained from the first author upon request. We alternated between Mplus and Mx because Mplus estimates the polychoric correlations very efficiently, while Mx's matrix-based syntax is very convenient in fitting models involving high dimensional Cholesky decompositions. R was used for its data simulation features.

the second panel of Figure 2 therefore need not preclude accurate dimensionality assessment. However, one might consider situations in which the data generating structure is less consistent with the common pathway model; in the following examples we consider such more severe violations of the common pathway structure.

Dataset 2. In Model 2, the $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_E$ matrices are both two-dimensional, but the items cluster differently with regard to A and E influences (e.g., clusters of items that form a unidimensional structure with respect to additive genetic influences, are two-dimensional with respect to unique environmental influences). Note that the data generating structure may still be accommodated by a common pathway model with 4 phenotypic factors, each affecting 3 items. However, as common pathway analyses are confirmatory in nature and predicated on the results of phenotypic analyses, we first investigated whether phenotypic EFA correctly indicated the number of phenotypic factors needed to account for the covariance structure.

The results of the EFA are shown in Table 2. Here, both the 1-factor and the 2-factor solution were clearly rejected by the χ^2 statistic, but in the 3-factor solution both the χ^2 and the RMSEA indicated a perfect fit (χ^2=0, df=33, p=1, RMSEA=0, CFI=1, TLI=1). In the 4-factor solution the same was the case, although the model (based on promax rotation; presented in Table 2) does not correspond to the data generating structure. Moreover, in the 4-factor solution none of the items appear to be best represented by the third factor, and only one item loads substantially (factor loading above √.025) on the fourth factor. Considering the fit statistics and the factor structure in Table 2, it appears that in a standard situation of dimensionality assessment the 3-factor solution would be a compelling choice.

Based on this 3-factor EFA solution (detailed in Table 2), we specified a 3-factor common pathway model and a corresponding independent pathway model, depicted in Figure 3. In both of these models, the phenotypic covariation in twin 1 (twin 2) is a function of three mutually correlated A (C, E) factors (i.e., $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ are 3 x 3 matrices with freely estimated off-diagonal elements). Although inclusion of cross-loadings improves model fit, we specify simple structure models given our focus on dimensionality assessment. For the common pathway model, the fit measures were χ^2 (577) = 2158, p<.001, RMSEA=.052, CFI=.944, TLI=.944, and for the independent pathway model χ^2 (507) = 1148, p<.001, RMSEA=.036, CFI=.976, TLI=.974. Additional analyses showed that inclusion of cross-loadings (as indicated by the EFA solution) improves model fit for both the common pathway and the independent pathway model; however, even then, the parameter estimates remain somewhat biased. Thus, even assuming lack of simple structure, these models are still unable to precisely convey the actual A, C and E effects on the items. If one considers the generating model (3rd panel Figure 2), it is clear why this is the case: a model that assumes equal clustering of items with regard to A, C, and E effects (as does any model based on phenotypic factor analysis) cannot adequately describe the data generating mechanism. Although we detail only the results based on the 3-factor EFA solution, none of the EFA solutions presented in Table 2 correctly convey the genetic and environmental effects on the items. It is interesting to note that the misspecification of the phenotypic models (in the sense that none accurately represented the data generating structure) was not evident in the fit measures associated with the models; the fit measures associated with all but the 1-factor the EFA solution indicated an excellent fit.

Table 2

Results of phenotypic EFA of Datasets 1-3

EFA Dataset 1

	χ^2	df	p	RMSEA
1f	325.3	54	0	.0003
2f	0	43	1	0
3f	0	33	1	0
4f	0	24	1	0

EFA Dataset 2

	χ^2	df	p	RMSEA
1f	470.0	54	0	.0004
2f	133.6	43	1	.0002
3f	0	33	1	0
4f	0	24	1	0

EFA Dataset 3

	χ^2	df	p	RMSEA
1f	543.8	54	0	.0004
2f	328.5	43	0	.0004
3f	196.8	33	0	.0004
4f	89.5	24	0	.0003
5f	0	16	1	0

2-factor EFA solution Dataset 1

Factor	1	
correlations:	.04	1

Uniqueness: .9 .85 .8 .75 .7 .65
.6 .55 .5 .45 .4 .35

Promax-rotated factor loadings:

	P1	P2
x1		.309
x2		.378
x3		.437
x4		.489
x5		.535
x6		.578
x7	.633	
x8	.671	
x9	.708	

3-factor EFA solution Dataset 2

Factor	1		
correlations:	.24	1	
	-.68	.16	1

Uniqueness: .9 .85 .8 .75 .7 .65
.6 .55 .5 .45 .4 .35

Promax-rotated factor loadings:

	P1	P2	P3
x1		.191	.253
x2		.234	.310
x3		.270	.358
x4		.507	.121
x5		.555	.132
x6		.600	.143
x7	.521	.170	
x8	.553	.181	
x9	.583	.190	

5-factor EFA solution Dataset 3

Factor	1				
correlations:	.09	1			
	-.06	.09	1		
	.04	.04	.23	1	
	-.02	-.01	.26	.25	1

Uniqueness: .9 .85 .8 .75 .7 .65
.6 .55 .5 .45 .4 .35

Promax-rotated factor loadings:

	P1	P2	P3	P4	P5
x1					.289
x2		-.167		.281	.108
x3	.267	-.209	.309		.127
x4	-.208	.249	.227		.358
x5	.157		.243	.429	
x6			.562		
x7		.639			.297
x8		.402		.449	-.109
x9	.311	.398	.367		-.113

x10	.742		x10	.798	-.105	.352	x10	.362	.365	-.175		.495
x11	.775		x11	.833	-.109	.367	x11	.478		-.191	.474	
x12	.807		x12	.867	-.114	.382	x12	.782		.209		
Prop var	.263	.107	Prop var	.251	.102	.063	Prop var	.102	.084	.065	.058	.050
Cum var	.263	.371	Cum var	.251	.353	.416	Cum var	.102	.186	.250	.309	.359

4-factor EFA solution Dataset 2

Promax-rotated factor loadings:

	P1	P2	P3	P4
x1				.459
x2		.285	.215	
x3		.329	.249	
x4		.507		
x5		.555		
x6		.600		
x7	.596	.107		
x8	.632	.113		
x9	.666	.12		
x10	.737		.51	
x11	.770		.533	
x12	.801		.554	

Based on this 3-factor EFA solution (detailed in Table 2), we specified a 3-factor common pathway model and a corresponding independent pathway model, depicted in Figure 3. In both of these models, the phenotypic covariation in twin 1 (twin 2) is a function of three mutually correlated A (C, E) factors (i.e., $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ are 3 x 3 matrices with freely estimated off-diagonal elements). Although inclusion of cross-loadings improves model fit, we specify simple structure models given our focus on dimensionality assessment. For the common pathway model, the fit measures were χ^2 (577) = 2158, p<.001, RMSEA=.052, CFI=.944, TLI=.944, and for the independent pathway model χ^2 (507) = 1148, p<.001, RMSEA=.036, CFI=.976, TLI=.974. Additional analyses showed that inclusion of cross-loadings (as indicated by the EFA solution) improves model fit for both the common pathway and the independent pathway model; however, even then, the parameter estimates remain somewhat biased. Thus, even assuming lack of simple structure, these models are still unable to precisely convey the actual A, C and E effects on the items. If one considers the generating model (3rd panel Figure 2), it is clear why this is the case: a model that assumes equal clustering of items with regard to A, C, and E effects (as does any model based on phenotypic factor analysis) cannot adequately describe the data generating mechanism. Although we detail only the results based on the 3-factor EFA solution, none of the EFA solutions presented in Table 2 correctly convey the genetic and environmental effects on the items. It is interesting to note that the misspecification of the phenotypic models (in the sense that none accurately represented the data generating structure) was not evident in the fit measures associated with the models; the fit measures associated with all but the 1-factor the EFA solution indicated an excellent fit.

The second part of the analyses was aimed at directly addressing the dimensionality issue, without reference to the phenotypic factor structure. To this end, the A, C, and E components of the observed covariance structure were estimated from the data, and the dimensionality of each of those components was separately evaluated using EFA. The A, C, and E covariance components (i.e., the 12 x 12 $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ matrices) may be estimated from twin data by fitting the model given on page 35. These analyses were carried out in Mx (Neale, 2000). Subsequently, each of the three covariance matrices was subjected to EFA. As we do not assume any phenotypic model and make no predictions about the dimensionalities of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ covariance components, this approach is purely exploratory.

The results of the EFA are given in Table 3 and Figure 4. As apparent from both the Table and the scree-plots in the Figure, the results correctly indicate the order of the $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ matrices to be 2, 1, and 2, respectively. The estimated factor loadings of the corresponding EFA solutions with 2 A, 1 C, and 2 E factors, shown in the lower panel of Figure 4, correspond exactly to the parameters of the generating model.

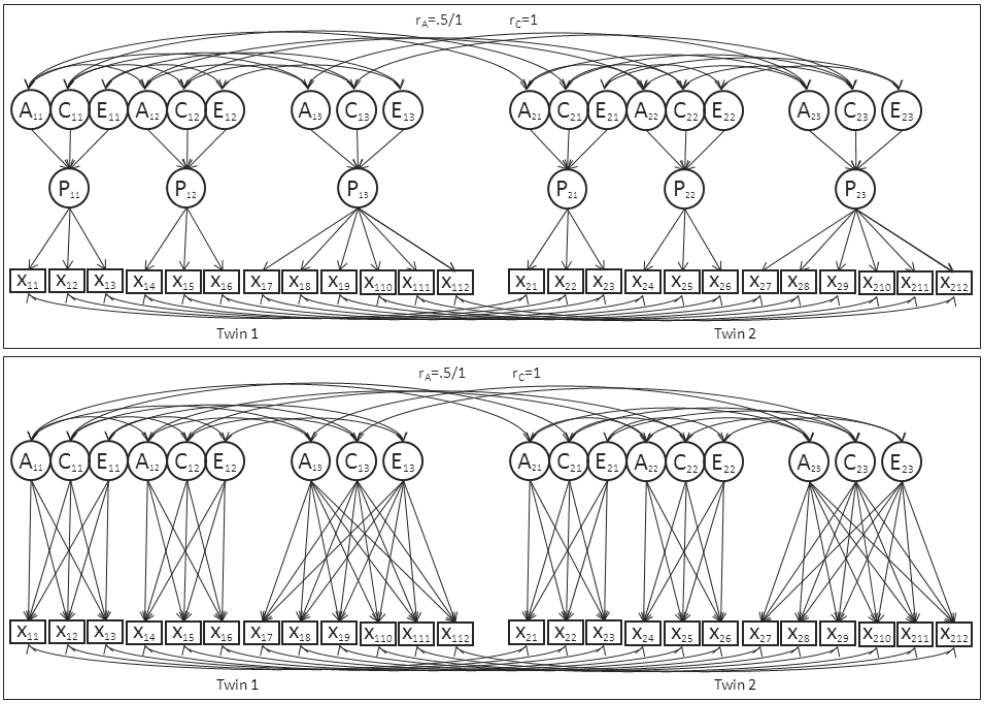

Figure 3. A common pathway (upper panel) and an independent pathway (lower panel) model based on the phenotypic EFA of Dataset 2.

Dataset 3. In Model 3, the A, C, and E structures differ appreciably from each other. The results of phenotypic EFA, shown in Table 2, indicate that a model with five phenotypic latent factors provides an adequate description of the data. However, as evident from the Table, the pattern of factor loadings in this model is inconsistent with a simple structure; thus deciding on the number of actual latent dimensions underlying the data and the nature of the factors is complicated. Given that none of the EFA solutions in Table 2 can correctly convey the genetic and environmental effects on the items, we do not detail the possible confirmatory common and independent pathway models one may fit to the data given the EFA results. Instead, we present the solution obtained by the EFA of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ variance components (Figure 5 and Table 3). As evident from both the Table and the Figure, a 2 A, 3 C, and 2 E structure is clearly supported by the EFA results, and both the factor loading structure and the values of the factor loading parameters are recovered correctly (lower panel Figure 5).

Finally, we note that the chosen dimensionalities of the $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, $\mathbf{\Psi}_E$ matrices represent only one instance of a possible violation of the common pathway model. In the present simulation, the values within the $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices are still consistent with a common pathway model (i.e., the non-zero elements of $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices satisfy the proportionality constraint $a_i/a_{i+1} = c_i/c_{i+1} = e_i/e_{i+1} = k$, where i=1, ...11, and k is a constant). In other words, the correlation between the non-zero values in the $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices is 1, i.e. the factor loadings are collinear. It is possible to further violate the common pathway structure by manipulating the correlations between the values in the $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_C$, and $\mathbf{\Lambda}_E$ matrices. However, this violation is less detrimental to model fit than are the differences in dimensionalities of $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, $\mathbf{\Psi}_E$ matrices.

Table 3
Fit statistics obtained in EFA of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ matrices in Datasets 2 and 3

		Dataset 2				Dataset 3			
		χ^2	df	p	RMSEA	χ^2	df	p	RMSEA
A	1 f	519.7	54	0	.0004	519.7	54	0	.0004
	2 f	0	43	1	0	0	43	1	0
C	1 f	0	54	1	0	1157.7	54	0	.0006
	2f	0	43	1	0	495	43	0	.0005
	3f	0	33	1	0	0	33	1	0
E	1 f	1302.9	54	0	.0007	1302.9	54	0	.0007
	2 f	0	43	1	0	0	43	1	0

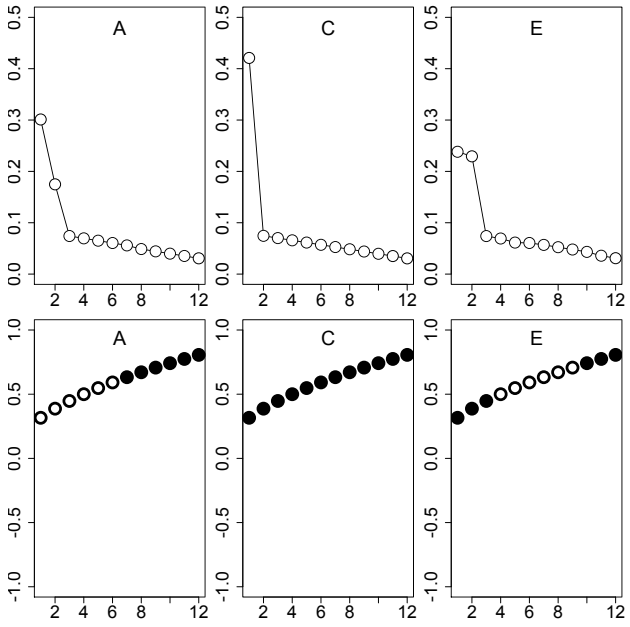

Figure 4. Dataset 2: Normalized eigenvalues for the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ matrices (upper panel) and factor loadings obtained in the EFA solutions with 2 A, 1 C, and 2 E factors (lower panel). Colors code for different latent factors.

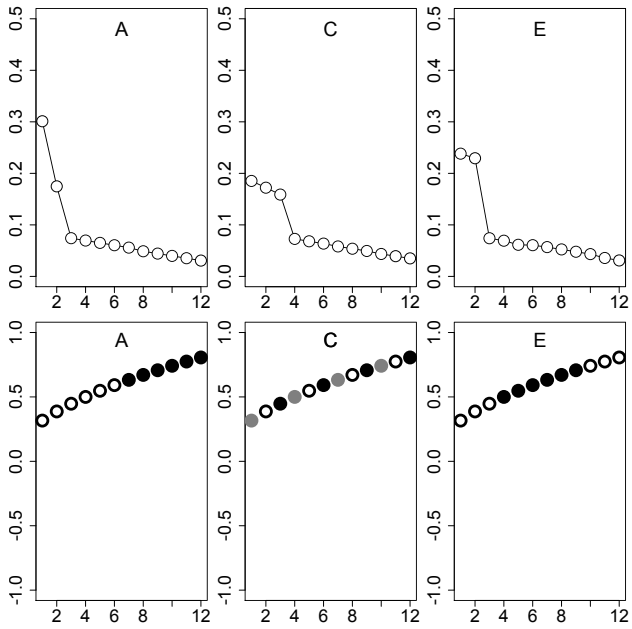

Figure 5. Dataset 3: Normalized eigenvalues for the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ matrices (upper panel) and factor loadings obtained in EFA solutions with 2 A, 3 C, and 2 E factors (lower panel). Colors code for different latent factors.

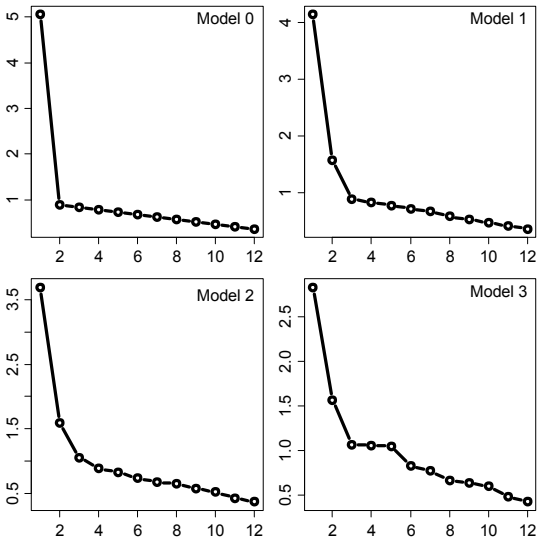

Figure 6. Eigenvalues of the phenotypic covariance matrices for Models 0-3.

The present simulation shows that the clustering of the items with respect to genetic and environmental effects is required to be identical for a unidimensional latent variable model to hold. This is in line with the theoretical derivation presented earlier in this chapter. In addition, it shows that if genetic and environmental effects do not cluster identically, psychometric analyses may fail to correctly indicate the dimensionality of the latent space. In these cases, the data will either show a significant degree of indeterminacy with respect to alternative dimensional hypotheses, or will support an incorrect latent structure. However, attending to the genetic and environmental antecedents of the items succesfully resolved the dimensionality issue. We now apply this methodology to an empirical dataset.

Illustration: childhood internalizing problems

Internalizing problems concern conditions such as anxiety, depression, and somatization. Dimensionality assessment has traditionally been difficult for such problems. For instance, current diagnostic systems like the DSM-IV (American Psychiatric Association, 1994) distinguish anxiety and mood disorders as separate categories, but there is a significant amount of evidence to suggest that the overlap between such disorders is larger than can be reasonably expected were such a categorical distinction between types of disorders correct (e.g., Brady & Kendall, 1992). This is supported by genetic analyses, which univocally suggest that the genetic effects that impact anxiety and depression are shared, while the unique environmental effects are not (see e.g. Kendler et al., 1987; Kendler, Neale, Kessler, Heath, & Eaves, 1992; Middeldorp, Cath, Van Dyck, & Boomsma, 2005). This presents an extraordinarily difficult task for the test constructor. For how should items that probe different anxiety and mood related problems be allocated to subscales? Can we reasonably expect a clear outcome of dimensionality assessment in this case? In the present example, we show that such an outcome is unrealistic given the genetic and environmental background of internalizing problems. In addition, we show how the use of genetic information uncovers a complex dimensional pattern that can be used to further the psychometric understanding of test scores.

Data

The data were obtained from the Netherlands Twin Register at the VU University Amsterdam (Bartels, van Beijsterveldt, et al., 2007; Boomsma et al., 2006), and consist of maternal ratings of 11,565 twins (including 2,085 MZ and 3,599 DZ complete twin pairs) of mean age 10.1 (SD = .4) on the *Internalizing* grouping of the Dutch version of the Child Behavior Checklist for Ages 4-18 (CBCL/4-18; Achenbach, 1991; Verhulst, Van der Ende, & Koot, 1996). The *Internalizing* grouping of the CBCL is a scale designed to measure disturbances in intropunitive emotions and moods in children, and consists of 3 subscales: *Anxious/Depressed* (AD), *Withdrawn* (W), and *Somatic Complaints* (SC), comprising 31 discrete items (listed in Appendix 2C) in total. Responses are given on a three-point scale.[11]

[11] Returning to the aforementioned assumptions of the twin design: in the present study, we tested a number of these assumptions, including absence of rater bias and absence of recruitment bias. The issue of rater bias was addressed by comparing the standard deviations observed in our sample to those of normative samples (Verhulst et al., 1996). These were found to differ only slightly: the ratios of our standard deviations to those of normative samples are .91, .83, and .95, for the *Anxious/Depressed, Withdrawn,* and *Somatic Complaints* scales, respectively. The issue of rater bias in the present data has been addressed in the past. Bartels, Boomsma, Hudziak, van Beijsterveldt, and van den Oord (2007) report that in a subset (N=7718) of the present sample, the estimate of the upper bound of the phenotypic variance that may contain rater bias is ~ .14.

Descriptive statistics

The item distributions were positively skewed, with response rates ranging from 54.8% to 96.8% (M = 84.8, SD = 10.3) for the response 0 (symptom not present), 2.9% to 41.2% (M = 13.2, SD = 9.4) for the response 1 (symptom somewhat/sometimes present), and 0.2% to 6.4% (M = 1.5, SD = 1.3) for the response 2 (symptom very/often present). MZ and DZ twin item correlations and the distribution of inter-item correlations are depicted in Appendix 2C.

Analyses

As in the preceding example, the analyses consisted of two parts. In the first set of analyses, the phenotypic dimensionality of the dataset was assessed using EFA, and the solutions obtained in EFA were tested in a confirmatory manner, by 1) specifying and fitting simple structure phenotypic models based on the EFA results[12], and b) subsequently using these simple structure models as a basis for specifying genetic common and independent pathway models. In common pathway models, the variance in the phenotypic factors obtained in EFA was decomposed into A, C, and E components. The independent pathway models were based on the common pathway models, in the sense that they retain the structure employed in the common pathway models (i.e., they contain the same number of A, C, and E factors, affecting the same clusters of items), but dispose of the psychometric factors, i.e. allow for the items to load directly on the A, C, and E factors. Thus, the common pathway models represent a special case of (i.e., are nested under) the independent pathway models. By comparing the fit of these common and independent pathway models, we address the focal question of whether one can interpret the phenotypic common factors substantively and causally.

In the second set of analyses, independent pathway modeling was applied in an exploratory manner. In particular, the analyses consisted of estimating the unconstrained genetic and environmental covariance matrices (i.e., the 31 × 31 additive genetic, shared and unshared environmental covariance matrices $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$), and subjecting each of these covariance matrices to EFA to obtain an indication of their dimensionality, i.e., the order of the covariance matrices $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$.

As in the simulation example, the analyses were performed using Mplus, Mx, and R.[13] Given the discrete nature of the items, we fitted discrete factor models (i.e., we assumed the discrete indicator variables to be a realization of a continuous normal latent process, and fit models to polychoric correlations; Flora & Curran, 2004; Wirth & Edwards, 2007) using the robust weighted least squares estimator (WLSMV; L. K. Muthén & Muthén, 1998-2007). The polychoric correlations between the 31 items and between the 62 (31 per twin) items served as input in the phenotypic and the genetic factor analyses, respectively. In evaluating model fit, we used CFI, TLI, and RMSEA. As both our sample size and the models employed were large, the chi-square statistic was of limited use as an overall fit measure (Jöreskog, 1993), and was employed only to test local hypotheses concerning comparisons of nested models, as these comparisons are associated with a smaller approximation error.

[12] EFA and CFA were performed using split half validation. Cases were randomly assigned to either half of the sample; one half was subsequently used for EFA (N=5782) and the other for CFA (N=5783).

[13] The scripts used to perform the analyses may be obtained from the first author upon request.

Results

The initial analysis involved phenotypic EFA of the 31 items. The term 'phenotypic' here indicates that only the observed (phenotypic) covariation is analyzed, i.e. the analysis does not exploit the fact that the sample consists of familially related individuals.[14] EFA indicated two well-fitting models, a 3- and a 4-factor model (depicted in Figure 7a and 7b). Interestingly, in both of these models, the items originally belonging to the Anxious/Depressed scale cluster into those appearing to be more relevant to anxiety (*3. Fears doing something bad, 4. Must be perfect, 8. Nervous, tense, 9. Fearful, anxious, 10. Feels too guilty, 11. Self-conscious, 14. Worries*) and those more related to depression (*1. Lonely, 2. Cries a lot, 5. Feels unloved, 6. Others out to get him, 7. Feels worthless, 12. Suspicious, 13. Sad*). Note that the depiction in the Figure is simplified insofar as only the path with the highest factor loading is shown for each item. The factor loadings associated with the paths omitted from the Figure equal .05 on average; for comparison, the mean of the factor loadings for the depicted paths equals .58. Detailed results, including item content, factor loadings, factor correlations, and proportion of variance explained (R^2) are given in Appendix 2C.

Subsequently, based on the EFA results and the standard CBCL/4-18 model, a 3- and a 4-factor phenotypic model (Figure 7c and 7b) were specified and fitted to the data.

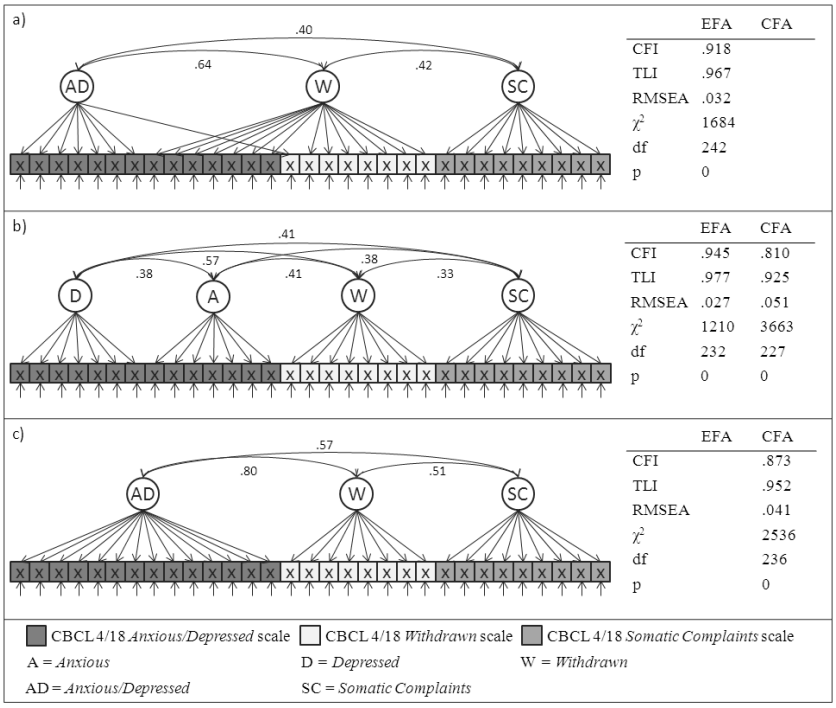

Figure 7. a) The 3-factor model based on EFA, b) the 4-factor model based on EFA, c) the standard CBCL/4-18 3-factor model. Right: fit indices obtained in EFA (geomin rotation)

[14] As treating observations from the same family as independent may result in biased estimates, we performed a correction for clustering available in MPlus, which has been shown to work well in this context (Rebollo, de Moor, Dolan, & Boomsma, 2006).

and CFA. N=5782 for EFA, N=5783 for CFA. For EFA solutions, only the path with the highest factor loading is shown for each item.

As is evident from the Figure, a firm distinction was hard to make between the fit of a model in which the anxiety and depression items represent a single dimension (Figure 7c), and a model in which they represent two distinct dimensions (Figure 7b). The additional solution provided by EFA (Figure 7a), in which items associated with anxiety load on the Withdrawn factor, obtained a similar fit. Whereas items pertaining to somatic complaints consistently form one dimension, the dimensionality of items pertaining to depression, anxiety, and withdrawn behavior therefore remains less clear. This is perhaps not surprising in the light of the well-established difficulty of distinguishing phenotypically the dimensions of anxiety and depression (e.g., see Clark & Watson, 1991).

In the next step, the results of phenotypic analyses were used as a basis for specifying genetic factor models. In common pathway models, the factor structure of the models tested in the phenotypic CFA (Figure 7b and 7c) was retained, and the contributions of the A, C, and E factors to the phenotypic latent factors investigated. The 3- and the 4-factor common pathway models specified in this way differ only minimally in terms of model fit: the respective fit measures were χ^2 (583) = 2030, p<.001, CFI=.952, TLI=.966, RMSEA=.030 and χ^2 (584) = 1811, p<.001, CFI=.959, TLI=.971, RMSEA=.027. In independent pathway modeling, the structure employed in the 3- and the 4-factor common pathway models was retained, but the psychometric factors are disposed of, i.e. the items were allowed to load directly on the A, C, and E factors. Again, the 3- and the 4-factor model differed only minimally in terms of model fit; the fit measures associated with the two models were χ^2 (534) = 1142, p<.001, CFI=.980, TLI=.984, RMSEA=.020, and χ^2 (542) = 1161, p<.001, CFI=.979, TLI=.984, RMSEA=.020, respectively.[15]

Returning to the focal question of whether the independent pathway models fit the data appreciably better than the corresponding common pathway models, we compared the general fit of the models and carried out likelihood ratio tests of the proportionality constraints mentioned above. These tests revealed both the 3- and the 4-factor-based independent pathway models to fit better than their common pathway counterparts (χ^2 difference tests:[16] χ^2 (25) = 1066, p<.001 for the 3-factor based models, and χ^2 (23) = 864, p<.001 for the 4-factor based models). This implies that the common pathway models, in which the latent variables mediate all of the A, C and E effects on individual phenotypic differences, fail to convey accurately the genetic and environmental effects on the items. Again, we note that the misspecification of the common pathway models was not evident in the fit measures associated with the models. Both common pathway models obtained a good fit, and the same is true of the phenotypic models.

In the second set of analyses, we employed EFA to evaluate the dimensionalities of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ covariance matrices. The results are shown in Figure 8. An inspection of scree plots indicates a 1-dimensional C structure. The structures of A and E matrices remain, however, somewhat less clear. To explore the A and E structures further, we use the EFA results as a basis for specifying a number of competing independent pathway models with varying A, C, and E dimensionalities, and fit these models to the phenotypic covariance

[15] Given that the fit of the 3- and the 4-factor model is virtually indistinguishable, in practice one might simply accept the 3-factor model on the basis of parsimony. However, given our interest in the specific reasons for the nearly identical fit, at this point we make no decisions on which model to accept and proceed with the analyses.

[16] For WLSMV estimators the standard approach of taking the difference between chi-square values and the corresponding degrees of freedom is not appropriate because the chi-square difference is not chi-square distributed (Muthén & Muthén, 1998-2007). We therefore performed chi-square difference testing using scaling correction factors (Satorra & Bentler, 2001).

matrix. An example of these confirmatory independent pathway models is depicted in Figure 9. A detailed overview of the fit measures and inter-factor correlations associated with each of the models is given in Appendix 2C (Table 6). Overall, a comparison of these models indicated a model with 2A, 1C, and 4E factors as the best-fitting model ($\chi^2(531) = 1082$, $p<.001$, CFI=.982, TLI=.986, RMSEA=.019). This model is depicted in Figure 9. It should, however, be noted that most of the models tested did not differ considerably in terms of model fit; therefore the structure in Figure 9 need not necessarily be conclusive. In addition, rejecting a common pathway model in favor of the corresponding independent model does not establish the structure employed in the independent model as in any way definitive, and there is, naturally, a possibility of other types of models (e.g., the mutualism model of van der Maas et al., 2006, or the network model of Cramer et al., 2010) providing a better account of the data. The use of the independent pathway model, as presented in this chapter, is merely instrumental to testing the mediation of external effects on item covariation by a latent variable. Furthermore, the present results do strongly suggest a unidimensional C structure, and multidimensional (but mutually differing) A and E structures.

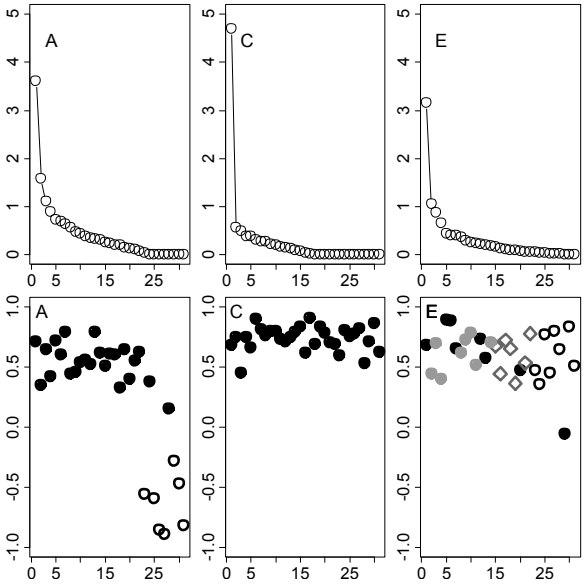

Figure 8. CBCL/4-18 dataset: Eigenvalues of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ matrices (upper panel) and factor loadings obtained in EFA solutions with 2 A, 1 C, and 4 E factors (lower panel). Shading/shapes code for different latent factors. Only the highest factor loading for each item is shown.

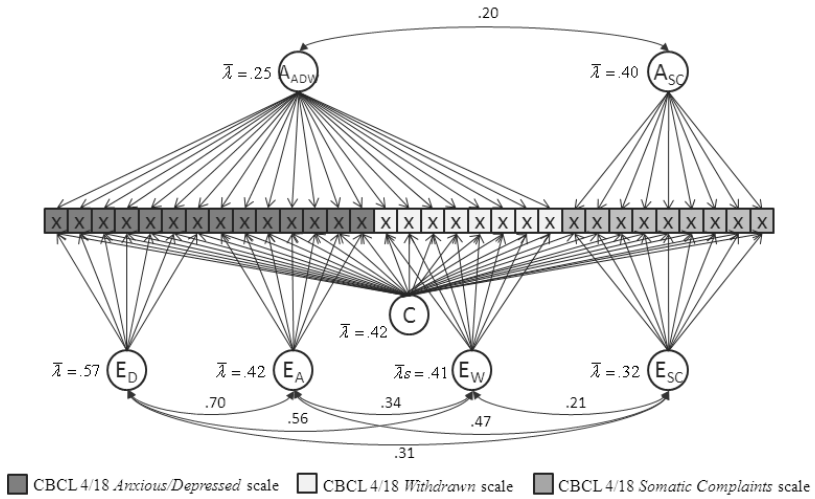

Figure 9. The 2A 1C 4E independent pathway model for the CBCL/4-18 *Internalizing* scale. Item residuals are not depicted but are estimated in the model.

In light of the present results, the results of the phenotypic analyses start to make more sense; the inability of phenotypic modeling to distinguish between several models appears to be due to the phenotypic structure being generated by three different sources: a 1-dimensional C, a 2-dimensional A, and a 4-dimensional E source.

Discussion

Even though the analysis and determination of dimensionality is of central importance in psychological science, currently available strategies for dimensionality assessment often leave the issue undecided. Building on ideas concerning genetic item analysis, as developed in quantitative genetics (Eaves, 1983; Heath, Jardine, Eaves, & Martin, 1989; Kendler et al., 1987; M. C. Neale, Lubke, Aggen, & Dolan, 2005; van den Berg, Glas, & Boomsma, 2007; Waller & Reise, 1992), the present chapter has outlined how genetic information may be brought to bear on the dimensionality assessment problem. In particular, the methodology outlined in this chapter may be used with genetically informative data to a) put latent variable hypotheses to a stronger test than is possible in purely phenotypic analyses, and b) gain insight into why dimensionality issues may be difficult to settle.

The methodology proposed in this chapter may therefore not only improve dimensionality assessment, but may also suggest explanations of why specific dimensional hypotheses are violated. While dimensionality assessment remains a difficult and to some extent subjective task, these methods therefore offer enhanced resolution relative to that possible in purely phenotypic analyses. Importantly, we do not claim that assessment of phenotypic dimensionality without incorporating genetic information cannot produce correct results, or that genetic analyses render standard methods obsolete. Rather, we think that genetic designs offer an underutilized and informative source of data that may help researchers to better understand the dimensionality of their constructs. Practically we envisage a situation in which phenotypic dimensionality research produces varied results, which will in practice simply result in disagreement concerning dimensionality. For

instance, this is the case for cognitive abilities, with respect to which there are competing models which differ in dimensionality (notwithstanding many decades of research). One solution to this is to collect larger data sets. However, the present chapter suggests to researchers that the greater resolution provided by larger data sets may not provide the answer. We propose that it might be useful to seek out twin data in order to investigate possible differences in dimensionality of genetic and environment influences on major constructs.

As mentioned previously, the logic underlying our approach is essentially the same as that involved in measurement invariance research and MIMIC modeling. Moreover, common and independent pathway model comparisons have been considered outside the context of genetics (see e.g. Carlson & Mulaik, 1993). However, what makes the analyses presented here different is that, unlike in standard MIMIC modeling, the A, C, and E factors determine the variance of the latent variable completely. Furthermore, the situation in which the common pathway model is rejected is at least as informative as that in which it is retained, as the information contained in genetically informative datasets allows one to examine the exact nature of violations of dimensional assumptions; something that is typically not the case in standard MIMIC modeling. In the twin model, one can establish whether or not the common pathway model fits and, in case of misfit, can arrive at a detailed account of the cause of misfit, thereby moving the question of dimensionality from the phenotypic level to the genetic level and the environmental level. This increased resolution (i.e., the possibility to view the lack of unidimensionality of the observed covariation structure as a function of the dimensionalities of its underlying genetic and environmental structures) is unique to the twin design and is, in our opinion, a particularly powerful aspect of the present method.

In our illustrative analyses, the incorporation of genetic information turned out to be highly informative. In standard phenotypic analyses, it proved difficult to decide whether a three- or four-dimensional latent structure underlies the data - a situation that is not uncommon in psychometric investigations into dimensionality, where one often has to decide between solutions that differ substantively but appear to be nearly equivalent statistically. Incorporating genetic information, however, suggested that the reason for the ambiguity in the data with respect to these structures is that *several* different models are correct, but apply to different sources of item covariation: a 2-factor model seems to better reflect additive genetic influences, whereas a 4-factor model better reflects unique environmental influences. Interestingly, common environmental influences appear to influence item scores across the board, suggesting that the common part of environmental variation varies along a single dimension.[17]

The question of how many dimensions are measured by the *Internalizing* scales of the CBCL can now be viewed from a new perspective, which may be surprising to the psychometrician: in terms of genetic variance the items appear to measure two dimensions, corresponding to genetic influences distinctly affecting symptoms of depression/anxiety/withdrawal, on the one hand, and somatic complaints, on the other. This implies, for instance, that genes act in a nonspecific way to influence the chance of developing depression-, anxiety-, and withdrawal-related symptomatology. Individual-specific environmental influences distinctly affect symptoms of depression, anxiety, withdrawal, and somatic complaints (thus, individual-specific environmental events may be e.g. specifically depressogenic or specifically anxiogenic), whereas environmental events

[17] It is possible that a unidimensional C component partly stems from method variance. For instance, variance due to rater bias, if not explicitly modeled, is absorbed by C (Neale & Cardon, 1992). Given data by multiple raters, it is possible to test for presence of rater bias.

shared by family members appear to either have a positive or a negative effect on the entire range of symptoms. In terms of, for instance, the anxiety/depression distinction, the present results suggest that these two syndromes share the same genetic basis, but are distinctly affected by individual-specific environmental events - a finding that is in line with prior genetic investigations into the dimensionality of anxiety and depression (e.g., Kendler et al., 1987; Kendler et al., 1992; Middeldorp et al., 2005).

It should be noted that the current results do not necessarily reflect upon the utility of the CBCL in the clinical context; we do not doubt its usefulness for diagnostic purposes, especially given that the broad structure found in our analyses is in line with the current item allocation of the CBCL. However, in the context of research one should bear in mind that the current scales may not measure three distinct sets of genetic, common environmental and individual-specific environmental influences, but possibly reflect a more complex underlying structure. Depending on the specific research goals, the results of this type of analysis may provide a basis for redefining the current scales to arrive at distinct measures of each of these sources of influences (e.g., if one's aim is to measure common environmental influences, one may view the item set as unidimensional and accordingly derive a single sumscore from the data).

Naturally, the results of this type of analyses are relevant not only to theories of psychopathology; we consider their implications to be much wider. For instance, theories in developmental psychology may benefit from investigating the individual differences in the development of behavior as a function of genetic and environmental influences, or examining how the various dimensions of environmental and genetic influences change and develop over time. Also, the results might have implications for genetics itself. Specifically, the search for genes affecting specific behaviors is often based on a composite measure of a phenotype, such as a sumscore. However, if these phenotypes are heterogeneous, analogous to the way the CBCL appears heterogeneous, using a total score as a basis for gene search would appear suboptimal, as the total score itself might not accurately reflect the genetic structure underlying the data (van der Sluis et al., 2010). We consider this issue to be important, because power to detect the effects of measured genes is likely to suffer if the phenotypic measure is not correctly defined (van der Sluis et al., 2010). Independent pathway item level analysis, as described in this chapter, offers possibilities for redefining the phenotypic scores in terms of genetic and environmental effects. This may in turn allow for using latent trait estimates derived from a model such as that in Figure 7, as a basis for gene search.

In addition to these practical benefits of the present methodology, there are important conceptual considerations that follow from the ideas presented in this chapter. For instance, latent variable models like the factor model can be viewed as incorporating hypotheses concerning a common cause structure that underlies item covariation (Borsboom, 2008; Borsboom et al., 2003; Haig, 2005a, 2005b). However, the question of whether latent variables hypothesized in a given context may be said to exist and have causal relevance is a point of dispute in many fields; one need only consider the fields of intelligence and personality research, where considerable controversy exists regarding the theoretical status of variables such as the *g*-factor and the five factors of personality. To the extent that such models survive the confrontation with genetic information, as described here, they may be considered more strongly corroborated than they could be in analyses of purely phenotypic data. However, if models for genetic and environmental effects have different structures, as was the case for the illustration data in this chapter, the common factors found in our phenotypic analysis may in fact be an amalgam of several different

genetic and environmental models. Clearly, in this case, the ascription of causal force to such amalgams is problematic.

In conclusion, we expect that the methodology proposed in this chapter may bear considerable fruit in disentangling dimensionality issues in the research areas where they have generated controversy, and shed light on the theoretical status of important hypothesized latent variables in intelligence, psychopathology, and personality research. The time is ripe for investigations along these lines. In the past decades, behavior genetics researchers have constructed large and well-archived twin and family registries that are perfectly suited for analyses such as those reported here (e.g., a 1998 review lists 16 twin registries in Europe alone; Boomsma, 1998; Busjahn, 2002). The datasets contained in those registries are typically obtainable via protocols for collaborative projects, and in some cases even publically available (e.g. Add Health; see Harris, Halpern, Smolen, & Haberstick, 2006). In addition, the development of psychometric software as well as the current speed of computers have led to a situation where the required statistical analyses have become feasible. In our view, this opens up a wealth of possibilities for refining and extending psychometric investigations beyond the analysis of purely phenotypic covariation.

Chapter 4

Three-and-a-half-Factor Model? The Genetic and Environmental Structure of the CBCL/4-18 Internalizing Grouping

Abstract

In the present chapter, multivariate genetic item analyses were employed to address questions regarding the ontology and the genetic and environmental etiology of the Anxious/Depressed, Withdrawn, and Somatic Complaints syndrome dimensions of the Internalizing grouping of the Child Behavior Checklist/6-18 (CBCL/6-18). Using common and independent pathway genetic factor modeling, it was examined whether these syndrome dimensions can be ascribed a realist ontology. Subsequently, the structures of the genetic and environmental influences giving rise to the observed symptom covariation were examined. Maternal ratings of a population-based sample of 17,511 Dutch twins of mean age 7.4 (*SD*=.4) on the items of the Internalizing grouping of the Dutch CBCL/6-18 were analyzed. Applications of common and independent pathway modeling demonstrated that the Internalizing syndrome dimensions may be better understood as a composite of unconstrained genetic and environmental influences than as causally relevant entities generating the observed symptom covariation. Furthermore, the results indicate a common genetic basis for anxiety, depression, and withdrawn behavior, with the distinction between these syndromes being driven by the individual-specific environment. Implications for the substantive interpretation of these syndrome dimensions are discussed.

Appendices can be obtained at http://sanjafranic.com/dissertation.

Based on: Franić, S., Dolan, C. V., Borsboom, D., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2014) Three-and-a-Half-Factor Model? The Genetic and Environmental Structure of the CBCL/6–18 Internalizing Grouping. *Behavior Genetics, 44*(3), 254-268.

The development of taxonomy of psychiatric symptoms has traditionally been challenging. Difficulties in delineating between diagnostic categories, arising from issues such as overlapping features of multiple disorders, inconsistent empirical evidence regarding the factor structure of psychometric instruments, definitional issues arising from high comorbidity rates, debates regarding dimensional vs. categorical conceptualization, and unknown degree of etiological overlap between symptoms or sets of symptoms, have notoriously hampered the attempts of arriving at a classification of psychopathology that would gain univocal support from empirical researchers and clinical practitioners alike. In children, these issues are further exacerbated by the developmental aspect of the disorders: for instance, the same disorder may manifest itself through different symptoms over time, while identical symptoms may reflect distinct, temporally changing underlying conditions. Symptoms of anxiety and depression, for instance, famously illustrate the aforementioned issues (Brady & Kendall, 1992; Brown, 1996; Clark & Watson, 1991; Mineka, Watson, & Clark, 1998; Rapee, Schniering, & Hudson, 2009). The definitional and etiological questions surrounding these disorders (and their high comorbidity rates) are as old as the systematic study of the disorders itself. Are these two highly comorbid disorders manifestations of a single syndrome, or separate entities with overlapping features? To what extent are their etiologies shared? Is their symptom overlap a reflection of inadequacies of the current diagnostic systems, or an indication of a shared etiology? These and similar questions have stimulated ample and diverse theoretical development, and motivated a vast amount of research. The theories range from those postulating anxiety and depression as different points along a single continuum, to those conceptualizing them as conceptually and empirically distinct phenomena (Clark, 1989).

This complexity, inherent to the study of psychiatric disorders, is further compounded by a lack of agreement in evaluating and understanding the structure of psychometric instruments used to assess psychopathology. The Child Behavior Checklist (CBCL, Achenbach & Rescorla, 2001), for instance, is one of the most widely used instruments in assessing childhood psychopathology. It has been translated into over 85 languages, and more than 6,000 publications from over 65 countries report its applications in both the practical and the research context. However, when faced with critical empirical and psychometric evaluations, the syndrome dimensions postulated in the CBCL do not always stand up to scrutiny. In possibly the most comprehensive critical psychometric/empirical evaluation of the CBCL to date, Hartman et al. (Hartman et al., 1999) demonstrated that the 8-factor cross-informant model of the CBCL (Achenbach & Rescorla, 2001, described below) fails to adequately describe the empirical data across multiple cultures under study, in both population-based and clinical samples. Furthermore, if violations of distributional assumptions, invariably present in the analysis of CBCL data, are taken into account when evaluating model fit, the conclusions of the studies indicating acceptable or nearly acceptable fit are often undermined (see Hartman et al., 1999). Upon close scrutiny, it therefore appears that the postulated 8-factor structure of the CBCL does not consistently survive critical confrontation with empirical data. This, naturally, raises questions about the instrument's validity: what do the CBCL syndrome dimensions measure, given the lack of unambiguous empirical support for the proposed 8-factor structure?

In the present chapter, we propose that multivariate genetic item analysis (e.g., Eaves, 1983; Franić, Dolan, Borsboom, Hudziak, et al., 2013; Heath, Eaves, & Martin, 1989; Heath, Jardine, et al., 1989; Kendler et al., 1987; M. C. Neale et al., 2005; van den Berg et al., 2007; Waller & Reise, 1992), as first applied to individual psychiatric symptoms by Kendler et al. (Kendler et al., 1987), can be used to illuminate some of the aforementioned issues.

Specifically, genetic item analyses can be employed to examine how some of the difficulties in delineating the CBCL syndrome dimensions may arise as a function of the complexity of the latent genetic and environmental structures that underlie the observed symptom covariation. In addition, the applications of this type of analysis can contribute to the discussion on whether the current CBCL dimensions may be conceptualized as well-defined, coherent entities exerting causal influence on item covariation (i.e., whether they can be ascribed a realist ontology; Borsboom et al., 2003), or are better considered an unconstrained amalgamation of genetic and environmental influences. In the present chapter, we focus on the Internalizing grouping of the CBCL (items of the CBCL pertaining to intropunitive emotions and moods), with the aim of answering two principal questions: 1. Can one interpret the Internalizing syndrome dimensions of the CBCL substantively and causally? 2. What is the structure of the genetic and environmental influences giving rise to the observed (i.e., phenotypic) symptom covariation? We do not place primary emphasis on detailed phenotypic dimensionality assessment, and use it mainly insofar as it serves as a gateway into exploring the latent genetic and environmental dimensionality.

Method

Data

The data were obtained from the Netherlands Twin Register at VU University Amsterdam (Bartels, Beijsterveldt, et al., 2007; van Beijsterveldt, Groen-Blokhuis, Hottenga, Franić, et al., 2013) and consist of maternal ratings of a population-based sample of 17,511 twins (including 3,023 MZ and 5,599 DZ complete twin pairs) of mean age 7.4 (SD = .4) on the Internalizing grouping of the Dutch version of the Child Behavior Checklist for Ages 6-18 (CBCL/6-18; Achenbach & Rescorla, 2001[18]). The CBCL/6-18 is a 140-item questionnaire used to assess problem behaviors and competencies in children, as reported by their parents. The cross-informant model of the CBCL (Achenbach & Rescorla, 2001) was derived through the application of principal components analysis, and consists of 8 correlated syndrome dimensions, broadly clustered into those pertaining to internalizing problems (the Internalizing grouping) and those pertaining to externalizing problems (the Externalizing grouping). The Internalizing grouping of the CBCL is a scale designed to measure disturbances in intropunitive emotions and moods in children, and consists of 3 subscales (i.e., syndrome dimensions): Anxious/Depressed (AD), Withdrawn (W), and Somatic Complaints (SC), containing 31 discrete items (listed in Figure 1) in total. Responses are given on a three-point scale: "not true", "somewhat or sometimes true", and "very true or often true". A path-diagrammatic representation of the three syndrome dimensions of the Internalizing grouping is given in Figure 1.

The analytic framework employed in the present study has been outlined in the preceding chapter, but is repeated in the *Approach* section below for completeness. The reader familiar with Chapter 3 may skip to the *Analyses* section.

[18] The study had permission to permission to use, reproduce and reformat the CBCL.

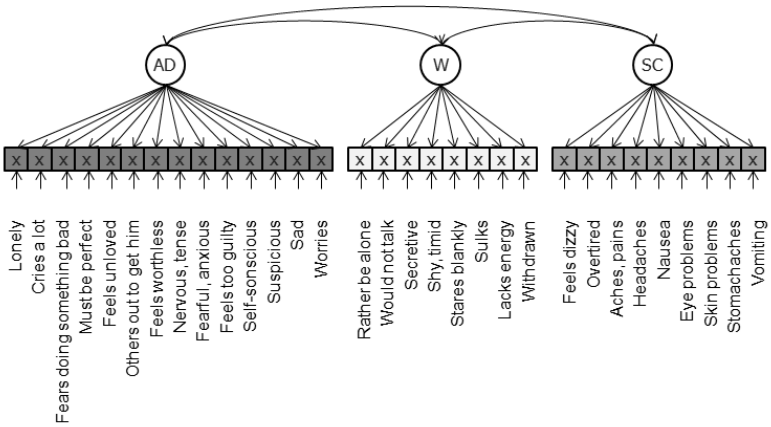

Figure 1. The syndrome dimensions and item content of the CBCL 6/18 Internalizing grouping. AD = Anxious/Depressed, W = Withdrawn, SC = Somatic Complaints.

Approach

Genetic covariance structure modeling (Martin & Eaves, 1977) is the application of structural equation modeling (Bollen, 1989; Kline, 2005) to data collected in genetically informative samples, such as samples of twins (Franić et al., 2012; M. C. Neale & Cardon, 1992). In the classical twin design, the sample consists of monozygotic (MZ) and dizygotic (DZ) twin pairs. DZ twins share an average of 50% of their segregating genes, while MZ twins share their segregating DNA (Falconer & Mackay, 1996; J. P. van Dongen et al., 2012). In the present analyses, the covariance structure of the phenotypes (i.e., observed traits, symptoms) is modeled as a function of latent factors representing three sources of individual differences: additive genetic (A), shared environmental (C) and individual-specific environmental (E) sources. Additive genetic influences are modeled by one or more A factors, which represent the total additive effects of genes relevant to the phenotypes. Based on quantitative genetic theory (Falconer & Mackay, 1996; Mather & Jinks, 1971), the A factors are known to correlate 1 across MZ twins and .5 across DZ twins. Environmental influences affecting the phenotype of both twins in an identical way, thereby increasing their similarity beyond what is expected based on genetic resemblance alone, are represented by one or more C factors. Therefore, by definition, the C factors correlate unity across twins (regardless of zygosity). All environmental influences causing the phenotype of two family members to differ are represented by one or more E factors. Thus, by definition, the E factors are correlated 0 across twins.[19] The expected covariance structure in a multivariate twin model is thus:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E & \Pi\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C \\ (\Pi\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C)^t & \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E \end{bmatrix}$$

In addition, the phenotype may be influenced by non-additive genetic effects (D), which are the result of interactions of alleles within the same locus (genetic dominance) or across different loci (epistasis). These will not be modeled in the present chapter, as the classical twin design does not allow for simultaneous estimation of A, D, C, and E effects. The choice between modeling C and D effects was informed by preliminary univariate item analyses, which showed most of the items to conform better to an ACE than to an ADE model. We note, however, that this does not exclude the presence of non-additive genetic influences (Keller and Coventry, 2005).

where, given p phenotypes, $\mathbf{\Sigma}_{11}$ ($\mathbf{\Sigma}_{22}$) is the p x p covariance matrix of twin 1 (twin 2), $\mathbf{\Sigma}_{12}$ ($\mathbf{\Sigma}_{21}$) is the twin 1 - twin 2 p x p covariance matrix, and $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ are the additive genetic, shared environmental, and unique environmental p x p covariance matrices, respectively. The coefficient P is the additive genetic twin correlation (1 for MZ twins, .5 for DZ twins).

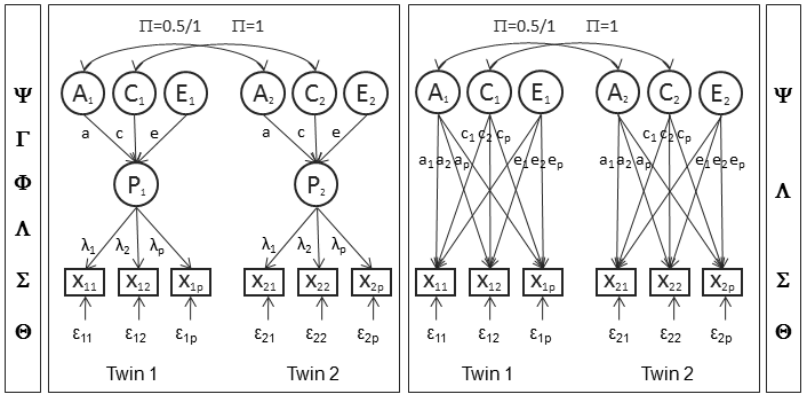

Figure 2. A common pathway (left) and an independent pathway (right) genetic factor model. Matrix names on the sides correspond to notation in the text.

Figure 2 depicts two examples of the multivariate twin models used in the present study. The first model is a *common pathway model* (Kendler et al., 1987), also known as the psychometric factor model (McArdle & Goldsmith, 1990). In a common pathway model, all of the A, C, and E influences on the item responses are mediated by a latent variable, henceforth referred to as the psychometric factor (factors P_1 and P_2 in Figure 2). P_1 and P_2 may be viewed as latent phenotypic factors, e.g. 'anxiety' or 'depression'. In common pathway models, the psychometric factor acts as a mediator of the genetic and environmental effects, and the factor loadings represent common pathways from the A, C, and E factors to the observed item responses.

The second model is the *independent pathway model* (Kendler et al., 1987), also known as the biometric factor model (McArdle & Goldsmith, 1990). This model is represented in the right panel of Figure 2. In the independent pathway model, there is no phenotypic latent variable that mediates the genetic and environmental effects on the item responses. Rather, the A, C, and E factors influence item responses directly. In terms of the phenotypic (i.e., observed) covariance matrix of the item responses (i.e., $\mathbf{S}_{11} = \mathbf{S}_{22}$), we can convey the common and the independent pathway models, respectively, as follows:

$$\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22} = \mathbf{\Lambda\Phi\Lambda}^t + \mathbf{\Theta}_{cp} = \mathbf{\Lambda}(\mathbf{\Phi}_A + \mathbf{\Phi}_C + \mathbf{\Phi}_E)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp} = \mathbf{\Lambda}(\mathbf{\Gamma}_A\mathbf{\Gamma}_A{}^t + \mathbf{\Gamma}_C\mathbf{\Gamma}_C{}^t + \mathbf{\Gamma}_E\mathbf{\Gamma}_E{}^t)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp}$$
$$\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}{}^t = \mathbf{\Lambda}(\mathbf{\Phi}_A + \mathbf{\Phi}_C)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp21} = \mathbf{\Lambda}(\mathbf{\Pi\Gamma}_A\mathbf{\Gamma}_A{}^t + \mathbf{\Gamma}_C\mathbf{\Gamma}_C{}^t)\mathbf{\Lambda}^t + \mathbf{\Theta}_{cp21}$$

and

$$\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22} = \mathbf{\Lambda}_A\mathbf{\Psi}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Psi}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Lambda}_E\mathbf{\Psi}_E\mathbf{\Lambda}_E{}^t + \mathbf{\Theta}_{ip} = \mathbf{\Lambda}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Lambda}_E\mathbf{\Lambda}_E{}^t + \mathbf{\Theta}_{ip}$$
$$\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}{}^t = \mathbf{\Pi\Lambda}_A\mathbf{\Lambda}_A{}^t + \mathbf{\Lambda}_C\mathbf{\Lambda}_C{}^t + \mathbf{\Theta}_{ip21}.$$

Here $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ are the covariance matrices of the A, C, and E factors in the two models. In the common pathway model the covariance matrix of the psychometric factor, $\mathbf{\Phi}$,

equals $\mathbf{\Phi}_A + \mathbf{\Phi}_C + \mathbf{\Phi}_E$, i.e. $\mathbf{\Gamma}_A\mathbf{\Gamma}_A{}^t + \mathbf{\Gamma}_C\mathbf{\Gamma}_C{}^t + \mathbf{\Gamma}_E\mathbf{\Gamma}_E{}^t$, where $\mathbf{\Phi}_A$, $\mathbf{\Phi}_C$, and $\mathbf{\Phi}_E$ denote the A, C, and E variance components of $\mathbf{\Phi}$, and $\mathbf{\Gamma}_A$, $\mathbf{\Gamma}_C$, and $\mathbf{\Gamma}_E$ are the vectors of factor loadings $\mathbf{\Gamma}_A$ = [a], $\mathbf{\Gamma}_C$ = [c], $\mathbf{\Gamma}_E$ = [e]. Note that in both models the diagonal matrices $\mathbf{\Theta}$ (denoted $\mathbf{\Theta}_{cp}$ and $\mathbf{\Theta}_{ip}$, as they may vary over the models) contain the residual variances of the items in the model. The residual covariance matrices may be subjected to their own decomposition, i.e., $\mathbf{\Theta} = \mathbf{\Theta}_A + \mathbf{\Theta}_C + \mathbf{\Theta}_E$ and $\mathbf{\Theta}_{21} = \mathbf{\Pi\Theta}_A + \mathbf{\Theta}_C$ (M. C. Neale & Cardon, 1992).

In the present chapter, we distinguish between genetic factor models (introduced above), and phenotypic factor models. By 'phenotypic factor model' we refer to the factor model as usually formulated and applied in psychological research. The term 'phenotypic' is used because the model is applied only to the observed (i.e., phenotypic) covariation; no genetic information is used.[20] The 8-factor cross-informant model of the CBCL and the 5-factor model of personality (McCrae & Costa, 2003; McCrae & John, 1992) are examples of a phenotypic factor model.

The common pathway model bears a number of similarities to the phenotypic factor model. Notably, both the phenotypic factor model and the common pathway model are based on the premise that all covariation in item responses is attributable to one or more latent variables. In phenotypic factor modeling, this can be formulated in terms of measurement invariance: influences of all external variables affecting covariation in item responses run only via the latent variable (Mellenbergh, 1989; Meredith, 1993). Likewise, in common pathway modeling one assumes that all of the A, C, and E influences on item covariation run only via the psychometric factor. That is, there are no direct effects of A, C, and E on the items. The assumption of full mediation of external influences by a latent variable has strong implications. For instance, different external variables affecting a set of item responses via the same latent variable exert the same magnitude of influence relative to each other on all the items that depend on that latent variable. For instance, if an A and a C variable affect a set of items via the same psychometric factor, then the magnitude of influence exerted by the variable A on any individual item will be a scalar multiple of the magnitude of influence exerted by the variable C on that same item, and this scalar multiple (k) will be a constant across all the items depending on this psychometric factor. This means that one can derive a common pathway model from an independent pathway model by imposing proportionality constraints on the factor loadings, such that $a_1/a_2 = c_1/c_2 = e_1/e_2 =$ k (following the notation in the right panel of Figure 2).

Thus, the common pathway model makes explicit an assumption of the phenotypic latent variable model concerning the sources of item covariation – all influences on item covariation run via the phenotypic latent variable. This means, barring cases of model equivalence, that a latent variable model cannot hold unless the corresponding common pathway model holds (Franić, Dolan, Borsboom, Hudziak, et al., 2013). Because any given latent variable hypothesis implies a corresponding common pathway model, a refutation of that common pathway model constitutes evidence against the latent variable hypothesis.

For this reason, one may test the latent variable hypothesis by comparing the fit of a common pathway model to that of a corresponding independent pathway model. Specifically, if a model in which all of the A, C, and E factors exert direct influence on the phenotype fits the data statistically better than a model in which these influences are mediated by a phenotypic latent variable, this would provide evidence against the hypothesis that the effects on the observed item covariation are completely mediated by the phenotypic latent variable. In that case the latent factors employed in the phenotypic factor

[20] This is the standard application of the factor model to data collected in unrelated subjects, or when no information is available on genetic relatedness. We note, however, that if genome-wide DNA marker data are available in unrelated subjects, these could be used in a GTCA-like approach (Yang et al., 2011) to explore genetic covariance structures.

model are no more than an amalgamation of the direct influences of the A, C, and E factors on the observed item responses. If, on the other hand, an independent pathway model does not fit the data better than the corresponding common pathway model, this would provide support for the structure employed in the common pathway model, and substantiation for the corresponding phenotypic latent variable model. Comparison of an independent pathway model and a common pathway model may be conducted using a likelihood ratio test, because, as shown above, a common pathway model can be derived from an independent pathway model by imposing appropriate proportionality constraints on the factor loadings (i.e., the models are nested).

Analyses

In the present study, the outlined methodology was used to examine the substantive interpretability of the Internalizing syndrome dimensions of the CBCL (Anxious/Depressed, Withdrawn, Somatic Complaints). The phenotypic dimensionality of the 31 items of the Internalizing grouping was assessed using exploratory (EFA) and confirmatory (CFA) factor analysis. In this part of the analyses, the data were treated as if the sample consisted of genetically unrelated individuals. As treating observations from the same family as independent may result in biased test statistics, we performed a correction for clustering available in MPlus, which has been shown to work well in this context (Rebollo, de Moor, Dolan, & Boomsma, 2006). The EFA was performed using the oblique geomin rotation. Split-half validation was used, i.e. the EFA was performed on one randomly selected half of the sample (N=8756), and CFA on the other (N=8755).

Based on the results of the phenotypic dimensionality assessment, a common pathway model was formulated: in this model, the phenotypic factors obtained in the EFA and the CFA were retained, and their variation decomposed into A, C, and E components, as illustrated in the top panel of Figure 4. Subsequently, an independent pathway model was specified. This model is equal to the common pathway model in the number of the latent A, C, and E factors (i.e., the dimensions of the $\mathbf{\Psi}_A$, $\mathbf{\Psi}_C$, and $\mathbf{\Psi}_E$ matrices are equal across the two models), but it disposes of the phenotypic factors, i.e., it allows for the items to load directly on the A, C, and E factors. By comparing the fit of the common and the independent pathway model, we address the first focal question of whether one can interpret the syndrome dimensions of the CBCL Internalizing grouping substantively and causally.

To address the second research question, namely one concerning the dimensionality and the factor structure of the genetic and environmental effects that underlie the observed symptom covariation, independent pathway modeling was employed in an exploratory manner. First, the covariance matrix of the 31 symptoms was decomposed into A, C, and E components, i.e., the unconstrained 31x31 $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ matrices were estimated. Subsequently, EFA was applied to each of these matrices to obtain an indication of their dimensionality. Examining the dimensionality and the factor structure of the genetic and environmental effects which jointly act to produce the observed symptom structure provides insight into the observed symptom covariation, as the structure emerging in the phenotypic analyses (EFA, CFA) depends directly on the structure and relative magnitude of the underlying genetic and environmental components; for instance, a strongly prevailing unidimensional C component will make the phenotypic structure appear unidimensional.

The analyses were performed using Mplus (L. K. Muthén & Muthén, 2007), Mx (M. C. Neale, 2000), and R (R Core Team, 2013R). Given the discrete nature of the items, we fitted discrete factor models (i.e., we assumed the discrete indicator variables to be a

realization of a continuous normal[21] latent process, and modeled polychoric correlations; Flora & Curran, 2004; Wirth & Edwards, 2007) using the robust weighted least squares estimator (WLSMV; L. K. Muthén & Muthén, 1998-2007). The polychoric correlations between the 31 items and between the 62 (31 per twin) items served as input in the phenotypic and the genetic factor analyses, respectively. In evaluating model fit, the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA) were employed. As both the sample size and the models employed were large, the chi-square statistic was of limited use as an overall fit measure (Jöreskog, 1993), and was used only to test local hypotheses concerning comparisons of nested models, as these comparisons are associated with a smaller approximation error.

Results

Phenotypic analyses (EFA and CFA). The results of phenotypic dimensionality assessment are presented in Figure 3 and Table 1. EFA produced two well-fitting solutions: a 3- and a 4-factor solution (Figure 3). Interestingly, in both solutions, the items of the Anxious/Depressed scale appear to cluster into those pertaining to anxiety ('Fears doing something bad', 'Must be perfect', ' Nervous, tense', 'Fearful, anxious', 'Feels too guilty', 'Self-conscious') and those pertaining to depression ('Lonely', 'Cries a lot', 'Feels unloved', 'Others out to get him', 'Feels worthless', 'Suspicious', 'Sad', 'Worries'). In contrast to the Anxious/Depressed scale, the Somatic Complaints scale displayed a clearly unidimensional structure. The same is true of the Withdrawn scale, with the exception of the item 'Sulks', which consistently clustered with the items pertaining to depression, and the item 'Shy, timid', which in the 4-factor solution cross-loaded highly on the 'Anxious' factor.

The 4-factor solution, in which anxiety and depression form separate clusters, and the standard CBCL cross-informant model containing the Anxious/Depressed, Withdrawn, and Somatic Complaints scales, were subsequently tested in CFA. The models and the fit measures are shown in Figure 3. As can be seen from the Figure, the two models differed only minimally in terms of model fit: CFI = .877 vs. .891, TLI = .944 vs. .950, RMSEA = .037 vs. .035 for the 3- vs. the 4-factor models, respectively. In the light of the well-established difficulty in distinguishing phenotypically the dimensions of anxiety and depression, this finding is perhaps not entirely unexpected.

Genetic covariance structure modeling. Based on the results of the phenotypic dimensionality assessment, a 3- and a 4-factor common pathway model were formulated. These are depicted in the top panel of Figure 4. In both models, the common factors obtained in the phenotypic analyses (Anxious/Depressed, Withdrawn, and Somatic Complaints for the 3-factor model, and Anxious, Depressed, Withdrawn, and Somatic Complaints for the 4-factor model) were retained, and the contributions of the A, C, and E factors to their variation were assessed. As can be seen in Figure 4, the fit of the two common pathway models was virtually indistinguishable: CFI = .947 vs. .952, TLI = .962 vs. .966, RMSEA = .028 vs. .026 for the 3- vs. the 4-factor model, respectively.

Subsequently, based on the two common pathway models, the two independent pathway models depicted in the lower panel of Figure 4 were formulated. In these models, the A, C, and E factors employed in the common pathway analyses were retained, but the psychometric factors were disposed of, i.e. the items were allowed to load directly on the A, C, and E factors. Again, the fit of the two independent pathway models was virtually

[21] Tests of departures from underlying bivariate normality indicated that the normality assumption was tenable for all items.

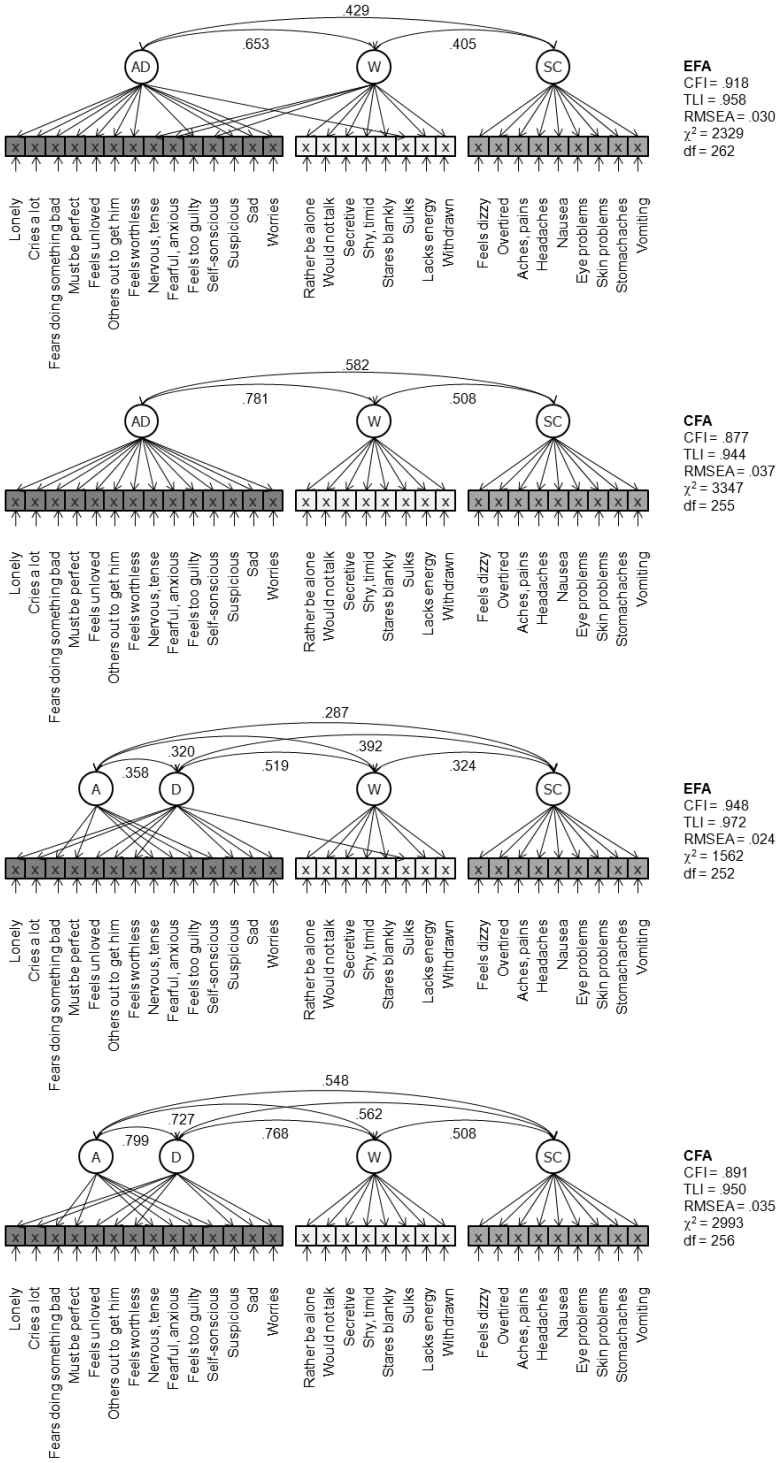
.429
.653
.405
AD
W
SC
EFA
CFI = .918
TLI = .958
RMSEA = .030
χ² = 2329
df = 262
Lonely
Cries a lot
Fears doing something bad
Must be perfect
Feels unloved
Others out to get him
Feels worthless
Nervous, tense
Fearful, anxious
Feels too guilty
Self-sonscious
Suspicious
Sad
Worries
Rather be alone
Would not talk
Secretive
Shy, timid
Stares blankly
Sulks
Lacks energy
Withdrawn
Feels dizzy
Overtired
Aches, pains
Headaches
Nausea
Eye problems
Skin problems
Stomachaches
Vomiting
.582
.781
.508
AD
W
SC
CFA
CFI = .877
TLI = .944
RMSEA = .037
χ² = 3347
df = 255
Lonely
Cries a lot
Fears doing something bad
Must be perfect
Feels unloved
Others out to get him
Feels worthless
Nervous, tense
Fearful, anxious
Feels too guilty
Self-sonscious
Suspicious
Sad
Worries
Rather be alone
Would not talk
Secretive
Shy, timid
Stares blankly
Sulks
Lacks energy
Withdrawn
Feels dizzy
Overtired
Aches, pains
Headaches
Nausea
Eye problems
Skin problems
Stomachaches
Vomiting
.287
.320
.392
.358
.519
.324
A
D
W
SC
EFA
CFI = .948
TLI = .972
RMSEA = .024
χ² = 1562
df = 252
Lonely
Cries a lot
Fears doing something bad
Must be perfect
Feels unloved
Others out to get him
Feels worthless
Nervous, tense
Fearful, anxious
Feels too guilty
Self-sonscious
Suspicious
Sad
Worries
Rather be alone
Would not talk
Secretive
Shy, timid
Stares blankly
Sulks
Lacks energy
Withdrawn
Feels dizzy
Overtired
Aches, pains
Headaches
Nausea
Eye problems
Skin problems
Stomachaches
Vomiting
.548
.727
.562
.799
.768
.508
A
D
W
SC
CFA
CFI = .891
TLI = .950
RMSEA = .035
χ² = 2993
df = 256
Lonely
Cries a lot
Fears doing something bad
Must be perfect
Feels unloved
Others out to get him
Feels worthless
Nervous, tense
Fearful, anxious
Feels too guilty
Self-sonscious
Suspicious
Sad
Worries
Rather be alone
Would not talk
Secretive
Shy, timid
Stares blankly
Sulks
Lacks energy
Withdrawn
Feels dizzy
Overtired
Aches, pains
Headaches
Nausea
Eye problems
Skin problems
Stomachaches
Vomiting

Figure 3. Results of the phenotypic EFA and CFA. In EFA solutions only the highest factor loading for each item is depicted (the omitted factor loadings equal .057 on average; the depicted factor loadings equal .57 on average).

indistinguishable: CFI = .977 vs. .976, TLI = .982 vs. .982, RMSEA=.019 vs. .010 for the 3- vs. the 4-factor-based model, respectively.

Addressing the first focal question of whether an independent pathway model fits the data appreciably better than a common pathway model, we compared the general fit of the models, and carried out likelihood ratio tests of the proportionality constraints mentioned above[22]. These tests revealed both the 3- and the 4-factor-based independent pathway models to fit the data better than their common pathway versions (χ^2=1554, df=24, p<.0001 for the 3-factor-based models, χ^2=1084, df=21, p<.0001 for the 4-factor-based models). This implies that the common pathway models, in which phenotypic latent variables mediate all of the A, C and E influences, fail to convey entirely accurately the genetic and environmental effects on the items.

In the second set of analyses, EFA was employed to evaluate separately the dimensionality and the factor structures of the genetic and environmental influences that underlie the observed symptom covariation. Specifically, we evaluated the dimensionalities of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ covariance matrices. The results are shown in Figure 5. An inspection of the scree plots in the Figure clearly indicates a 1-dimensional C structure. The structures of A and E matrices remain, however, somewhat less clear. To explore the A and E structures further, the present EFA results were used as a basis for specifying a number of competing independent pathway models with varying dimensionalities of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ covariance matrices. An overview of these models, including the fit measures and inter-factor correlations, is given in Table 1 in Appendix 3. Overall, a comparison of the models suggested a model with 2A, 1C, and 4E factors as the best-fitting model with acceptable inter-factor correlations (CFI=.978, TLI=.983, RMSEA=.018). This model is depicted in Figure 6, and parameter estimates are given in Table 2. It should, however, be noted that the models did not differ considerably in terms of model fit; therefore the structure in Figure 6 need not necessarily be conclusive. What the present results do strongly suggest, however, is a unidimensional C structure, and multidimensional (but mutually differing) A and E structures. These structures may also be discerned in Figure 7, which gives a graphical representation of the $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$, and $\mathbf{\Sigma}_E$ covariance matrices (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012).

Finally, the results of variance component estimation are given in Table 2. Overall, around 50% of the variance in the CBCL Internalizing symptoms is explained by the common A, C, and E factors, the remaining half being due to residual (symptom-specific) factors. The overall symptom heritability (defined as the heritability due to both the common and the symptom-specific factors) is 50% on average. The mean proportions of the phenotypic variance explained by the C and E factors are 20% and 30%, respectively (last 3 columns Table 2). These proportions are relatively stable across all symptom clusters, with symptoms of depression being somewhat less heritable than the others (41% vs. 51%-65% on average). Interestingly, the high item heritability is predominantly due to the item-specific, rather than the common A factors, while the C component is primarily due to the common C factor, with the item-specific factors accounting for a negligible portion of the variance.

[22] For WLSMV estimators the standard approach of taking the difference between chi-square values and the corresponding degrees of freedom is not appropriate because the chi-square difference is not chi-square distributed (Muthén & Muthén, 1998-2007). We therefore performed chi-square difference testing using scaling correction factors (Satorra & Bentler, 2001).

Table 1
Standardized factor loadings obtained in the phenotypic EFA and CFA

Item	EFA (N=8756)							CFA (N=8755)						
	3-factor solution			4-factor solution				3-factor model			4-factor model			
	AD	W	SC	D	A	W	SC	AD	W	SC	D	A	W	SC
Lonely	.734	-.068	.001	.670	-.054	.090	.016	.680			.706			
Cries a lot	.438	.039	.103	.390	.008	.124	.111	.555			.578			
Fears doing bad	.393	.165	.049	.309	.471	-.050	.005	.580				.619		
Must be perfect	.314	.174	.062	.245	.453	-.042	.019	.499				.530		
Feels unloved	.988	-.268	-.055	.885	.008	-.103	-.061	.708			.736			
Others out to get him	.852	-.141	.001	.767	-.006	.013	.004	.695			.725			
Feels worthless	.819	.012	-.028	.720	.249	.006	-.052	.791			.821			
Nervous, tense	.277	.303	.142	.215	.363	.148	.122	.634				.678		
Fearful, anxious	.197	.426	.121	.139	.388	.250	.105	.650				.695		
Feels too guilty	.499	.183	.046	.411	.470	-.019	.004	.710				.759		
Self-conscious	.079	.622	.032	-.002	.569	.347	-.002	.620				.664		
Suspicious	.628	.032	.028	.556	.030	.139	.037	.640			.664			
Sad	.729	.027	.010	.653	.056	.126	.017	.768			.796			
Worries	.484	.191	.100	.406	.366	.061	.072	.740			.767			
Rather be alone	.141	.408	-.039	.139	-.048	.474	.006		.542				.541	
Would not talk	-.001	.675	-.031	.004	.026	.678	.027		.661				.662	
Secretive	-.019	.812	-.083	-.033	.090	.806	-.029		.743				.743	
Shy, timid	-.191	.795	-.017	-.228	.424	.562	-.019		.569				.572	
Stares blankly	.174	.457	.050	.180	-.096	.546	.112		.668				.667	
Sulks	.403	.065	.106	.356	.000	.154	.117		.547				.546	
Lacks energy	.134	.339	.050	.150	-.134	.441	.099		.567				.566	
Withdrawn	.141	.718	-.145	.138	.026	.740	-.082		.778				.778	
Feels dizzy	.205	-.011	.478	.193	-.001	.031	.476			.620				.620
Overtired	.216	.164	.318	.197	.035	.192	.329			.752				.753
Aches, pains	.026	.018	.726	.023	.038	.015	.720			.778				.778

Headaches	-.050	.024	.709	-.042	.017	.022	.703	.639	.638
Nausea	.022	-.089	.883	.032	-.046	-.043	.876	.741	.741
Eye problems	-.025	.077	.334	-.015	.001	.078	.337	.380	.380
Skin problems	-.017	.060	.280	-.023	.072	.023	.274	.302	.302
Stomachaches	.004	.020	.732	-.007	.120	-.036	.719	.722	.722
Vomiting	-.049	-.017	.724	-.040	-.009	-.004	.718	.609	.609
Factor determinacies	.963	.963	.963	.953	.873	.935	.946		

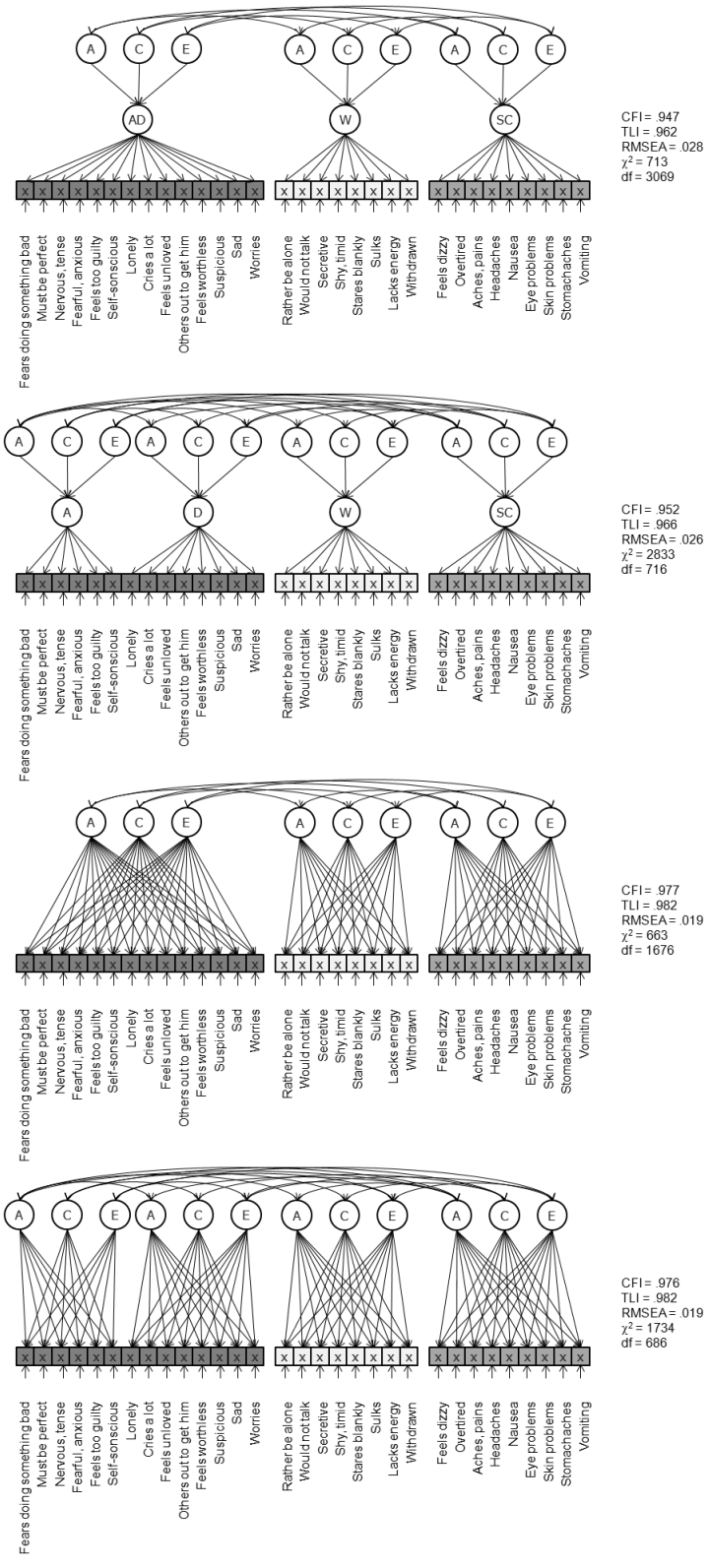
A
C
E
AD
W
SC
CFI = .947
TLI = .962
RMSEA = .028
χ² = 713
df = 3069
Fears doing something bad
Must be perfect
Nervous, tense
Fearful, anxious
Feels too guilty
Self-sonscious
Lonely
Cries a lot
Feels unloved
Others out to get him
Feels worthless
Suspicious
Sad
Worries
Rather be alone
Would not talk
Secretive
Shy, timid
Stares blankly
Sulks
Lacks energy
Withdrawn
Feels dizzy
Overtired
Aches, pains
Headaches
Nausea
Eye problems
Skin problems
Stomachaches
Vomiting
A
D
W
SC
CFI = .952
TLI = .966
RMSEA = .026
χ² = 2833
df = 716
CFI = .977
TLI = .982
RMSEA = .019
χ² = 663
df = 1676
CFI = .976
TLI = .982
RMSEA = .019
χ² = 1734
df = 686

Figure 4. The common (upper panel) and independent (lower panel) pathway models fitted to the data.

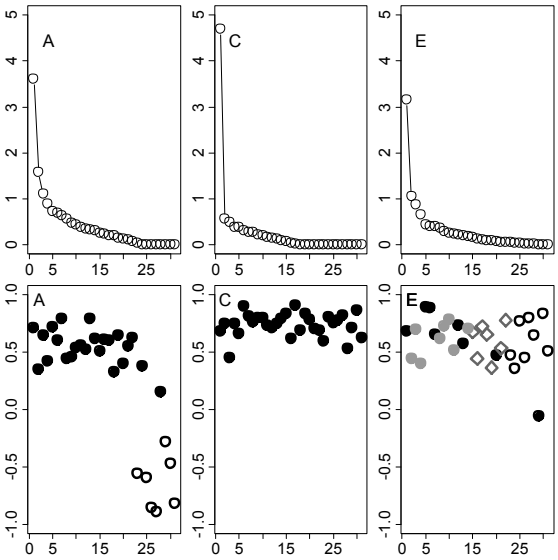

Figure 5. Eigenvalues of the $\mathbf{S}_A$, $\mathbf{S}_C$, and $\mathbf{S}_E$ matrices (upper panel) and factor loadings obtained in EFA solutions with 2 A, 1 C, and 4 E factors (lower panel). Colors/shapes code for different latent factors. Only the highest factor loading for each item is shown.

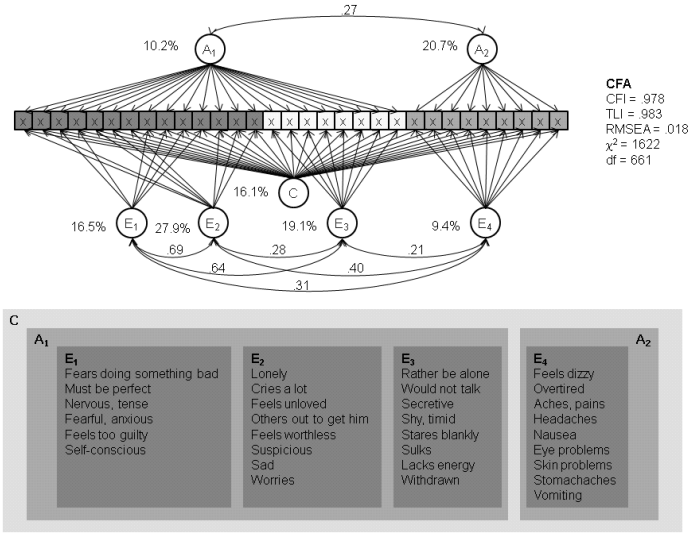

Figure 6. The 2A 1C 4E independent pathway model. Item residuals are not depicted but are included in the model. The mean percentages of item variance explained by each factor are given. Items are listed below, and their allocation to factors is indicated by the color of the panels.

Table 2

The 2A 1C 4E model: Proportions of item variance explained by the common A, C, and E factors (first three columns), by the item-specific A, C, and E factors (next three columns), by all common factors relevant to the item ($\lambda_{common}^2=\lambda_A^2+\lambda_C^2+\lambda_E^2$), by all residual factors relevant to the item ($\lambda_{residual}^2=\lambda_{resA}^2+\lambda_{resC}^2+\lambda_{resE}^2$), and by all the A, C, and E factors relevant to the item, respectively (Total $\lambda_A^2=\lambda_A^2+\lambda_{resA}^2$, etc). The $\overline{\lambda}^2$ rows give the mean proportion of item variance explained per item cluster (Depressed, Anxious, Withdrawn, and Somatic Complaints, respectively), and the Overall $\overline{\lambda}^2$ row gives the mean proportion of item variance explained across all items.

Item	λ_{A1}^2	λ_C^2	λ_{E2}^2	λ_{resA}^2	λ_{resC}^2	λ_{resE}^2	λ_{common}^2	$\lambda_{residual}^2$	Total l_A^2	Total l_C^2	Total l_E^2
Lonely	0.01	0.17	0.36	0.37	0.02	0.08	0.54	0.46	0.37	0.18	0.44
Cries a lot	0.02	0.23	0.06	0.56	0.00	0.13	0.30	0.70	0.58	0.23	0.19
Feels unloved	0.00	0.16	0.45	0.39	0.00	0.00	0.61	0.39	0.39	0.16	0.45
Others out to get him	0.01	0.24	0.34	0.38	0.00	0.04	0.58	0.42	0.39	0.24	0.38
Feels worthless	0.05	0.21	0.44	0.29	0.00	0.01	0.70	0.30	0.34	0.21	0.45
Suspicious	0.04	0.30	0.10	0.37	0.08	0.11	0.44	0.56	0.41	0.38	0.21
Sad	0.03	0.27	0.30	0.36	0.01	0.03	0.60	0.40	0.39	0.28	0.33
Worries	0.11	0.21	0.19	0.31	0.10	0.07	0.51	0.49	0.42	0.32	0.27
$\overline{\lambda^2}$	0.03	0.22	0.28	0.38	0.03	0.06	0.54	0.46	0.41	0.25	0.34
			λ_{E1}^2								
Fears doing something bad	0.08	0.10	0.21	0.36	0.26	0.00	0.39	0.62	0.44	0.36	0.21
Must be perfect	0.06	0.08	0.18	0.42	0.00	0.26	0.32	0.68	0.48	0.08	0.44
Nervous, tense	0.12	0.19	0.11	0.40	0.00	0.18	0.42	0.58	0.52	0.19	0.29
Fearful, anxious	0.18	0.16	0.12	0.36	0.00	0.18	0.46	0.54	0.53	0.16	0.30
Feels too guilty	0.08	0.22	0.30	0.35	0.06	0.00	0.59	0.41	0.43	0.28	0.29
Self-conscious	0.42	0.08	0.07	0.25	0.00	0.17	0.58	0.42	0.67	0.09	0.24
$\overline{\lambda^2}$	0.16	0.14	0.17	0.36	0.05	0.13	0.46	0.54	0.51	0.19	0.30

			λ^2_{E3}								
Rather be alone	0.04	0.09	0.22	0.37	0.00	0.28	0.35	0.65	0.41	0.09	0.50
Would not talk	0.14	0.15	0.15	0.40	0.16	0.00	0.44	0.56	0.54	0.31	0.15
Secretive	0.21	0.11	0.32	0.31	0.00	0.05	0.64	0.36	0.52	0.11	0.37
Shy, timid	0.43	0.03	0.04	0.33	0.00	0.17	0.50	0.50	0.77	0.03	0.20
Stares blankly	0.04	0.21	0.23	0.48	0.00	0.05	0.47	0.53	0.51	0.21	0.28
Sulks	0.03	0.25	0.02	0.49	0.00	0.21	0.30	0.70	0.52	0.25	0.23
Lacks energy	0.02	0.17	0.12	0.66	0.00	0.03	0.31	0.69	0.68	0.17	0.15
Withdrawn	0.14	0.09	0.43	0.37	0.00	0.00	0.65	0.37	0.51	0.09	0.43
$\overline{\lambda^2}$	0.13	0.14	0.19	0.43	0.02	0.10	0.46	0.54	0.56	0.16	0.29

	λ^2_{A2}		λ^2_{E4}								
Feels dizzy	0.07	0.22	0.08	0.25	0.04	0.34	0.37	0.63	0.32	0.26	0.43
Overtired	0.03	0.37	0.08	0.41	0.00	0.11	0.47	0.53	0.44	0.37	0.19
Aches, pains	0.22	0.15	0.24	0.30	0.09	0.00	0.61	0.39	0.52	0.23	0.24
Headaches	0.26	0.09	0.11	0.23	0.00	0.30	0.47	0.53	0.49	0.09	0.42
Nausea	0.46	0.11	0.10	0.09	0.13	0.10	0.68	0.32	0.55	0.25	0.20
Eye problems	0.03	0.04	0.10	0.70	0.00	0.14	0.17	0.83	0.73	0.04	0.23
Skin problems	0.06	0.05	0.00	0.64	0.00	0.24	0.11	0.89	0.70	0.05	0.24
Stomachaches	0.30	0.14	0.13	0.25	0.00	0.19	0.56	0.44	0.55	0.14	0.31
Vomiting	0.44	0.10	0.00	0.00	0.33	0.12	0.54	0.46	0.44	0.44	0.12
$\overline{\lambda^2}$	0.21	0.14	0.09	0.32	0.07	0.17	0.44	0.56	0.53	0.21	0.27

Overall $\overline{\lambda^2}$	0.13	0.16	0.18	0.37	0.04	0.12	0.47	0.53	0.50	0.20	0.30

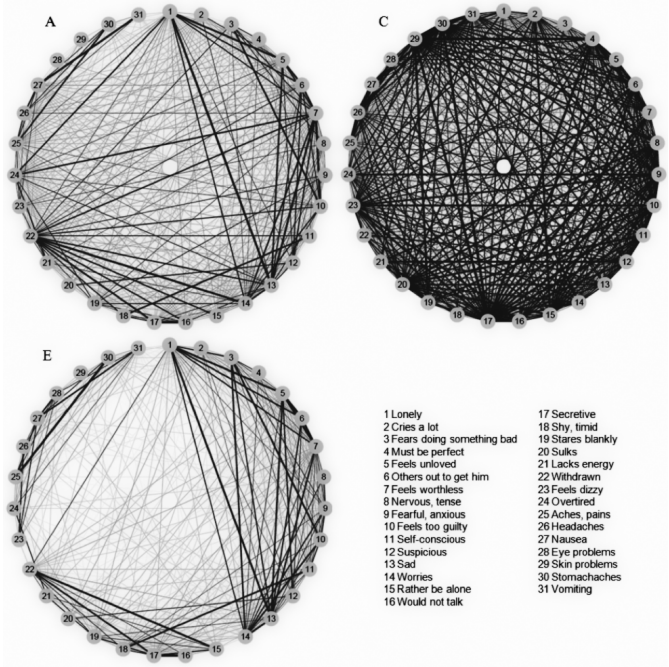

Figure 7. Graphical representation of the $\mathbf{\Sigma}_{\mathbf{A}}$, $\mathbf{\Sigma}_{\mathbf{C}}$, and $\mathbf{\Sigma}_{\mathbf{E}}$ matrices. Nodes (i.e., circles) represent symptoms. The thickness of the edges (i.e., of the lines connecting the nodes) represents the strength of correlations between the symptoms. For instance, the thickness of the line connecting Item 1 ("Lonely") to item 13 ("Sad") in the "A" graph represents the magnitude of the additive genetic correlation between these two symptoms.

Discussion

The present chapter aimed at answering two principal questions: one pertaining to the ontological nature of the syndrome dimensions postulated in the CBCL cross-informant model, and the other pertaining to the factor structure of the genetic and environmental influences that underlie the observed symptom covariation.

The first question relates to a longstanding discussion in philosophy of science. The latent variable model, arguably the predominant measurement model in psychology, invariably invokes a latent variable which is hypothesized to underlie a set of observed variables (i.e., item responses, symptoms). The ontological nature of such latent variables has long been a subject of debate. On the theoretical side of the debate, broadly speaking, two principal (and mutually opposing) accounts of the latent variable are commonly invoked. In the realist view, the latent variable represents a real entity which is assumed to exist independently of measurement, and is characterized by a causal relationship with its indicators: for instance, *because* a child is depressed, it exhibits symptoms such as excessive crying and feelings of sadness and worthlessness. The opposing, constructivist account, posits the latent variable as nothing more than a statistical construct used to simplify observations; in this view, this construct need not exist independently of measurement (Borsboom et al., 2003).

Empirical contributions to this debate have, to our knowledge, been scarce, although the existence and causal relevance of specific latent constructs such as depression and general intelligence have long been a source of controversy. Genetic factor modeling, as applied in the present chapter, may inform the discussion from an empirical perspective: by comparing the fit of a common pathway model, in which the latent phenotypic variables mediate all genetic and environmental effects on item covariation (the model therefore incorporating a realist hypothesis concerning the nature of those variables, or at least bring consistent with a realist perspective), to the fit of an independent pathway model (which bears no realist commitment regarding the phenotypic variable), one may test the latent variable hypothesis.

In the present case, neither the common pathway model including the three Internalizing syndrome dimensions of the CBCL (Anxious/Depressed, Withdrawn, and Somatic Complaints), nor the common pathway model postulating anxiety and depression as separate entities, survived confrontation with the independent pathway models. This invites reconsideration of the substantive interpretation of the dimensions in question, as it follows that these dimensions are better understood as a composite of unconstrained genetic and environmental influences than as well-defined entities that plausibly exist independently of measurement and statistical procedures (e.g., as natural kinds, Kendler, Zachar, & Craver, 2011).

This does not necessarily undermine the practical utility of the CBCL; we do not doubt its usefulness for diagnostic purposes, especially given that the broad structure found in our analyses is in line with the current item allocation of the CBCL. Furthermore, the reasons for rejecting the common pathway structure may be local (due only to a subset of observed variables) and therefore the violation may be accommodated by addition of parameters or by the removal of offending variables. What the present results do suggest, however, is that the three syndrome dimensions, as currently defined, do not appear to represent homogeneous entities in the Borsboom et al. (Borsboom et al., 2003) sense, but are rather an amalgam of several different genetic and environmental structures. Clearly, the ascription of causal forces to such amalgams is problematic.

The second research question pertains to the structure of the genetic and environmental influences that give rise to the observed symptom covariation. Interestingly, the results suggest mutually differing additive genetic, common environmental and unique environmental structures. The 2-dimensional additive genetic structure distinctly affects symptoms of anxiety, depression, and withdrawal, on the one hand, and somatic complaints, on the other. The 4-dimensional unique environmental structure affects each of these symptom clusters distinctly, while the common environment acts uniformly across the entire range of internalizing symptoms. This replicates the findings of previous multivariate investigations into the genetic and environmental sources of symptom covariation, which demonstrate a common genetic diathesis for anxiety and depression, with the distinction between these disorders being driven by the individual-specific environment (e.g., Kendler et al., 1987; Kendler et al., 1992; Middeldorp et al., 2005).

The present results put the aforementioned difficulties in delineating between the diagnostic categories of anxiety and depression into a clearer perspective. Anxiety and depression appear to share a common genetic basis: a single set of genes affects the individual differences in predisposition to developing general anxiety-, depression- and withdrawal-related symptomatology. Previous research and theoretical work have amply demonstrated a possibility of a general factor accounting for shared symptoms of anxiety, depression, and possibly more broad neurotic symptomatology (with more specific factors accounting for the specific subtypes of symptoms) (e.g., Clark & Watson, 1991). This general

factor can conceivably be identified with the shared genetic basis found in the present analyses. While this shared basis constitutes a broad genetic vulnerability which may predispose children to developing general internalizing symptomatology, the specific form of symptomatology (anxiety, depression or withdrawal) may depend on the children's unique environmental influences. The common family environment[23], interestingly, exerts an overall protective or predisposing effect on the entire set of internalizing symptoms, either increasing or lowering the chance of developing internalizing psychopathology across the board.

If one takes into account not only the structure, but also the relative magnitude of the A, C, and E influences found in the present analyses, an interesting picture emerges. Consider the set of items pertaining to anxious, depressed, and withdrawn behaviors. Under the model depicted in Figure 6, this item set is influenced by a unidimensional A and a unidimensional C structure. These unidimensional latent structures, which act to make the symptoms act alike (i.e., covary), collectively explain around a quarter of their total phenotypic variance (10.2% and 17% of the relevant item variance is explained by the A_1 and by the C factor, respectively). The factors which facilitate the clustering of these symptoms into three separate groups (in particular, the E_1, E_2, and E_3 factors) explain around 22% of their phenotypic variance. The remainder (~50%) of the phenotypic variance is explained by item-specific factors. Given the balance in the magnitude of influence that these mutually differing structures exert on the item set, the inability of the phenotypic modeling to distinguish between several different models is not surprising. In fact, one could wonder how the phenotypic analyses could converge on a single model, if several different models, each equally relevant to the phenotypic structure, are correct.

Finally, it should be mentioned that problems regarding the validity and reliability of children's self-reports and the consequent use of raters (parents, teachers) in the assessment of children's behavior may complicate assessment and subsequent interpretation. Rater bias (i.e., systematic effects on ratings originating from rater characteristics) is a widely recognized problem in research involving informants. In the context of twin and family studies, unmodeled rater bias is known to result in an overestimation of the shared environmental variance (M. C. Neale & Cardon, 1992). Previous studies on internalizing symptoms have demonstrated a modest to nonexistent role of shared environment in the development of anxiety (Gregory & Eley, 2007; Hettema, Neale, & Kendler, 2001; Legrand, McGue, & Iacono, 1999; Rapee et al., 2009), and a modest to moderate role of shared environment in the development of depression (Boomsma et al., 2005b; Rice, Harold, & Thapar, 2002b). Although this is consistent with the present findings, the extent to which our estimate of the shared environmental component is confounded by rater bias remains to be examined.

In summary, the present chapter utilized genetic item analyses to examine the ontology and the genetic and environmental etiology of the latent constructs 'Anxious/Depressed', 'Withdrawn', and 'Somatic Complaints', as defined in the CBCL/6-18 cross-informant model. The results 1) invite reconsideration of the substantive interpretation of these latent constructs, and 2) consistently with results of previous studies, demonstrate that additive genetics, common environment, and individual-unique environment each exert a distinct and mutually differing pattern of influence on internalizing symptoms. These results provide an informative context to the discussion on the phenotypic delineation between different syndromes or disorders, and contribute to our understanding of both the nature of the Internalizing syndrome dimensions and the etiology of internalizing behavior.

[23] See Carey 2009 for an alternative interpretation of C.

Chapter 5

The Big Five Personality Traits: Psychological Entities or Statistical Constructs?

Abstract

The present study employed multivariate genetic item-level analyses to examine the ontology and the genetic and environmental etiology of the Big Five personality dimensions, as measured by the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992; H. A. Hoekstra, Ormel, & De Fruyt, 1996). Common and independent pathway model comparison was used to test whether the five personality dimensions fully mediate the genetic and environmental effects on the items, as would be expected under the realist interpretation of the Big Five. In addition, the dimensionalities of the latent genetic and environmental structures were examined. Item scores of a population-based sample of 7900 adult twins (including 2805 complete twin pairs; 1528 MZ and 1277 DZ) on the Dutch version of the NEO Five Factor Inventory were analyzed. Although both the genetic and the environmental covariance components displayed a 5-factor structure, applications of common and independent pathway modeling indicate that they do not comply with the proportionality constraints entailed in the common pathway model. Implications for the substantive interpretation of the Big Five are discussed.

Based on: Franić, S., Borsboom, D., Dolan, C. V., & Boomsma, D. I. (2013) The Big Five Personality Traits: Psychological Entities or Statistical Constructs? *Behavior Genetics*, advance online publication, doi: 10.1007/s10519-013-9625-7, 1-14.

Over the past century, one of the most influential approaches to personality description has been the Five Factor (FF) approach. Predicated on the lexical approach to personality description, reflected in Cattell's (Cattell, 1943, p.483): "All aspects of human personality which are or have been of importance, interest, or utility have already become recorded in the substance of language", the FF approach is based on the idea that identification of basic dimensions of human personality is possible via the application of factor analytic techniques to verbal descriptors of human traits.

The beginnings of the FF approach can be traced to Allport & Odbert's (Allport & Odbert, 1936) selection of 4,504 psychological trait terms from the 1925 unabridged Webster's New International Dictionary. Cattell (Cattell, 1943, 1945) augmented this list in the 1940s by adding "the substance of all syndromes and types which psychologists have observed and described in the past century or so", and subsequently abbreviated it to a set of 35 variables - a factor analysis of which produced 12 "primary" factors. In the early 1960s, Tupes and Christal (Tupes & Christal, 1992) performed a series of factor analyses on Cattell's variables and observed five recurrent orthogonal factors, which they denoted Surgency/Extraversion, Agreeableness, Dependability, Emotional Stability, and Culture (French, 1953). Through Norman's (Norman, 1963, 1967) further addition to, and subsequent abbreviation of, Allport & Odbert's original list, and further selection of adjectives from this list by Goldberg (Goldberg, 1977, 1980, 1983, 1990, 1992), a set of variables with a clearer five-factor orthogonal structure was produced. Goldberg (Goldberg, 1980, 1982) denoted these five factors "the Big Five".

In a parallel research program, following a cluster analysis of Cattell's Sixteen Personality Factor (16PF) Questionnaire in which three factors were extracted – Neuroticism, Extraversion, and Openness to Experience – McCrae and Costa (McCrae & Costa, 1983) developed a 144-item, 18-facet, 3-dimensional questionnaire, which they termed the NEO Inventory. After linking their Neuroticism and Extraversion factors to those from the previous lexically based research (e.g., Goldberg, 1980, 1981, 1982, 1983), they fully adopted the FF approach, and consequently developed measures of Agreeableness and Conscientiousness. The addition of these scales to the NEO Inventory resulted in the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985), and the subsequent implementation of facets to measure these two new factors yielded the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992). The NEO Five Factor Inventory (NEO-FFI) is a shorter, 60-item version of the NEO-PI-R.

The FF approach has been extraordinarily influential: numerous behavior genetics studies have assessed the heritabilities of the Big Five (and more recently sought associations with measured genetic variants; de Moor et al., 2010), neural and clinical correlates of the five factors have been examined (e.g., DeYoung et al., 2010; Nigg et al., 2002), and the model has found wide practical application, for instance in the field of personnel selection (Schmit & Ryan, 1993). A Google Scholar search for "Five Factor model personality" returns nearly two million hits, and a Google search of the same term returns around 121 million.

Notwithstanding its popularity, however, a plethora of issues have been raised concerning the conceptual, empirical and statistical foundations of the FF approach (e.g., J. Block, 1995). Lack of formal theory underpinning the approach and the possibility of empirical analyses being shaped by prior conceptual commitments are some of the most prominent ones. Concerns have been raised over the orthogonality of the factor solutions, their proposed simple structure, and even the number of factors being significantly impacted by the selection of input variables and choices of factor rotations, which ultimately might have rested more on the authors' conceptual beliefs than on mathematical/statistical

criteria. In addition, the degree of arbitrariness involved in Cattell, Norman, and Goldberg's selection of trait terms and construction of clusters remains unknown. The model has received additional criticism for failing to account for intra-individual personality structure and personality functioning. A factor analysis of common English terms describing laptop computers, for instance, might yield size, processing speed, random-access memory capacity, storage capacity, and operating system as five orthogonal factors; however, one may wonder to what extent these factors are informative about the physical structure of a laptop computer, or its functional architecture (Cervone, 2005). The model has also been criticized on psychometric grounds for a number of problems including failure of orthogonality (J. H. Block & Block, 1980; Costa & McCrae, 1992; Goldberg, 1992; Mroczek, 1992; Peabody & Goldberg, 1989), the presence of cross-loadings (J. Block, 1995; Costa & McCrae, 2008; Parker, Bagby, & Summerfeldt, 1993), low validity coefficients (Pervin, 1994), lack of reproducibility of the five-factor structure from other personality inventories (Caprara, Barbaranelli, & Comrey, 1995; Hahn & Comrey, 1994), and lack of fit in confirmatory context (McCrae, Zonderman, Costa, Bond, & Paurnonen, 1996; Parker et al., 1993). The FF model is derived through, and based on, exploratory techniques such as exploratory factor analysis (EFA) and principal components analysis (PCA); in the confirmatory factor analysis (CFA) context, however, the model typically obtains unsatisfactory fit.

Another, arguably more fundamental issue, concerns a possible misinterpretation of principal components (Markus & Borsboom, 2013) and, more broadly, the ontological nature of the five factors. Being generated in a formative model, the components obtained in PCA are efficient statistical summaries of the data. Their standard interpretation amongst FF model proponents, however, is of a realist nature; they are considered to be behavior-generating entities (e.g., extraversion causes party-going behavior; McCrae & Costa, 2008). This possible misinterpretation of principal components, along with some of the other criticism listed above, has prompted questions about whether the Big Five factors are truly a discovery, as advocated by its proponents, or should rather be seen as a set of statistical constructs emanating from factor analysis of possibly preselected sets of variables.

In the present chapter, we address the last issue using quantitative genetic methodology. As outlined in Chapter 3 (see also Franić, Dolan, Borsboom, Hudziak, et al., 2013), quantitative genetic methods can be used to test hypotheses regarding the ontological nature of latent variables. In particular, we address the question of whether the realist interpretation of the Big Five personality factors (in which the factors represent entities causing the observed item responses) is supported by the data, or whether the factors would more correctly be interpreted as statistical constructs. To this end, we examine the dimensionality of the latent genetic and environmental structures underlying the observed covariation in NEO-FFI items. Behavior genetic studies have been performed on personality data before (e.g., Bouchard Jr & Loehlin, 2001; Loehlin, 1989; Loehlin & Martin, 2001; Plomin & Caspi, 1990), but item-level analyses, which enable us to address the specific research question above, have seldom been undertaken on NEO-FFI or NEO-PI-R data (Johnson & Krueger, 2004).

Method

Data

The data were obtained from the Netherlands Twin Register at VU University Amsterdam (Willemsen et al., 2013), and consist of item scores of a population-based sample of 7900

adult twins (including 2805 complete twin pairs; 1528 MZ and 1277 DZ) on the Dutch version of the NEO-FFI (Costa & McCrae, 1992; H. A. Hoekstra et al., 1996). The participants were aged between 18 and 86 (M = 32.3, SD = 12.7) at time of measurement. 68.3% of the participants were female. The NEO-FFI is a 60-item personality questionnaire consisting of 5 subscales: Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C). Item content is given in Table 1. The responses are given on a 5-point scale ('strongly disagree', 'disagree', 'neutral', 'agree', 'strongly agree').

Initially, the sample consisted of 8090 twins, and missingness was limited to .9%. In treating missingness, we adopted the guidelines outlined in the NEO-FFI manual (Costa & McCrae, 1992; H. A. Hoekstra, et al., 1996): if missingness per participant exceeded 15%, the participant's scores were excluded from the analyses. The application of this criterion reduced the missingness to .4%, and the sample size to N=7900. The remaining missing values were assigned the 'neutral' value of 3. Application of LISREL's (Jöreskog & Sörbom, 2006) test for underlying bivariate normality indicated no significant departures from normality for any of the items. The MZ and DZ twin item correlations are depicted in Figure 2.

Table 1

Item content of the NEO-FFI. Item numbering in the parentheses corresponds to that in the text/Tables/Figures. Reverse-scored items are marked with an asterisk

Item no.		Item content	Scale
1	(n1)	Not a worrier *	Neuroticism
6	(n2)	Feels inferior	
11	(n3)	Goes to pieces under stress	
16	(n4)	Rarely lonely or blue *	
21	(n5)	Tense, jittery	
26	(n6)	Feels worthless	
31	(n7)	Rarely fearful or anxious *	
36	(n8)	Angry at the way people treat him	
41	(n9)	Easily discouraged	
46	(n10)	Seldom sad or depressed *	
51	(n11)	Feels helpless	
56	(n12)	Ashamed	
2	(e1)	Likes having many people around	Extraversion
7	(e2)	Laughs easily	
12	(e3)	Not cheerful or light-hearted *	
17	(e4)	Enjoys talking to people	
22	(e5)	Likes to be where the action is	
27	(e6)	Prefers to do things alone *	
32	(e7)	Bursting with energy	
37	(e8)	Cheerful, vivacious	
42	(e9)	Not a cheerful optimist *	
47	(e10)	Leads a fast-paced life	
52	(e11)	Very active	
57	(e12)	Rather go his own way than lead others *	

3	(o1)	Doesn't waste time daydreaming *	Openness
8	(o2)	Sticks to a single way of doing things *	
13	(o3)	Intrigued by patterns	
18	(o4)	Thinks controversial speakers only confuse students *	
23	(o5)	Not affected by poetry *	
28	(o6)	Tries new foods	
33	(o7)	Doesn't notice moods different environments produce *	
38	(o8)	Looks to religious authorities for moral decisions *	
43	(o9)	Excited by poetry or art	
48	(o10)	Little interest in speculating about nature of universe *	
53	(o11)	Wide range of intellectual interests	
58	(o12)	Enjoys playing with theories	
4	(a1)	Courteous	Agreeableness
9	(a2)	Often gets into arguments *	
14	(a3)	Some consider him selfish or egotistical *	
19	(a4)	Prefers cooperation to competition	
24	(a5)	Cynical, skeptical *	
29	(a6)	Thinks people will take advantage *	
34	(a7)	Most people like him	
39	(a8)	Some consider him cold or calculating *	
44	(a9)	Business-like, unsentimental *	
49	(a10)	Thoughtful, considerate	
54	(a11)	Shows if he doesn't like people *	
59	(a12)	Prepared to manipulate *	
5	(c1)	Keeps belongings neat and clean	Conscientiousness
10	(c2)	Good at pacing himself	
15	(c3)	Not very methodical *	
20	(c4)	Performs tasks conscientiously	
25	(c5)	Has a clear set of goals	
30	(c6)	Wastes time before settling down to work *	
35	(c7)	Works hard	
40	(c8)	Follows through on commitments	
45	(c9)	Not dependable *	
50	(c10)	Productive	
55	(c11)	Unable to get organized *	
60	(c12)	Strives for excellence	

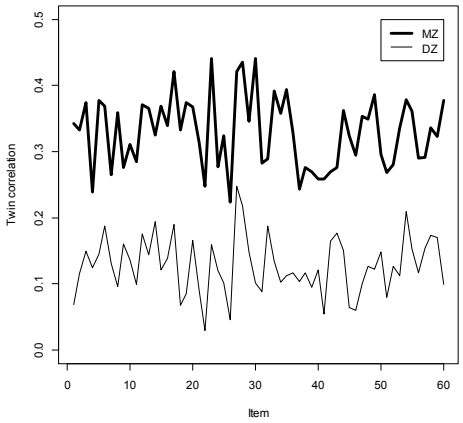

Figure 2. Phenotypic MZ and DZ polychoric twin item correlations.

The analytic framework employed in the present study has been outlined in Chapters 3 and 4, but is repeated below (*Approach* section) for completeness. The reader familiar with Chapters 3 or 4 may skip to the *Analyses* section.

Approach

Genetic covariance structure modeling (Martin & Eaves, 1977) is the application of structural equation modeling (Bollen, 1989; Kline, 2005) to data collected in genetically informative samples, such as samples of twins (Franić et al., 2012; M. C. Neale & Cardon, 1992). In the classical twin design, the sample consists of monozygotic (MZ) and dizygotic (DZ) twin pairs. DZ twins share an average of 50% of their segregating genes, while MZ twins share nearly their entire genome (Falconer & Mackay, 1996; J. P. van Dongen et al., 2012). The covariance structure of the phenotypes (i.e., observed traits) is modeled as a function of latent factors representing several sources of individual differences: additive genetic (A), non-additive genetic (D), shared environmental (C), and individual-specific environmental (E) sources. Additive genetic influences are modeled by one or more A factors, which represent the total additive effects of genes relevant to the phenotype. Non-additive genetic influences are modeled by one or more D factors, representing the total non-additive effects of genes relevant to the phenotype. Non-additive effects arise from interactions of alleles within the same locus (genetic dominance) and/or across different loci (epistasis). Based on quantitative genetic theory (Falconer & Mackay, 1996; Jinks & Fulker, 1970), the A factors are known to correlate 1 across MZ twins and .5 across DZ twins, and D factors are known to correlate 1 across MZ twins and .25 across DZ twins. Environmental influences affecting the phenotype of both twins in an identical way, thereby increasing their similarity beyond what is expected based on genetic resemblance alone, are represented by one or more C factors. Therefore, by definition, the C factors correlate unity across twins (regardless of zygosity). All environmental influences causing the phenotype of two family members to differ are represented by one or more E factors. Thus, by definition, the E factors are uncorrelated across twins.

The classical twin design does not allow for simultaneous estimation of A, C and D effects (Keller & Coventry, 2005); two of these sources of individual differences can be

modeled at most.[24] Assuming, for instance, an ADE model, the expected covariance structure in a multivariate twin model is:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}_E & r_A\boldsymbol{\Sigma}_A + r_D\boldsymbol{\Sigma}_D \\ (r_A\boldsymbol{\Sigma}_A + r_D\boldsymbol{\Sigma}_D)^t & \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}_E \end{bmatrix}$$

where, given p phenotypes, $\boldsymbol{\Sigma}_{11}$ ($\boldsymbol{\Sigma}_{22}$) is the p x p covariance matrix of twin 1 (twin 2), $\boldsymbol{\Sigma}_{12}$ ($\boldsymbol{\Sigma}_{21}$) is the twin 1 - twin 2 p x p covariance matrix, and $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_D$ and $\boldsymbol{\Sigma}_E$ are the additive genetic, non-additive genetic, and unique environmental p x p covariance matrices, respectively. The coefficients r_A and r_D are the additive and the non-additive genetic twin correlations, respectively (MZ: $r_A = r_D = 1$; DZ: $r_A = 1/2$, $r_D = 1/4$).

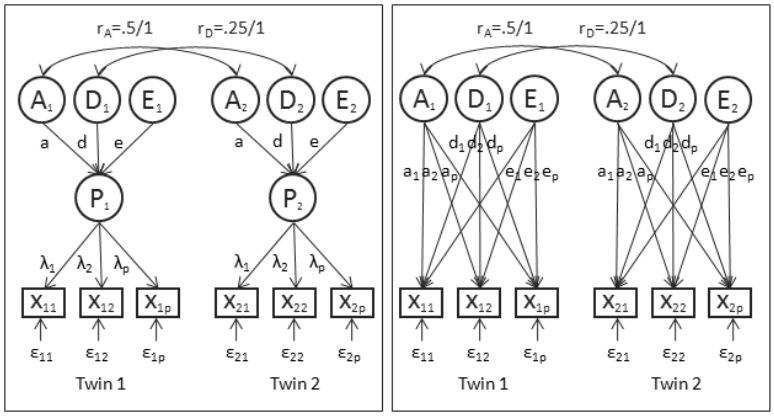

Figure 1. A common (left) and an independent (right) pathway model.

Figure 1 gives two examples of the multivariate twin models used in the present study. The first model in Figure 1 is a *common pathway model* (Kendler et al., 1987), also known as the psychometric factor model (McArdle & Goldsmith, 1990). In a common pathway model, all of the A,C(D), and E influences on the item responses are mediated by a latent variable, also referred to as the psychometric factor (factors P_1 and P_2 in Figure 1). P_1 and P_2 may be viewed as latent factors obtained in standard psychological research, e.g. 'neuroticism' or '*g*'. The second model in Figure 1 is an *independent pathway model* (Kendler et al., 1987), also known as the biometric factor model (McArdle & Goldsmith, 1990). In the independent pathway model, there is no phenotypic latent variable that mediates genetic and environmental effects on the item responses. Rather, the A, C(D), and E factors influence item responses directly.

In the present text, we distinguish between genetic factor models (introduced above), and phenotypic factor models. By 'phenotypic factor model', we refer to the factor model as usually formulated and applied in psychological research. The term 'phenotypic' is used to indicate the model as applied to observed (i.e., phenotypic) covariation; no genetic information is used. The 8-factor cross-informant model of the CBCL (Achenbach 1991) and the FF model of personality (McCrae & Costa, 2003; McCrae & John, 1992) are examples of a phenotypic factor model.

[24] Other designs, e.g., the nuclear twin family design, the stealth design, or the cascade design permit simultaneous estimation of A, C, D and E effects (Keller, Medland, & Duncan, 2010).

The common pathway model bears a number of similarities to the phenotypic factor model. Notably, both the phenotypic factor model and the common pathway model are based on the premise that all covariation in item responses is attributable to one or more latent variables. In phenotypic factor modeling, this hypothesis can be formulated in terms of measurement invariance: all external variables that produce covariation in item responses exert their influence via the latent variable (Mellenbergh, 1989; Meredith, 1993). Likewise, in common pathway modeling, one assumes that all of the A, C(D), and E effects on item covariation are mediated by the psychometric factor. That is, there are no direct effects of A, C(D), and E on the items.

The assumption of full mediation of external influences by the latent phenotypic variable(s) has strong implications. For instance, different external variables affecting a set of item responses via the same latent variable exert the same magnitude of influence relative to each other on all the items that depend on that latent variable. For instance, if an A and an E variable affect a set of items via the same psychometric factor, then the magnitude of influence exerted by the variable A on any individual item will be a scalar multiple of the magnitude of influence exerted by the variable E on the same item, and this scalar multiple (k) will be a constant across all the items depending on the same psychometric factor. This means that one can derive a common pathway model from an independent pathway model by imposing proportionality constraints on the factor loadings, such that $a_1/a_2 = d_1/d_2 = e_1/e_2 = k$ (following the notation in the right panel of Figure 1).

Thus, the common pathway model makes explicit an assumption of the phenotypic latent variable model concerning the sources of item covariation: all influences on item covariation are mediated by the phenotypic latent variable. Barring exceptional cases of model equivalence, this means that a latent variable model cannot hold unless the corresponding common pathway model holds (Franić, Dolan, Borsboom, Hudziak, et al., 2013). Because any given latent variable hypothesis implies a corresponding common pathway model, a refutation of that common pathway model would constitute evidence against the latent variable hypothesis.

For this reason, one may test the latent variable hypothesis by comparing the fit of a common pathway model to that of a corresponding independent pathway model. Specifically, if a model in which all of the A, C(D), and E factors exert direct influence on the phenotype fits the data statistically better than a model in which these influences are mediated by a phenotypic latent variable, this provides evidence against the hypothesis that the effects on the observed item covariation are completely mediated by the phenotypic latent variable. In that case, the latent factors employed in the phenotypic factor model may be no more than an amalgamation of the direct influences of the A, C(D), and E factors on the observed item responses. If, on the other hand, an independent pathway model does not fit the data better than the corresponding common pathway model, this would provide support for the structure employed in the common pathway model, and substantiation for the corresponding phenotypic latent variable hypothesis. Comparison of an independent pathway model and a common pathway model may be conducted using a likelihood ratio test, because, as shown, a common pathway model can be derived from an independent pathway model by imposing appropriate proportionality constraints on the factor loadings (i.e., the models are nested).

Analyses

In the first phase of the analyses, the phenotypic structure of the NEO-FFI was examined using exploratory (EFA) and confirmatory (CFA) factor analysis. Here, the data were treated

as if the sample consisted of unrelated individuals. To correct for the clustering in the data due to the genetic relatedness, we employed a correction for clustering available in MPlus (L. K. Muthén & Muthén, 1998-2007; Rebollo et al., 2006). EFA and CFA were performed using split-half validation: EFA was performed on one randomly selected half of the sample (N=3950), and CFA on the other (N=3950). In EFA, 3-6 factor solutions with the oblique geomin rotation were tested. We opted for an oblique criterion because the NEO-PI-R and NEO-FFI data conform appreciably better to a model with oblique factors, despite the initial idea of orthogonality (Goldberg, 1993; Mroczek, 1992; Peabody & Goldberg, 1989). The best-fitting substantively interpretable model indicated by EFA was subsequently tested in CFA.

In the second phase of the analyses, the results of the phenotypic analyses were used as a basis for specifying multivariate common and independent pathway genetic factor models. Here, only data on complete twin pairs (1528 MZ and 1277 DZ twin pairs) were used. The genetic and environmental etiology of the items was first examined using univariate modeling: a number of competing models (ACE, ADE, AE) were fitted to each of the 60 items, and likelihood ratio testing was employed to determine the best model for each item. The same approach was used on subscale level: univariate (ACE, ADE, AE) models were fitted to each of the five subscales. The results of these preliminary analyses were subsequently used as a basis for specifying multivariate common and independent pathway models.[25]

To address the central question concerning the ontological nature of the latent personality factors, the common and independent pathway models were compared against each other using likelihood ratio testing. Finally, to explore the structure of the genetic and environmental influences on the NEO-FFI items in a hypothesis-free fashion, the 60x60 phenotypic covariance matrix was decomposed into 60x60 genetic and environmental correlation matrices, and each of these matrices was subjected to EFA. The genetic and environmental correlations matrices were obtained in a standard twin model using Cholesky decompositions in Mx (Neale 2000). We used the phenotypic 120x120 (60 per twin) polychoric correlation matrix as input, because Pearson product moment correlations based on discrete data tend to be slightly biased (Dolan, 1994).

The analyses were carried out using Mplus 5 (L. K. Muthén & Muthén, 1998-2007), Mx, and R (R Core Team, 2013). Given the discrete nature of the items, we fitted discrete factor models (i.e., we assumed the discrete indicator variables to be a realization of a continuous normal latent process, and modeled polychoric correlations; Flora & Curran, 2004) using the robust weighted least squares estimator (WLSMV; L. K. Muthén & Muthén, 1998-2007). The polychoric correlations between the 60 items and between the 120 (60 per twin) items served as input in the phenotypic and the genetic factor analyses, respectively. In evaluating model fit, the Tucker Lewis Index (TLI)[26] and the Root Mean Square Error of Approximation (RMSEA) were used. Cut-off values of >.90 (TLI) and <.08 (RMSEA) were employed as criteria for acceptable fit. As both our sample size and the models employed were large, the chi-square statistic was of limited use as an overall fit measure (Jöreskog, 1993), and was employed only to test local hypotheses concerning comparisons of nested models, as these comparisons are associated with a smaller approximation error.

[25] Although item-specific residual factors can be subjected to their own AC(D)E decomposition, in the present chapter this was not done given our focus on dimensionality assessment and the common/independent pathway model comparison. The covariances among the residuals between the twins were however added. These covariances were estimated separately in the MZs and DZs, given the possible genetic residual effects.

[26] TLI is an incremental fit index based on the difference in fit of a baseline model with uncorrelated variables and the fitted model. The standard rule of thumb was formulated for the analyses of scale scores, not item score. As item scores tend to correlate to a lesser extent than scale scores (often based on multiple items), the standard TLI rule of thumb is hard to satisfy. See e.g. Kenny (2012).

Results

The results of the phenotypic EFA are given in Tables 2 and 3. As evident from Table 2, the 5- and the 6-factor phenotypic solution both fitted adequately (TLI>.94, RMSEA<.055). However, as the 6-factor solution was difficult to interpret substantively, in further analyses we focused on the 5-factor solution. This solution, detailed in Table 3, resembles closely Costa and McCrae's (Costa & McCrae, 1992) FF model. Thus, the basic structure of the FF model replicated well in our sample.

Based on the EFA results, a 5-factor model (corresponding exactly to Costa and McCrae's (Costa & McCrae, 1992) FF model) was formulated and tested in CFA. The fit measures, given in Table 2, indicated a suboptimal fit. This is not unexpected considering the literature, which frequently reports a misfit of the FF model to empirical data (e.g., McCrae et al., 1996; Parker et al., 1993). To examine the extent to which the misfit is due to presence of cross-loadings, in the next step we freed all the cross-loadings with a modification index (MI) larger than 50, and re-fitted the model. This resulted in an acceptable model fit (χ2= 9708, df=499, TLI=.899, RMSEA=.068). However, the modified model contained 94 cross-loadings.

Table 4 shows the factor loadings, residual variances and inter-factor correlations associated with the simple structure 5-factor model. The average variance explained by the factors ranges from 22% (O and A factors) to 42% (N factor). The factor correlations between Openness and the other factors are generally low (r<.12). The correlations between Neuroticism and the other factors are substantial and negative (from -.41, to -.62), and the remaining factors (Extraversion, Agreeableness, and Conscientiousness) are substantially and positively intercorrelated; from .45 to .48. This is line with the literature, which frequently reports substantial correlations between the five factors (e.g., J. Block, 1995).

In the first step of the genotypic analyses, the genetic and environmental etiology of the items was examined in a univariate fashion. The same was done on the subscale level, with the subscale scores being defined as the sum scores across the relevant items. Overall, none of the items or scales contained a detectable C component. With regard to the A, D, and E influences, the items displayed two major patterns: some appeared additive genetic and unique environmental in origin (AE model), while for the rest neither additive nor non-additive genetic influences could be detected (E model). On subscale level, only the Agreeableness scale displayed a significant D component, and the remaining scales conformed to an AE model. As another set of our preliminary analyses showed that the D component did not exceed 5% for any of the items (M=2.1%, SD=1%), and that a D component was only detected for a limited number of items, D was not modeled in the subsequent analyses. Considering that the power to detect sources of variation is greater at the subscale level, which conformed predominantly to an AE model, we proceeded with the multivariate analyses using an AE model.

Table 2

Fit measures for the 3-6 -factor geomin-rotated EFA solutions and the 5-factor CFA model

Method	Factors	$\chi 2$	df	TLI	RMSEA
EFA	3f	19932	535	.822	.094
	4f	13918	599	.887	.075
	5f	8266	648	.940	.055
	6f	6827	642	.951	.049
CFA	5f	17222	436	.789	.099

Table 3

Standardized factor loadings (λ), residual variance (μ_R = mean residual variance) and inter-factor correlations in the geomin–rotated 5-factor phenotypic EFA solution. The highest loading for each item is highlighted. Factor loadings smaller than .10 are omitted

Item	λ_N	λ_E	λ_O	λ_C	λ_A	Res
n1	.55				.15	.67
n2	.63					.55
n3	.65				.13	.56
n4	.57	-.15				.61
n5	.76					.43
n6	.75			-.10		.37
n7	.66				.18	.54
n8	.56				-.28	.61
n9	.57		-.11	-.28		.49
n10	.67	-.12			.18	.49
n11	.59			-.27		.48
n12	.56			-.13		.64
						μ_R = .54
e1		.67	-.11			.58
e2	-.18	.65				.49
e3	-.45	.53			.11	.39
e4		.63			.13	.53
e5		.61			-.21	.61
e6	-.19	.32		-.15	.21	.80
e7	-.22	.34		.21	-.26	.64
e8	-.31	.71				.26
e9	-.43	.44				.50
e10	.	.23		.14	-.29	.83
e11	-.15	.41		.41	-.19	.48
e12	-.26	.14	.11		-.14	.86
						μ_R = .58
o1	.14		.26	-.32	.12	.79
o2	-.11	-.13	.22	-.20		.91
o3		-.11	.60			.64
o4	-.19		.36		.16	.82

o5			.62		.16	.60
o6		.15	.26			.90
o7			.31	.10	.25	.81
o8	-.11		.15			.96
o9			.67			.53
o10			.53			.71
o11	-.13	.14	.53	.17	-.14	.59
o12			.58		-.21	.63
						μ_R = .74
a1	.13	.18		.27	.25	.80
a2	-.27	-.13			.43	.71
a3				.16	.53	.63
a4		.13			.38	.83
a5	-.37				.44	.62
a6	-.43		.12		.34	.70
a7	-.13	.36		.20	.20	.68
a8		.11			.63	.58
a9		.20			.58	.63
a10	.21	.29		.39	.40	.55
a11					.42	.81
a12			-.16	.12	.54	.65
						μ_R = .68
c1		-.14	-.11	.57	.14	.66
c2	-.16			.62		.55
c3		-.17		.49		.74
c4	.10			.45	.29	.70
c5				.62	-.24	.56
c6	-.19	-.18	-.15	.55		.58
c7		.19		.67	-.10	.49
c8				.56	.20	.62
c9	-.14		-.12	.41	.29	.66
c10		.12		.61	-.12	.53
c11	-.46			.46	.11	.44
c12	.16	.19	.13	.38	-.27	.72
						μ_R = .60

Factor correlations:

	N	E	O	A
E	-.25			
O	.08	.13		
A	-.31	.24	.07	
C	-.06	.03	.00	.09

Table 4

Standardized factor loadings (λ), residual variance (μ_R = mean residual variance), and inter-factor correlations (right) in the phenotypic 5-factor model

Item	λ_N	Res	Item	λ_E	Res	Item	λ_O	Res	Item	λ_A	Res	Item	λ_C	Res
n1	.52	.73	e1	.43	.81	o1	.32	.90	a1	.41	.84	c1	.45	.80
n2	.65	.58	e2	.61	.63	o2	.16	.97	a2	.47	.78	c2	.66	.56
n3	.63	.60	e3	.82	.33	o3	.61	.63	a3	.57	.67	c3	.40	.84
n4	.62	.61	e4	.58	.66	o4	.32	.90	a4	.36	.87	c4	.43	.82
n5	.73	.47	e5	.40	.84	o5	.70	.51	a5	.67	.55	c5	.51	.74
n6	.78	.39	e6	.33	.89	o6	.28	.92	a6	.48	.77	c6	.60	.64
n7	.63	.61	e7	.55	.70	o7	.35	.88	a7	.64	.59	c7	.58	.67
n8	.52	.73	e8	.85	.28	o8	.11	.99	a8	.42	.83	c8	.52	.73
n9	.68	.54	e9	.76	.42	o9	.76	.42	a9	.30	.91	c9	.54	.71
n10	.69	.52	e10	.08	.99	o10	.55	.69	a10	.55	.70	c10	.67	.55
n11	.70	.51	e11	.65	.57	o11	.47	.78	a11	.19	.96	c11	.84	.30
n12	.59	.65	e12	.33	.89	o12	.50	.75	a12	.31	.91	c12	.24	.94
		μ_R=.58			μ_R=.67			μ_R=.78			μ_R=.78			μ_R=.69

	N	E	O	A
E	-.62			
O	.12	.08		
A	-.41	.46	.10	
C	-.53	.45	-.03	.48

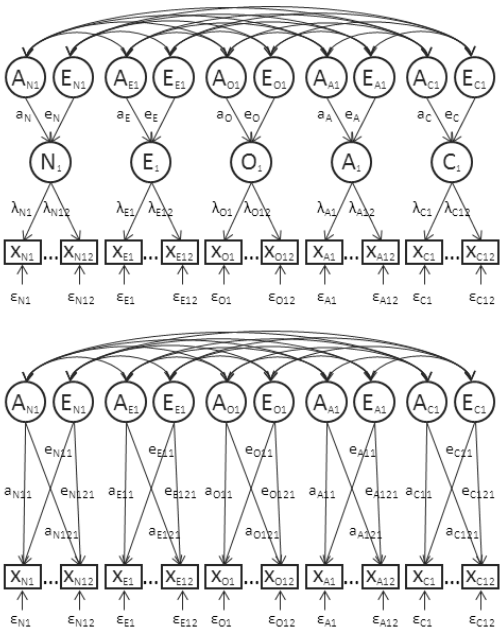

Figure 3. The common (upper panel) and independent (lower panel) pathway models fitted to the NEO-FFI data. The models are only partially depicted; the full models include a 'twin1' and a 'twin2' part, analogous to Figure 1. The 'within twin' A factors are mutually correlated, as are the 'within twin' E factors. The item-specific factors were modeled as correlated over twin 1 and twin 2 (i.e., the 60x60 twin 1 - twin 2 residual covariance matrix is diagonal).

To test the mediation of the genetic and environmental influences by the latent personality factors, in the next step a common pathway and an independent pathway AE model were tested (Figure 3). In the common pathway model, the variation in the latent personality factors was decomposed into additive genetic and unique environmental components. Additive genetic influences explained around half of the variance in the latent traits (.48, .48, .58, .43, and .47 for the N, E, O, A, and C factors, respectively), the remainder of the trait variance being determined by unique environmental factors. The fit measures associated with the model were: $\chi2$=112786, df=14776, TLI=.832, RMSEA=.069[27]. The independent pathway model was formulated by disposing of the phenotypic factors employed in the common pathway model. The fit measures associated with this model were: $\chi2$=94852, df=14721, TLI=.862, RMSEA=.062[4]. As the difference between chi-square values obtained using the WLSM estimator is not chi-square distributed (L. K. Muthén & Muthén, 1998-2007), the comparison of the common and the independent pathway model was carried out using a chi-square difference test with scaling correction factors (Satorra & Bentler, 2001). The resulting chi-square difference was $\Delta\chi2$=123646, df=55. Additionally, the comparison was performed using maximum likelihood estimation with robust standard errors (MLR; L. K. Muthén & Muthén, 1998-2007). The results converged with those obtained using the WLSM estimator (common pathway: $\chi2$=40477, df=14195, TLI=.699, RMSEA=.036;

[27] As MPlus output obtained using the WLSMV estimator could not be used for subsequent chi-square difference testing due to the non-linear constraints in the model, estimation was performed using the WLSM estimator.

independent pathway: $\chi 2$=35423, df=14140, TLI=.756, RMSEA=.033; chi-square difference: $\Delta\chi 2$=3115, df=55). The significant difference between the fit of the two models indicates incomplete mediation of the genetic and environmental influences by the latent personality factors.

In the light of the well-established presence of cross-loadings in the NEO-PI-R and the NEO-FFI (J. Block, 1995; McCrae & Costa, 2008; Parker et al., 1993), an additional test was performed: a common and an independent pathway model based on the phenotypic model with 94 cross-loadings were formulated and fitted to the data. Due to the computational intensity of fitting these models using the WLSMV estimator, the MLR estimator was used. The resulting fit measures were $\chi 2$= 31176, df=14101, TLI=.803, RMSEA=.029, and $\chi 2$= 25831, df= 13952, TLI=.862, RMSEA=.025, respectively. Consistently with the results obtained for the simple structure models, the fit of the two models differed significantly ($\Delta\chi 2$=4034, df=149), indicating incomplete mediation of the genetic and environmental effects by the latent personality factors, despite the assumption of simple structure being discarded.

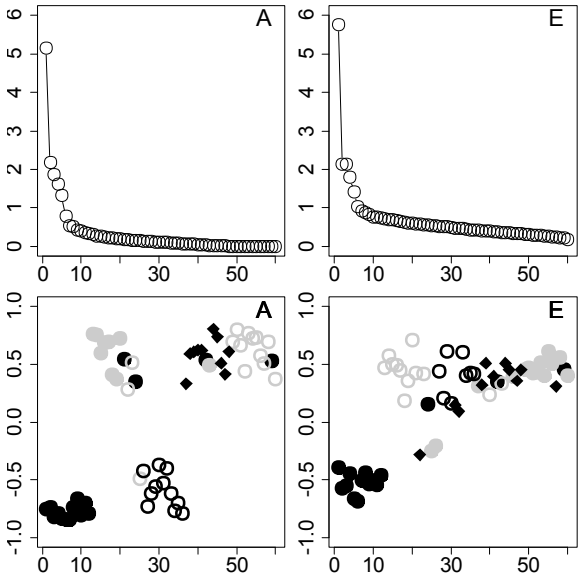

Figure 4. Eigenvalues of $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_E$ matrices (upper panel) and factor loadings obtained in EFA solutions with 5 A and 5 E factors (lower panel). Shapes/shading code for different latent factors. Only the highest factor loading for each item is shown.

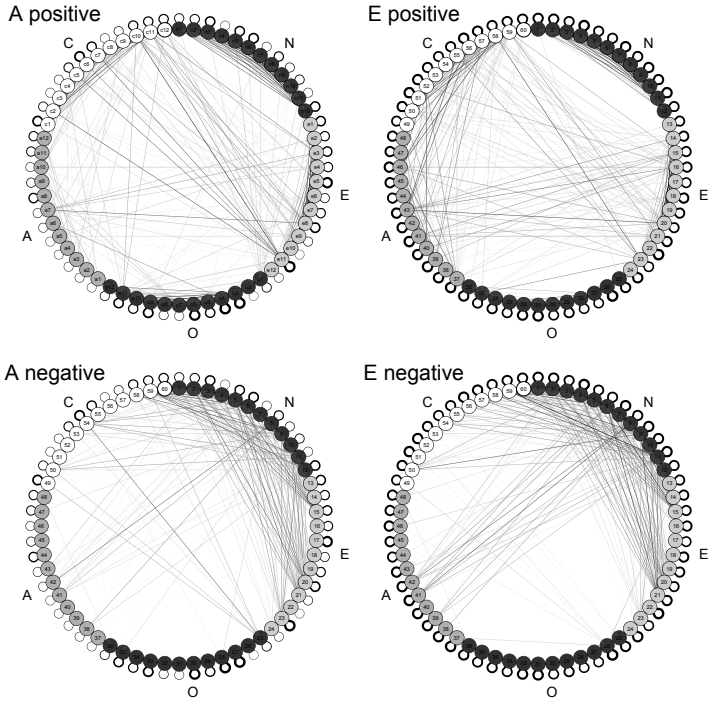

Figure 5. Graphical representations (Epskamp et al., 2012) of the A (left) and E (right) covariance components of the NEO-FFI. Positive (upper panel) and negative (lower panel) covariances are shown separately. Nodes (i.e., circles) represent items. The thickness of the edges (i.e., of the lines connecting the nodes) represents the magnitude of covariance between the items.

Finally, to further explore the structures of the genetic and environmental influences on the item covariation, the 60x60 phenotypic polychoric covariance matrix was decomposed into 60x60 additive genetic and unique environmental matrices, and the dimensionality of these two covariance matrices was assessed using EFA (geomin rotation). The results are given in Figure 4. As evident from the Figure, the scree plots (upper panel) for the A and the E matrix both indicate a 5-factor model. Furthermore, the factor structures of the additive genetic and the unique environmental influences (lower panel Figure 4) resemble very closely the 5-factor phenotypic structure of the NEO-FFI. This can also be seen in Figure 5, which depicts the pattern and the magnitude of the A and E intercorrelations between the items; as evident, the A and the E covariance structure resemble each other to a high degree. Finally, the magnitudes of the A and E variance components of each of the 60 items are depicted in Figure 6; on average, these are .33 and .67, respectively.

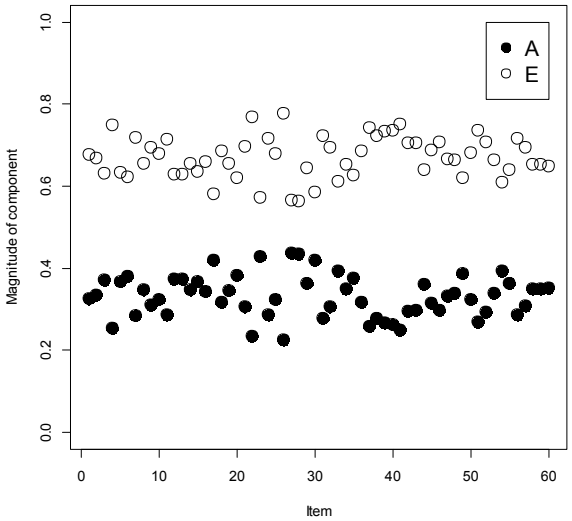

Figure 6. Magnitude of the A and E variance components of the 60 items of NEO-FFI.

Discussion

In the present study, we tested the hypothesis that the Big Five factors are causally efficient entities, which serve to mediate the genetic and environmental effects on phenotypic data. This hypothesis was tested by comparing the fit of independent pathway models to the fit of common pathway models. If the latent variables in the FF model indeed act as causes of behavior, which fully mediate genetic effects, the independent and common pathway models should fit equally well. If, however, the latent variables are merely statistical constructs that organize phenotypic correlations but do not have the status of causally efficient entities, the independent pathway model should show superior fit. In addition to these hypothesis tests, the structures and the dimensionalities of the latent genetic and environmental effects were examined in an exploratory factor analysis. Two findings emerged: 1) the constraints associated with the common pathway model were not tenable, i.e., the fit indices favored the independent pathway model, and 2) the rotated 5-factor structures as obtained in the EFA of genetic and environmental correlation matrices are quite similar.

The fact that our analyses favor the independent pathway model constitutes evidence against the realist interpretation of the Big Five dimensions. Even when we allow cross-loadings to be present, the magnitude of the test statistic based on the models is such ($\Delta\chi 2$= 4034, df=149, based on MLR) that the degree of misfit associated with the common pathway model is considerable. Perhaps, one could argue that both models fit well in view of the acceptable approximation error (common pathway model RMSEA: .029; independent pathway model RMSEA: .025). However, in our view, the acid test here is not the overall degree approximation error of the individual models. Rather, it is the model comparison, which reveals the specific source of approximation error, namely the proportionality constraints associated with the common pathway model. These are evidently untenable.

The fact that the exploratory factor analyses of the additive genetic and unshared environmental correlation matrices produced highly similar 5-factor models is interesting in its own right, and by no means a trivial finding. The phenotypic FF model does not imply five genetic and environmental factors to surface: the latter implies the former, but not vice versa, and several examples are known in which the structures diverge (Franić, Dolan, Borsboom, van Beijsterveldt, & Boomsma, 2014; Kendler et al., 1987). Thus, although the data unambiguously reject the proportionality constraints derived from the latent variable hypothesis, it is certainly not the case that the A and E covariance structures are radically different.

Therefore, although the formal tests indicate that the independent pathway model is preferable, the exploratory results do lend some credence to the latent variable hypothesis. One possible explanation for this finding is that, although the full mediation hypothesis is not precisely true, it does provide a reasonable approximation to the generating model. The specific reasons for rejecting the common pathway model may, for instance, be highly local (due only to a subset of observed variables), and thus the violation may be accommodated by the addition of parameters or by the removal of offending variables. A second possible explanation is that, even though we have fitted highly relaxed versions of the FF model, the models still embodied auxiliary hypotheses that were not exactly true (e.g., linearity, normality, continuity, discarding C and D effects) which may have produced misfit evident in the likelihood ratio tests (which are derived on the hypothesis that the least restricted model fits the data). A third possibility is that the similar structure of A and E matrices, as evidenced in the present chapter, is simply a chance finding that has little to do with the realist interpretation of the Big Five dimensions. This hypothesis is tenable, because the truth of an independent pathway model does not preclude that the genetic and environmental covariance structures comprise 5 factors, with or without configurally similar loadings.

In our view, the formal test on the proportionality of loadings should carry the primary weight of the evidence, as it was designed specifically to distinguish between the tested models. However, it is certainly notable that the A and E covariance matrices showed strikingly similar structures, and even though this equivalence is not a formal test of the common pathway hypothesis, it does confirm an indirect consequence of that hypothesis. Further research may investigate the relevance of this finding to the veracity of the FF model.

In the present analyses, the genetic and environmental variables are all-encompassing in the sense that they represent all (unmeasured) polygenic and unshared environmental influences. However, the mediation hypothesis can be formulated with respect to any measured variable. It is a drawback of much of the research concerning the covariates of the Big Five dimensions that they generally involve Big Five subscale scores rather than items. We consider the demonstration of the mediatory role of, say, neuroticism in the relationship between a covariate (e.g., sex) and the neuroticism items, to be a stronger result than the demonstration of a sex difference in the neuroticism scale scores. In this regard the present results are relevant to gene-finding studies (e.g., genome-wide association studies; GWAS). If a measured genetic variant has its effect on the common factor “neuroticism”, then its effect is present in all the relevant items, and the interpretation of the gene as a "gene for neuroticism" is tenable. This is not so if the effect is limited to a subset of the items, or perhaps even a single item.

Chapter 6

Childhood and Adolescent Anxiety and Depression: Beyond Heritability

Abstract

The present chapter reviews the methodology of behavior genetics studies addressing research questions that go beyond simple heritability estimation, and illustrates using representative studies on childhood and adolescent anxiety and depression. The classical twin design and its extensions have been employed to investigate age and sex differences in the genetic determinants of complex traits and disorders, the role of genetic factors in explaining comorbidity, gene-environment interaction, and the effects of social interaction among family members. The review provides relatively consistent evidence for: a) small to negligible sex differences in the genetic etiology of childhood anxiety and depression, b) a substantial role of genetic factors in accounting for the temporal stability of these disorders, c) a genetic basis for the comorbidity between anxiety and depression, d) a possible role of interaction between genotype and environment in affecting liability to these disorders, e) a role of genotype-environment correlation, and f) a minor, if any, etiological role of sibling interaction. Implications for treatment are discussed.

Based on: Franić, S., Middeldorp, C. M., Dolan, C. V., Ligthart, L., & Boomsma, D. I. (2010) Childhood and Adolescent Anxiety and Depression: Beyond Heritability. *Journal of the American Academy of Child and Adolescent Psychiatry, 49(8)*, 820-829.

An important role of twin and family studies in psychiatry has been to establish the contributions of genetic and environmental factors to the observed or phenotypic individual differences in psychiatric disorders. Genetic research, frequently employing the classical twin design, has demonstrated the pervasive importance of both genetic and environmental factors in complex psychiatric disorders and related traits (e.g., personality). These findings can be seen as guiding molecular genetic studies, as there would be no use of searching for genes that influence individual differences in a behavior that is not influenced at least in part by genetic factors. Twin studies have evolved, however. At present, thanks to developments in statistical methodology and the establishment of large twin registries in which both measured genotypes and environmental variables are available, psychiatric genetic research is moving beyond the relatively simple task of assessing the contributions of genetic and environmental factors. Present research often focuses on more subtle issues, such as how genetic and environmental influences are modulated by age and sex, or how gene expression is affected by the environment.

In the present chapter, we review the recent findings of genetic research in the area of childhood and adolescent anxiety and depression. The aim is not to provide a comprehensive overview of all existing literature, but rather to introduce the reader to the methods employed in behavior genetics while presenting the results of representative research relevant to the current key issues. To this end, we first introduce the classical twin design and a number of recent extensions thereof. These extensions serve to address specific issues, which we discuss in the context of research findings. The issues include 1) sex and age differences in the genetic etiology of childhood anxiety and depression, 2) the nature of the comorbidity of these disorders, 3) the interplay between genes and the environment, and 4) social interactions among family members. We conclude with a discussion of the clinical implications of recent findings.

Classical twin design and heritability estimation

The classical twin design may be employed to decompose the variance of a phenotype into components due to genetic (G), shared environmental (C), and unique environmental (E) factors. Genetic and environmental variances are attributable to the contribution of an unspecified number of genes and environmental exposures, respectively. As genes come in pairs, each autosomal genetic locus comprises two alleles, one contributed by each parent. The alleles may be the same (the individual is a homozygote) or different (the individual is a heterozygote). The phenotypic effects of alleles may add up (additive genetic influences; A) or interact (non-additive genetic influences; D). Because dominance effects are rarely observed in genetic studies of anxiety and depression (Gregory & Eley, 2007), we limit our discussion primarily to additive genetic effects. Common environmental factors (C) are shared among family members, and contribute to their similarity (for instance, parental socio-economic status or parenting style may increase similarity between two siblings growing up in the same home). Unique environmental factors generate differences between family members.

Table 1 details the decomposition of the phenotypic variance. Under the assumption that the phenotype is affected by additive genetic, common environmental, and unique environmental factors (ACE model), the variance (V) is decomposed as follows: $V = V_A + V_C + V_E$. The phenotypic correlation between monozygotic (MZ) twins is a function of their staring their additive genetic (as they develop from the a single fertilized egg) and common environmental influences. On average, dizygotic (DZ) twins share 50% of their segregating genes, as do other first-degree relatives. Based on the phenotypic (i.e., observed) degrees of

resemblance between MZ and DZ relatives, one may estimate the contribution of the additive genetic variance to the total trait variance (i.e., the heritability), as detailed in table 1: $h^2 = 2(r_{MZ} - r_{DZ})$. Table 1 also details the estimation of c^2 and e^2 (the proportions of variance explained by shared and non-shared environmental factors, respectively).

Table 1

Decomposition of phenotypic variance into additive genetic, shared environmental and unique environmental components, under an ACE model.

Expression	Meaning
$V = V_A + V_C + V_E$	The variance of an observed trait (i.e., phenotype), may be decomposed into additive genetic, shared environmental, and unique environmental components (V_A, V_C, and V_E, respectively)
$r_{MZ} = (V_A + V_C) / V$	Correlation between monozygotic (MZ) twins
$r_{DZ} = (½V_A + V_C) / V$	Correlation between dizygotic (DZ) twins
$V_A = 2V(r_{MZ} - r_{DZ})$	Contribution of additive genetic factors to the phenotypic variance
$V_C = r_{MZ} V - V_A$	Contribution of shared environmental factors to the phenotypic variance
$V_E = V - r_{MZ} V$	Contribution of unique environmental factors to the phenotypic variance
$h^2 = 2(r_{MZ} - r_{DZ})$ $= 2(V_A + V_C - ½V_A - V_C) / V$ $= 2(V_A - ½V_A) / V = V_A / V$	Standardized contribution of additive genetic factors to the phenotypic variance (i.e., the heritability coefficient; h^2)

For the classical twin design to yield correct and generalizable estimates of the magnitudes of genetic and environmental influences, a number of assumptions have to hold (for a review we refer to G. Carey, 2009; Falconer & Mackay, 1996; Keller & Coventry, 2005; Martin et al., 1997; Plomin et al., 2008). First, it is assumed that the environmental sharing is equal in MZ and DZ twins – an assumption that may be challenged, for instance, due to the assumption that MZ twins may experience more similar environments than DZ twins. However, with regard to environmental aspects relevant to psychopathology, it has been shown that greater environmental sharing in MZs is a consequence, rather than a cause of their phenotypic similarity (Martin et al., 1997). Second, results obtained on twin samples are assumed to generalize to non-twin populations. This assumption appears to hold, as estimates of environmental and genetic parameters obtained in twin studies tend to differ little from those obtained in the general population (Martin et al., 1997). Other assumptions of the classical twin design include (but are not limited to) the absence of assortative mating (i.e., the tendency to mate with individuals with a phenotype similar to one's own), genotype-environment interaction (i.e., dependency of genetic effects on the environment and vice versa), and genotype-environment correlation (i.e., non-random placement of genotypes in the range of available environments). These effects may be assessed if the appropriate data are collected and, when assumptions are not met, may be modeled. For instance, data on parents or spouses of twins can be employed to examine and model possible effects of assortative mating.

The classical twin design does not allow for the simultaneous estimation of the additive and dominance genetic, and common and unique environmental effects.

Consequently, one of these sources of variance is typically assumed to be absent. The choice may be informed by model comparison (e.g., an ACE against an ADE model), or by inspection of phenotypic MZ and DZ correlations: a DZ correlation greater than half of the MZ correlation suggests an effect of C on the trait; a DZ correlation smaller than half of the MZ correlation suggests D.

Beyond heritability

Using the classical twin design, researchers have established the role of genetic and environmental factors in virtually all psychiatric traits. Presently, thanks to several developments over the last half of the past century, it is possible to move beyond the basic question of heritability. These developments include the large-scale collection of phenotypic data by means of well-validated standardized questionnaires and the collection of extensive genotypic and environmental data. Furthermore, the realization that the resolution of relatively subtle effects requires large sample sizes has provided the impetus for the establishment of large twin and family registries. In case of children, survey data obtained from teachers, in addition to parental ratings, form an important source of information concerning childhood development and psychopathology. Finally, an important development have been the possibilities offered by advances in statistical and psychometric modeling.

These developments are now coming together and researchers are looking into more subtle aspects of genetic and environmental effects. These include the dependency of these effects on sex and age. Age or developmental effects are important in understanding the role of genetics in the development of psychopathology; also, such effects negate the perception of the heritability of a phenotype or disorder as a fixed entity. In addition, the availability of measured genotypes and environments allow researchers to address the often neglected issue of gene-environment interaction in the development of psychopathology. Finally, given the genotype data, researchers are increasingly turning to the detection of individual genes associated with particular phenotypes by means of genome-wide association (GWA) studies (McCarthy et al., 2008; Teare & Barrett, 2005).

Sex differences in the genetic architecture of anxiety and depression

Can heritability differ across the sexes? Are genetic effects in males and females attributable to the same genes (whose effect may be modulated by sex), or different genes? The former suggests a quantitative sex difference in the genetic architecture; the latter a qualitative difference. Both quantitative and qualitative effects can result in sex differences in the estimated variance components. Quantitative differences are apparent if estimates of genetic and environmental variances differ in males and females; qualitative differences can be detected by testing whether the genetic correlation in opposite-sex DZ twins is significantly different from its expected value based on the correlations in same-sex DZ twins. A correlation lower than expected suggests that genetic risk factors differ across the sexes. An analogous method can be used to establish whether different shared environmental factors are implicated in the trait in males and females.

The applications of this methodology to childhood anxiety and depression have generally revealed small to negligible differences in heritabilities in boys and girls. However, the results do appear to vary with the exact definition of the phenotype (e.g., general anxiety vs. separation anxiety). In addition, the variation in the results may depend on age or measurement scale (e.g., a continuous variable vs. a dichotomy). The majority of

studies included in the present review employed continuous measures of anxiety and depression, defined as summed scores on standardized questionnaires such as the Child Behavior Checklist (CBCL; Achenbach, 1991). In case a categorical measure of psychopathology was used, this was noted in the text.

Recently, Lamb et al. (Lamb et al., 2010) examined anxious depression and withdrawn behavior (defined on a 3-point scale) in 12, 14, and 16 year-olds, and found no evidence for differential heritabilities across the sexes. Similarly, Rettew et al. found the genetic and environmental influences on neuroticism to be of equal importance in 12-17 year-old male and female adolescents (Rettew et al., 2006). A similar result was obtained by Rice et al. (Rice et al., 2002b) for parental ratings of depression in 8-17 year-olds. For self-rated symptoms, however, the same study obtained evidence for greater genetic and a smaller shared environmental influence in boys. For separation anxiety in 3-18 year-olds the opposite pattern was observed, i.e., a greater genetic and a smaller shared environmental influence was found in girls (Hettema et al., 2001). A similar result was obtained by Happonen et al., using teacher ratings of depression symptoms in 11 and 12 year-olds (Happonen et al., 2002). For self-, parent-, and peer-ratings, however, the same study found equal heritabilities across sexes. In addition, several recent studies examined possible qualitative sex differences. Rettew et al. obtained some support for qualitative sex differences with respect to neuroticism in adolescent twins (Rettew et al., 2006). However, in a study of anxiety and depression in 3 and 12 year-olds, Boomsma et al. concluded that the same genes were expressed in girls and boys (Boomsma et al., 2005b).

In conclusion, most studies suggest that sex differences in the genetic architecture of childhood anxiety and depression, if present, are small. The detection of sex effects may depend on other factors such as age and definition of phenotype, and additional research may elucidate the role of such factors.

Age differences in the genetic architecture of anxiety and depression

Genetic and environmental effects may be age-dependent. For instance, there is no reason to assume that environmental or genetic influences on anxiety or depression at age 5 are identical to those at age 10. To establish such differences one can study twins who differ with respect to age. However, the correct interpretation of any established difference requires a study of twins in a longitudinal design, in which phenotypes are measured in the same twins at two or more occasions. Such data allow one to determine the role of genetic and environmental factors in the stability of individual differences over time. For instance, one may address the question of whether the same genes contribute to individual differences over time (which would result in a high degree of genetic stability), or whether, for instance, shared environmental effects tend to diminish over time.

Studies of age-related changes in anxiety and depression have generally demonstrated these disorders to be moderately stable during childhood. However, many children who initially display relatively high levels of anxiety or depression go on develop normally, whereas some other children, who displayed initial normal development, go on to develop anxiety or depression at a later age (Middeldorp & Boomsma, 2009). Furthermore, studies on age-related differences generally report an influence of shared environment on anxiety and depression during childhood that fades as children enter adolescence (Hettema et al., 2001; Middeldorp & Boomsma, 2009; Rapee et al., 2009; Rice et al., 2002b; Sullivan, Neale, & Kendler, 2000). In adulthood, the influence of shared environment disappears, and the relative influence of genetic factors increases; as a consequence, the heritability of these traits increases with age. Similar results have been obtained for related phenotypes, such as

withdrawn behavior and obsessive-compulsive symptoms (Hudziak et al., 2004; Lamb et al., 2010; Van Grootheest et al., 2008). A recent meta-analysis supported the increase in heritability of anxiety and depression after childhood, and found this increase to be greater for anxiety-related symptoms than for symptoms of depression (Bergen, Gardner, & Kendler, 2007).

In addition, longitudinal studies have established qualitative age-related differences. In a study of anxious/depressed behavior as reported by parents, Boomsma et al. found relatively low stability of genetic effects on anxious depression at ages 3 to 7, and an increase in genetic stability from ages 7 to 12 (Boomsma, Van Beijsterveldt, Bartels, & Hudziak, 2008). Kendler et al. reported a high genetic stability of anxiety and depression throughout ages 8 to 20 (Kendler, Gardner, & Lichtenstein, 2008). However, additional, age-specific genetic influences emerged in adolescence and early adulthood. Hoekstra et al. (R. A. Hoekstra, Bartels, Hudziak, Van Beijsterveldt, & Boomsma, 2008) examined withdrawn behavior, a trait found to predict later anxiety and depression (Goodwin, Fergusson, & Horwood, 2004), and observed considerable stability of genetic influences throughout childhood. Finally, Kendler et al. studied the development of situational, social, animal, and blood/injury fears, and found that genetic effects on these fears become more fear-specific as a function of age, and that in late adolescence new genetic influences relevant to social fears emerge (Kendler, Gardner, Annas, & Lichtenstein, 2008).

In summary, most longitudinal studies report a small to moderate temporal stability of childhood anxiety and depression. While genetic factors appear to account for most of this stability, additional age-specific genetic factors have also been found to emerge over time, possibly accounting for some of the temporal instability of these disorders. It should, however, be mentioned that the role of the informant remain a contentious issue. Estimates of genetic influences on psychopathology are known to depend on whether the data are obtained from the child itself or from an informant such as a parent (Rapee, Barrett, Dadds, & Evans, 1994). This is especially relevant for studies of age-related changes, as such studies often use informants (parent, teachers) to assess the behavior of young children, and self-report in older children.

Comorbidity

Twin designs may be employed to address the question of whether the co-occurrence of two disorders has a genetic or an environmental basis, or arises from a direct causal interaction between the disorders (e.g., depression directly causing anxiety). To address the question of comorbidity, MZ and DZ correlations on different measures (e.g., depression in twin 1 and anxiety in twin 2) may be compared; a comparatively higher MZ correlation indicates a genetic basis for the comorbidity.

Multiple studies have demonstrated comorbidity within anxiety disorders, as well as between anxiety disorders and depression, both in children and adults (Angold et al., 1999; Brady & Kendall, 1992; Middeldorp et al., 2005). This comorbidity is explained partly by shared genetic risk factors (Middeldorp, et al., 2005). For instance, Eley at al. studied five anxiety-related syndromes (general distress, separation anxiety, fears, obsessive-compulsive behaviors, and shyness/inhibition) and found moderate genetic overlap among these syndromes (Eley et al., 2003). Silberg et al. found that depression in girls after age 14 was genetically correlated with earlier symptoms of simple phobias and overanxious disorder, but environmentally correlated with separation anxiety (Silberg, Rutter, & Eaves, 2001). Direct causality may also been inferred; for instance, prolonged anxiety may lead to a depressive episode. If such a direct causal relationship is present, all the genetic and

environmental effects on the causal disorder will also be present in the "caused" disorder. A recent study of direct causal effects of exercise on depression in adults yielded no evidence for direct causal effects (de Moor et al., 2008).

Genotype-environment interaction (GEI)

Most studies do not examine how genetic and environmental factors combine in affecting liability to illness. It is frequently assumed, implicitly or explicitly, that genes and the environment do not interact, i.e. that they affect the phenotype independently of each other. This is not necessarily the case. For instance, assume an environmental factor, say an important life event, and a single gene with 3 possible genotypes: AA, Aa, and aa. If the life event increases the risk of disease to the same extent in individuals with any genotype, the genotype and environment are said to have an additive effect (Figure 1a). However, if the average change in risk associated with the life event is different in individuals with different genotypes, the genotype and the environment interact (Figure 1b). For instance, the liability of individuals with the AA genotype may increase substantially as a result of the life event, whereas those with the aa genotype may barely be affected.

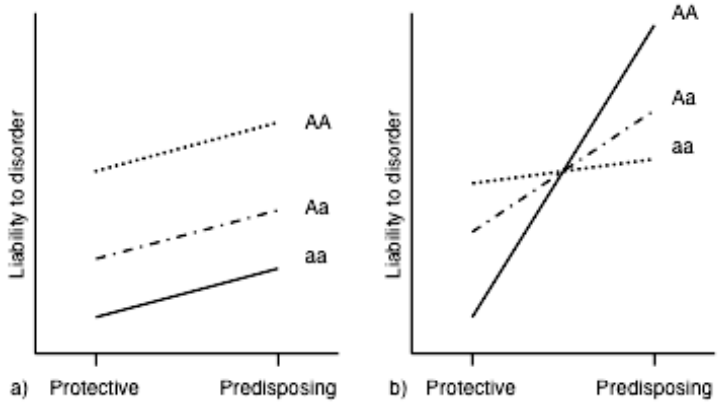

Figure 1. Liability to developing a disorder as a function of genotype (AA, Aa or aa) and environmental exposure (protective or predisposing). The predisposing environment is associated with an increase in disease liability. a) This increase is equal in individuals with the AA, Aa and aa genotype (additive effects). b) The increase is different in individuals with different genotypes. Individuals with the AA genotype have a disproportionately low chance of developing a disease in the protective environment, but suffer from a disproportional increase in liability when exposed to the predisposing environment (Kendler & Eaves, 1986).

GEI may be examined by estimating the relative contributions of genetic and environmental factors to a trait across different levels of environmental exposure (Eaves, 1982). A difference in the genetic contribution to a phenotype as a function of an environmental moderator (for instance, a higher heritability of depression in people with lower levels of social support) would constitute evidence for GEI. One may also estimate whether different genes are expressed across different levels of environmental exposure. This can be accomplished using a twin design in which each individual is measured at two different levels of environmental exposure (e.g. before and after an important life event). The correlation between the

measures in these two conditions is partitioned into components due to genetic and environmental factors. A low genetic component would suggest that different genes are expressed across the environmental exposures (Falconer & Mackay, 1996).

A number of recent studies of childhood psychopathology found evidence of GEI (Eley et al., 2004; Lau & Eley, 2008b; Silberg, Rutter, Neale, & Eaves, 2001). For instance, Silberg et al. found greater genetic effects on anxiety/depression in adolescent girls who had experienced recent negative life events than in those who had not (Silberg, Rutter, Neale, et al., 2001). Similar findings were obtained for separation anxiety symptoms in childhood and panic anxiety symptoms in adolescence. Lau and Eley (Lau & Eley, 2008b) observed that 15-year-old adolescent girls at a genetic risk for developing depression tend to experience more negative life events and maternal punitive discipline (an instance of genotype-environment correlation, addressed in the next section), and were at a higher risk of developing depressive symptoms in response to those events (GEI). Feinberg et al. (Feinberg, Button, Neiderhiser, Reiss, & Hetherington, 2007) found no evidence for interaction of parental negativity and warmth with the heritability of interview-assessed depression in adolescents, but showed that effects of individual-specific environment (E) on depression changed as a function of parental negativity. As parental negativity increased, the effects of unique environment on depression increased. The latter finding is consistent with a recent study that examined the moderating role of six environmental factors (mother- and father-child relationship problems, antisocial and pro-social peer affiliation, academic achievement and engagement, and stressful life events), and found that the effects of individual-specific environment on symptoms of depression and anxiety increased as the environmental adversity increased (Hicks, DiRago, Iacono, & McGue, 2009).

When measures of the genotype are available, it is possible to test for an interaction of the environment with a specific gene variant. For instance, Caspi et al. investigated the association between the serotonin transporter gene and depression in adults who had experienced stressful life events and those who had not (Caspi et al., 2003). Stressful life events were associated with depression, but only in individuals who carried at least one copy of the short allele of the serotonin transporter gene polymorphism (5-HTTLPR). However, a meta-analyses of replication efforts has failed to confirm this finding (Risch et al., 2009).

Finally, it should be noted that the variation in the phenotype due to GEI, when not explicitly accounted for, inflates either the heritability estimate or the estimate of unique environment (Purcell, 2002). In particular, unmodeled variance due to the interaction of genotype and shared environment (AxC) inflates the estimate of heritability, and unmodeled variance due to the interaction of genotype and unique environment (AxE) inflates the estimate of variance due to unique environmental factors. If they are expected, but not explicitly modeled, these effects should be borne in mind in interpreting the results of twin studies.

Genotype-environment correlation

In the traditional twin design, genetic and environmental contributions to individual differences are usually assumed to be independent or uncorrelated. However, the possibility of genotype-environment correlation (rGE) is widely recognized. In fact, three types of rGE are distinguished, namely passive, evocative, and active rGE. Passive rGE refers to the case in which children inherit genes and an environment that both predispose them to a given phenotypic outcome (Eaves, 1987). For an example of this, think of a parent who suffers from depression. This parent may pass on genes that predispose their child to develop

depression, but in addition may inadvertently create a depressogenic environment for the child (by being unresponsive, unhappy, demoralized, etc.). Evocative rGE refers to the situation in which person's genetically influenced characteristics evoke environmental reactions which exacerbate the characteristics (Plomin, Defries, & Loehlin, 1977). For instance, an anxious and withdrawn child, simply by behaving anxiously may elicit certain responses in other children (e.g. shunning) or in parents (more protective parenting), which contribute to the child's anxiety. Finally, active rGE refers to the situation in which individuals, as a consequence of certain characteristics, actively seek out or create environments which are conducive of these characteristics (Eaves, 1987; Plomin et al., 1977). For instance, a withdrawn child may actively avoid social situations, such as birthday parties and sports activities, and thereby create an environment that fosters the child's general withdrawal.

One way of detecting rGE is by decomposing the correlation between a measured environmental factor and the phenotype of interest into A, C, and E components. Any contribution of genetic factors to the observed correlation means that the same genetic factors influence the environmental phenotype and the trait, thus creating a correlation between the two (Eaves, 1987; Plomin, et al., 1977).

A number of recent studies of childhood and adolescent anxiety and depression obtained evidence of rGE. Kendler and Baker (Kendler & Baker, 2007) reviewed 55 studies and found that environmental variables, such as stressful life events, parenting, family environment, social support, peer interactions and marital quality, are all under significant genetic influence (with heritability estimates ranging from .07 to .39). Narusyte et al. examined the association between maternal emotional overinvolvement and adolescent internalizing problems, and found that the latter evoked the former (Narusyte et al., 2008). Kendler demonstrated that the (genetically influenced) temperamental traits of the children elicited parental warmth, protectiveness, and authoritarianism (Kendler, 1996).

Social interaction among family members

Like other siblings, MZ and DZ twins interact as they grow up together. In this process the behavior of one sibling may influence the behavior of the other. If variance in the behavior of interest is in part genetic, then via their interaction, the genotype of one sibling exerts an influence on the phenotypic behavior of the other. Such interaction effects may be cooperative or competitive, depending on whether the behavior of one sibling facilitates or inhibits the behavior of the other. Cooperation, or positive interaction, leads to increased phenotypic MZ and DZ twin resemblance, while competitive or negative interaction tends to decrease it. In addition, sibling interaction affects the total phenotypic variance in MZ and DZ twins. If the interaction is cooperative and there is some genetic influence on the trait, the variance in both MZ and DZ twins is increased, but the increase is greater in the MZs. If the interaction is competitive and there is genetic influence on the trait, both MZ and DZ variability is decreased, but the decrease is greater in the MZs. Therefore, depending on the pattern of MZ and DZ resemblance and the total phenotypic variance observed in MZ and DZ twins, interaction effects may be detected, and cooperative and competitive interaction distinguished (G. Carey, 1986).

With respect to childhood anxiety and depression, studies have shown sibling interaction to play a minor, if any, etiological role. No evidence for sibling interaction effects on internalizing problems was observed in 3 and 10-15 year-old twins (van den Oord, Boomsma, & Verhulst, 1994; van den Oord, Verhulst, & Boomsma, 1996; Van der Valk, Stroet, & Boomsma, 1998), and only one study indicated possible twin contrast effects on

separation anxiety and depression (Eaves et al., 1997). However, pervasive effects of sibling interaction have been demonstrated in externalizing disorders (G. Carey, 1992).

Discussion

The present review demonstrates a substantial role of genetic factors in the etiology of depression and anxiety in the normal distribution. Insofar as one subscribes to a dimensional model of psychopathology (e.g., Widiger, 2005) in which affected children are considered to occupy the extreme of the population distribution, these results are highly relevant. The implication of genetic factors per se clearly does not mean that the child's level of anxiety or depression is immutable, or that environmental interventions are essentially pointless. In fact, the implication is quite the opposite: the substantial heritability of anxiety may imply that the biological parents also display increased levels of anxiety, which could affect the child through passive rGE. This could mean that treatment should involve the parents and possibly siblings. Moreover, through reactive rGE, anxious and depressed children can elicit a certain parenting style, such as punitive discipline or overprotective behavior, which may, in turn, create an environment that sustains the symptoms. It is important to explain this mechanism to parents. Similarly, the child's diagnosis may stem from an interaction of a genetic predisposition and adverse environmental factors, such as a divorce, or being bullied at school. Such information further supports that a strategy aimed at improving the child's environment may yield improvements in their overall emotional behavioral health.

As longitudinal studies have demonstrated, the chance that childhood anxiety or depression symptoms are transient is substantial; however, in case of persistent or recurrent symptoms, it is feasible to assume that genetic factors (which may be correlated with environmental risk, or interact with the environment) may play a greater role in their stability. In case of persistent symptoms, in addition to addressing environmental factors, therapy should also focus on individual characteristics (such as attributional style, coping style, or the tendency to ruminate) that could maintain the symptoms (Ciesla & Roberts, 2007; Hammen, 1992; Lau & Eley, 2008a). Of course, one should bear in mind that the framework informed by the results of behavioral genetic studies is probabilistic and predicated on a dimensional model of psychopathology. However, we believe that this framework is useful for thinking about the way in which genes and environment may contribute to childhood depression or anxiety.

Establishing significant heritability for childhood psychiatric disorders has promoted the attempts at localization and identification of genes that contribute to risk. Identification of genetic variants through genome-wide association (GWA) studies, which are hypothesis-free, is a feasible and appealing option that is increasingly employed for childhood traits. Two examples include a successful association study for fetal growth and birth weight, and studies of childhood ADHD (Freathy, Mook-Kanamori, Sovio, & Prokopenko, 2010). Although ADHD studies need to be improved and replicated, they are starting to implicate processes such as neuronal migration and cell adhesion and division as potentially important in the etiology of ADHD (Banaschewski, Becker, Scherag, Franke, & Coghill, 2010). Some researchers have voiced major concerns at the feasibility of explaining heritability by GWA studies (the famous 'missing heritability' problem) and a diversity of solutions has been proposed (Eichler et al., 2010). One solution might be the study of identical twins. The search for differences in genotypes within discordant MZ twin pairs seems to be a promising approach in gene finding. With future possibilities for human genome sequencing of large numbers of individuals it may even become feasible to turn this

strategy around and sequence large numbers of unselected MZ pairs, after which differences within pairs can be correlated with differences in phenotypes (Zwijnenburg, Meijers-Heijboer, & Boomsma, 2010). Ultimately, once the relevant genes are identified and their function understood, one may be able to move towards an effective combination of personalized treatment that includes both pharmacological and environmental (family-based) intervention.

Chapter 7

Genetic and Environmental Stability of Intelligence in Childhood and Adolescence

Abstract

The present study examined the genetic and environmental contributions to the temporal stability of verbal, nonverbal and general intelligence across a developmental period spanning childhood and adolescence (5-18 years). Longitudinal twin data collected in four different studies on a total of 1748 twins, comprising 4641 measurement points in total, were analyzed using genetic adaptations of the simplex model. The heterogeneity in the type of instrument used to assess psychometric intelligence across the different subsamples and ages allowed us to address the auxiliary question of how to optimally utilize the existing longitudinal data in the context of gene-finding studies. The results were consistent across domains (verbal, nonverbal and general intelligence), and indicated that phenotypic stability was driven primarily by the high stability of additive genetic factors, that the stability of common environment was moderate, and that the unique environment contributed primarily to change. The cross-subscale stability was consistently low, indicating a small overlap between different domains of intelligence over time. The high stability of additive genetic factors justifies the use of a linear combination of scores across the different ages in the context of gene-finding studies.

Appendices can be obtained at http://sanjafranic.com/dissertation.

Based on: Franić, S., Dolan, C. V., van Beijsterveldt, C. E. M., Hulshoff Pol, H. E., Bartels, M., & Boomsma, D. I. (2014). Genetic and Environmental Stability of Intelligence in Childhood and Adolescence. *Twin Research and Human Genetics, 17*(03), 151-163.

Intelligence is one of the most frequently studied human behavioral traits. Over the past century it has motivated research across a diverse range of fields including not only the behavioral sciences, but also genetics, neuroscience, molecular biology, and economics. It is one of the strongest known determinants of major life outcome such as educational attainment, occupational success, health and longevity (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson, 1997b; Gottfredson & Deary, 2004; Neisser et al., 1996; Schmidt & Hunter, 2004). Over the past several decades, developments in multivariate statistical modeling coupled with the availability of large data sets collected in twins and relatives have allowed for the examination of the genetic and environmental etiology of individual differences in intelligence, and the more recent advances in genotyping and DNA sequencing have enabled the search the for specific genetic variants underlying the observed variation (e.g., Benyamin, Pourcain, et al., 2013b; Davies et al., 2011; Franić, Groen-Blokhuis, et al., 2013; Najmabadi et al., 2007). The findings emerging from twin and family studies have univocally indicated 1) a role of genetic factors in the etiology of intelligence (e.g., Bouchard & McGue, 1981; Deary, Spinath, & Bates, 2006; Plomin et al., 2008; Plomin & Spinath, 2004), and 2) an age-dependent pattern of heritability, with individual differences in late adolescence and adulthood being more strongly influenced by genetic factors than those in childhood (the heritability estimates typically ranging from ~20% in infancy to ~40-50% in middle childhood and ~60-80% in adulthood; e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & Van Baal, 1998; Deary et al., 2006; Haworth et al., 2009a; R. A. Hoekstra et al., 2007; McGue, Bouchard Jr, Iacono, & Lykken, 1993a; Petrill et et al., 2004; Plomin, 1986; Polderman et al., 2006). Environmental factors that contribute to similarity between family members (e.g., shared family environment) typically decline in etiological relevance throughout childhood and adolescence, while environmental factors that facilitate differentiation between family members appear to play a persistently modest to moderate role (e.g., Bartels et al., 2002; Boomsma & Van Baal, 1998; Haworth et al., 2009a). The temporal stability of intelligence (i.e., the conservation of the rank order of individuals over time) is estimated to be fairly high, with around 45-60% of the variance in childhood being preserved over any given ~2-year interval (e.g., Bartels et al., 2002). This continuity in the observed individual differences is attributable predominantly to genetic factors, i.e., to the expression of a single set of genes throughout development (e.g., Bartels et al., 2002; Bishop et al., 2003; Eaves et al., 1986; R. A. Hoekstra et al., 2007; Petrill et et al., 2004; Rietveld et al., 2003). In addition to contributing to stability, genetic factors also generate change: age-specific genetic factors emerge at different ages, partly accounting for the lack of complete temporal stability. Environmental influences shared amongst family members, insofar as they are relevant, contribute mostly to stability, whereas the unshared environment contributes predominantly to change.

The aim of the present study is to contribute to the existing body of literature by providing one of the most comprehensive examinations of the genetic and environmental etiology of the observed stability of intelligence to date. We analyzed longitudinal twin data collected in four different studies on a total of 1748 twins, measured across a developmental period spanning childhood and adolescence (5-18 years of age). In contrast to many of the previous examinations of the genetic and environmental stability of intelligence (but see R. A. Hoekstra et al., 2007; Rietveld et al., 2003), we examine the stability of verbal and nonverbal abilities separately. In addition, we examine the stability of general intelligence (*g*; Jensen, 1998; Spearman, 1904). Because the choice of the psychometric instrument used to assess intelligence is inevitably dependent on the age of the participant, and because we combine data from four different studies (comprising 14 different subprojects), there is considerable heterogeneity in the measurement instrument used to assess intelligence across

the different samples and ages. This is not dissimilar to the situation in many other data registries, where longitudinal measures are often collected using different instruments across the life span. In twin registries in particular, this issue becomes especially prominent in the context of gene-finding studies (e.g., Flint, 2013; Goldstein et al., 2013a; Visscher et al., 2012), where specific genetic variants contributing to the variation in the observed trait (i.e., the phenotype) are sought. Here, the definition of the 'observed trait', or phenotype, is of considerable relevance (e.g., van der Sluis et al., 2010): how does one define a single 'observed trait' to be used in the analyses, given multiple measures over time? The presence of longitudinal data collected using different psychometric instruments allows us to address the auxiliary question of how to optimally utilize the existing twin registry data on intelligence in the context of gene finding studies, i.e. of whether data summarization is likely to diminish the power to detect genetic effects (see e.g. Medland & Neale, 2010; Minică et al., 2010; van der Sluis et al., 2010).

In summary, the present study aims to 1) assess the observed stability of verbal abilities, nonverbal abilities, and general intelligence, and 2) study the observed stability as a function of the underlying genetic and environmental factors. The structure of the dataset allows for an evaluation of how the results replicate and integrate across the different samples, and the presence of measures collected using multiple psychometric instruments allows us to address the practical question of how to optimally utilize the existing data in the context of gene-finding studies. Although the terms 'intelligence' and 'cognitive ability' have each been given a multitude of definitions (e.g., Jensen, 1998; Spearman, 1904), in the present chapter we use the two terms interchangeably.

Method

Sample

The data were obtained from the Young Netherlands Twin Register (YNTR; van Beijsterveldt, Groen-Blokhuis, Hottenga, Franić, et al., 2013). YNTR is a population-based register of Dutch twins born after 1986, recruited at birth and measured longitudinally at ages 1 through 18. The sample consisted of 1748 twins (including 872 complete twin pairs; 399 monozygotic (MZ) and 473 dizygotic (DZ)), and was composed of four longitudinally measured subsamples (sample sizes: 544, 226, 552, and 426 individuals). A detailed structure of the data is given in Figure 1 in Appendix 4. The twins were measured longitudinally at ages 5-18. This generated 4641 data points in total: 1946, 808, 1076, and 811 data points were available for the four subsamples, respectively. 47.5% of the participants were male.

Measures

Cognitive abilities were assessed longitudinally, using the Revised Amsterdam Children Intelligence Test (RAKIT; Bleichrodt, Drenth, Zaal, & Resing, 1984), Wechsler Intelligence Scale for Children (WISC-R and WISC-III; Sattler, 1992; Van Haasen et al., 1986; Wechsler et al., 2002), Raven's Standard and Advanced Progressive Matrices (SPM, APM; Raven, Raven, & Court, 1998; Raven, 1960), and the Wechsler Adult Intelligence Scale (WAIS; Stinissen, Willems, Coetsier, & Hulsman, 1970; Wechsler, 1997), the choice of test being largely dependent on the participants' age. Subscale scores were derived following the guidelines in the tests' manuals (Bleichrodt et al., 1984; Sattler, 1992; Van Haasen et al., 1986; Wechsler et al., 2002; Stinissen et al., 1970; Wechsler, 1997): for RAKIT, a verbal (V) and a nonverbal (NV) score were defined; for the WISC and the WAIS, the Verbal Comprehension Index

(VCI), Perceptual Organization Index (POI), and Freedom from Distractibility Index (FDI) were defined. For Raven's SPM and APM, the total score (defined as the total number of items answered correctly) was used in the analyses. Because the variances of the subscale scores across the different tests were quite heterogeneous in magnitude, to ease subsequent computation we standardized by dividing each variable by the product of its standard deviation and √5. This resulted in variances of an equal order of magnitude across the different tests.

Approach

Genetic covariance structure modeling (Martin & Eaves, 1977) is the application of structural equation modeling (Bollen, 1989; Kline, 2005) to data collected in genetically informative samples, such as samples of twins (Franić et al., 2012; M. C. Neale & Cardon, 1992). In the classical twin design, the sample consists of MZ and DZ twin pairs. DZ twins share 50% of their segregating genes on average, while MZ twins share nearly their entire genome (Falconer & Mackay, 1996; J. P. van Dongen et al., 2012). The covariance structure of the phenotypes (i.e., observed traits) is typically modeled as a function of latent factors representing several sources of individual differences: additive genetic (A), shared environmental (C), and individual-specific environmental (E) sources[28]. Additive genetic influences are modeled by one or more A factors, which represent the total additive effects of genes relevant to the phenotype. Based on quantitative genetic theory (Falconer & Mackay, 1996; Mather & Jinks, 1971), the A factors are expected to correlate 1 across MZ twins and .5 across DZ twins. Environmental influences affecting the phenotype of both twins in an identical way, thereby increasing their similarity beyond what is expected based on genetic resemblance alone, are represented by one or more C factors. Therefore, by definition, the C factors correlate unity across twins (regardless of zygosity). All environmental influences causing phenotypic differences among family members to differ are represented by one or more E factors. Thus, by definition, the E factors are uncorrelated across twins. Assuming an ACE model, the expected covariance structure in a multivariate twin model is thus:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E & r_A\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C \\ (r_A\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C)^t & \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_E \end{bmatrix}$$

where, given p phenotypes, $\boldsymbol{\Sigma}_{11}$ ($\boldsymbol{\Sigma}_{22}$) is the p x p covariance matrix of twin 1 (twin 2), $\boldsymbol{\Sigma}_{12}$ ($\boldsymbol{\Sigma}_{21}$) is the twin 1 - twin 2 p x p covariance matrix, and $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_C$ and $\boldsymbol{\Sigma}_E$ are the additive genetic, shared environmental, and unique environmental p x p covariance matrices, respectively. The coefficient r_A is the correlation between the additive genetic factors in twin 1 and twin 2 (1 in MZ and .5 in DZ twins).

In the present study, the temporal stability of intelligence (i.e., the stability of individual differences in performance on intelligence tests over time) and the temporal stability of genetic and environmental influences on intelligence (i.e., the degree to which the observed stability is attributable to the continuity of the genetic/environmental factors that affect intelligence over time) were modeled using the simplex model (Guttman, 1954; Jöreskog, 1970). An example of a simplex model is depicted in Figure 1. In this model, the

[28] In addition, the trait may be influenced by non-additive genetic factors (D), which include genetic interactions within the same locus (genetic dominance) or across different genetic loci (epistasis). In the present chapter, non-additive genetic effects were not modeled because the classical twin design does not allow for the simultaneous estimation of A, C, and D effects, and both the existing literature and our preliminary analyses favored an ACE over an ADE model.

data at occasion t (t = 1…T) are regressed on data at the preceding measurement occasion (t – 1), and the regression coefficient $\beta_{t,t-1}$ obtained in this regression is used as an indicator of temporal stability. For instance, a high β in the regression of verbal abilities at age 7 on verbal abilities at age 5 would indicate that the individual differences in verbal abilities are highly stable across this age span, i.e., that the rank order of individuals is largely preserved. Thus, the variance of a measure at a given time point is modeled as a function of factors that are stable over time (e.g., the variance at time point t is a function of the variance at time point t-1 and of the regression coefficient $\beta_{t,t-1}$: $\sigma^2_{t-1}*\beta_{t,t-1}^2$) and newly-emerging factors that affect the phenotype at the given time point but were absent at the preceding time point. The variance of a measure at time point t can thus be expressed as: $\sigma^2_t = \beta_{t,t-1}^2*\sigma^2_{t-1} + \zeta_t$, where ζ_t denotes the variance due to innovation. A high $\beta_{t,t-1}$ in combination with low ζ_t indicates high temporal stability; conversely, a low $\beta_{t,t-1}$ and a high ζ_t indicate low stability, implying that the factors relevant to the phenotype at time t-1 decrease in relevance by time t, and newly emerging factors gain in relevance.

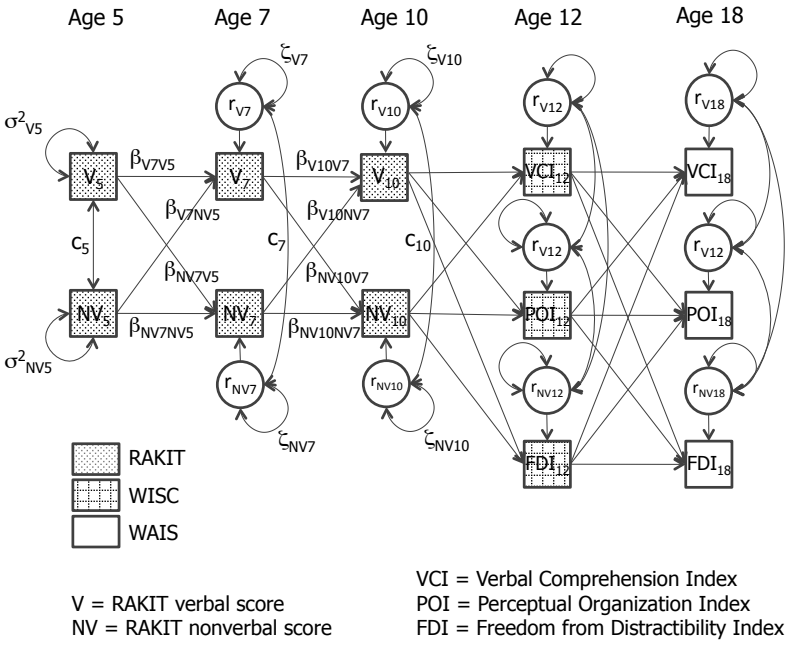

Figure 1. Phenotypic simplex model fitted to the data in Sample 1. Subscale scores on the RAKIT, WISC, and WAIS at five measurement occasions are modeled. For simplicity, parameter notation is only given for the first three measurement occasions. σ^2=variance, ζ=residual variance, c=(residual) covariance, β=regression coefficient. 'c' denotes covariance (between V5 and NV5) at the first measurement occasion, and residual covariance (i.e., covariance between the innovation factors) at subsequent measurement occasions.

In a simplex model with p observed variables, the expected p x p covariance matrix **Σ** equals $(\mathbf{I} - \mathbf{B})^{-1}\ \mathbf{\Psi}\ (\mathbf{I} - \mathbf{B})^{-1t}$, where **I** is a p x p identity matrix, **B** is a p x p matrix containing the autoregressive coefficients (βs) in the model, and **Ψ** is a p x p matrix containing the variances and covariances (for the first measurement occasion) and the residual variances

and covariances (for all the subsequent measurement occasions) of the observed variables. Thus, for the first three measurement occasions in Figure 1:

$$\Psi = \begin{bmatrix} \sigma^2_{V5} & c_5 & 0 & 0 & 0 & 0 \\ c_5 & \sigma^2_{NV5} & 0 & 0 & 0 & 0 \\ 0 & 0 & \zeta_{V7} & c_7 & 0 & 0 \\ 0 & 0 & c_7 & \zeta_{NV7} & 0 & 0 \\ 0 & 0 & 0 & 0 & \zeta_{V10} & c_{10} \\ 0 & 0 & 0 & 0 & c_{10} & \zeta_{NV10} \end{bmatrix},$$

where σ2 denotes variance, ζ denotes residual (innovation) variance, and c denotes (residual) covariance. Further,

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{V7V5} & \beta_{V7NV5} & 0 & 0 & 0 & 0 \\ \beta_{NV7V5} & \beta_{NV7NV5} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{V10V7} & \beta_{V10NV7} & 0 & 0 \\ 0 & 0 & \beta_{NV10V7} & \beta_{NV10NV7} & 0 & 0 \end{bmatrix},$$

where $\beta_{t,t-1}$ is the regression coefficient in the linear regression of a variable at time *t* on a variable at time *t*–1 (e.g., β_{V7V5} denotes the regression of variable V_7 on variable V_5).

To assess the contributions of genes and the environment to the observed stability and change in intelligence scores, a genetic adaptation of the simplex model was used (Boomsma, Martin, & Molenaar, 1989; Boomsma & Molenaar, 1987; Franić et al., 2012; M. C. Neale & Cardon, 1992). In genetic adaptations of the simplex model, in contrast to modeling a single time series, the phenotype is modeled as a function of several (genetic and environmental) latent time series. For instance, in a model containing only additive genetic and unique environmental latent factors (AE model; Figure 2)[29], the phenotypic variable V measured at age t, V_t, is related to the additive genetic and unshared environmental factors A_t and E_t (t=1,…,T), and simplex models, or first order autoregressions, are specified to account for the stability and change at the level of A and E (e.g., $\sigma^2_{Et}=\beta_{Et,t-1}{}^2*\sigma^2_{Et-1}+\zeta_{Et}$) [30]. The expected covariance structure of the phenotype(s) is thus:

$$\begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_A + \mathbf{\Sigma}_E & \mathbf{r_A}\mathbf{\Sigma}_A \\ \mathbf{r_A}\mathbf{\Sigma}_A & \mathbf{\Sigma}_A + \mathbf{\Sigma}_E \end{bmatrix},$$

where (assuming the latent factors are expressed on the same scale as the phenotype) the covariance matrices $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_E$ are modeled as follows:

$$\mathbf{\Sigma}_A = (\mathbf{I} - \mathbf{B}_A)^{-1}\mathbf{\Psi}_A(\mathbf{I} - \mathbf{B}_A)^{-1t},$$
$$\mathbf{\Sigma}_E = (\mathbf{I} - \mathbf{B}_E)^{-1}\mathbf{\Psi}_E(\mathbf{I} - \mathbf{B}_E)^{-1t}.$$

This means that one can assess the contributions of genetic and environmental factors to the observed stability and the change in stability. The phenotypic covariance between consecutive time points may be due to genetic influences ($\beta_{At,t-1} \neq 0$), environmental

[29] The Figure assumes that the variances of the latent innovation factors (ζ_A and ζ_E) are scaled to 1.

[30] Here, σ^2_{Et} and σ^2_{Et-1} are the variances of the unique environmental factors at times t and t-1, $\beta_{Et,t-1}$ is the regression of the E factor at time t on the E factor at t-1, and ζ_{Et} is the variance of the unique-environmental residual (i.e., innovation) at time t.

influences ($\beta_{Et,t-1} \neq 0$), or both ($\beta_{At,t-1} \neq 0$ and $\beta_{Et,t-1} \neq 0$). Likewise, any lack of stability may be due to either or both sources of individual differences. For instance, intermediate phenotypic stability (e.g., a correlation of .5) may be due to perfect genetic stability ($\zeta_{At} = 0$), in combination with complete environmental instability ($\beta_{Et,t-1} = 0$).

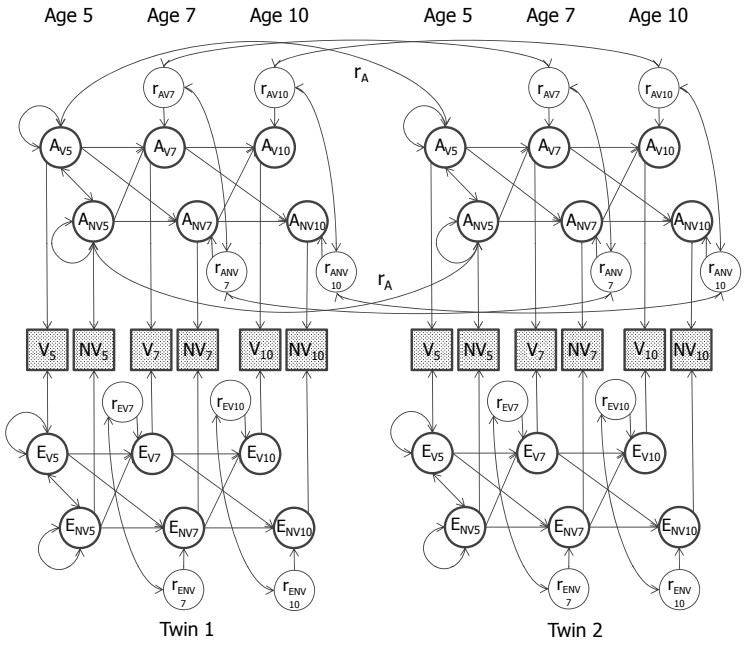

Figure 2. An example of an AE simplex model. Data observed at three measurement occasions are modeled as a function of additive genetic and unique environmental factors (A and E, respectively), and simplex models are specified to account for the stability and change at the level of A and E.

Analyses. The analyses were designed to examine the degree of phenotypic stability of intelligence, and assess the contributions of genes and the environment to the observed stability and change. This was achieved by fitting simplex models (described in detail below) to intelligence tests subscale scores: RAKIT V and NV, WISC and WAIS VCI, POI, and FDI, and Raven sum scores. In addition to modeling the subscale scores, the stability of general cognitive ability (*g*; Jensen, 1998, Spearman, 1904) was assessed. The *g* factor was defined as a first-order factor underlying performance on the different subscales at a given age, and the temporal stability of *g* was examined on both the phenotypic, and the genetic and environmental level. Thus, overall, four different types of models were fitted: a) phenotypic simplex models, b) phenotypic simplex models with a *g* factor, c) ACE simplex models, and d) ACE simplex models with a *g* factor. These models were fitted to each of the four samples separately, resulting in 16 distinct sets of results. To accommodate for any possible mean differences across the sexes, means were modeled separately for males and females in all analyses.

Phenotypic simplex models. The phenotypic simplex models fitted to each of the four samples are depicted in Figure 3. Here, age is given on the x-axis; thus the spatial distance between the variables at the different measurement occasions corresponds to the temporal

distance between these measurement occasions. The temporal distance between the measurement occasions is important, because the interpretation of the stability parameters depends on it: with equal stability over time, the *estimate* of stability between two measurement points decreases as a quadratic function of the temporal distance between those measurement points. For instance, with a constant stability of .5 between any two measurement points 2 years apart (e.g., from age 4 to age 6, from age 6 to age 8, etc.), the stability estimate would be $.5^2 = .25$ if one were to estimate the stability between two measurement points 4 years apart (e.g., age 4 to age 8). Therefore, the stability estimate is a function not only of the underlying stability, but also of the temporal distance between the measurement points used for estimation. Thus fact will be brought in mind while interpreting the results.

In all of the models depicted in Figure 3, each of the subscales measured at a given time point is specified to predict each of the subscales measured at the subsequent time point. Thus, not only the main regression paths (e.g., from RAKIT V at age 5 to RAKIT V at age 7), but also the cross-paths (e.g. from RAKIT V at age 5 to RAKIT NV at age 7), are estimated. In other words, any possible temporal contribution of one domain of intelligence to another is assessed. Although in Figure 3 we depict the phenotypic models for all the four samples, in subsequent text (the phenotypic model with a *g* factor, the ACE simplex model and the ACE model with a *g* factor) the models will only be illustrated for Sample 1. The structure of the models for the remaining samples, however, can be deduced from Figure 3: for instance, in the ACE simplex model for Sample 2, the structure of each of the three variance components (A, C, and E) is equal to the phenotypic structure for Sample 2 depicted in Figure 3.

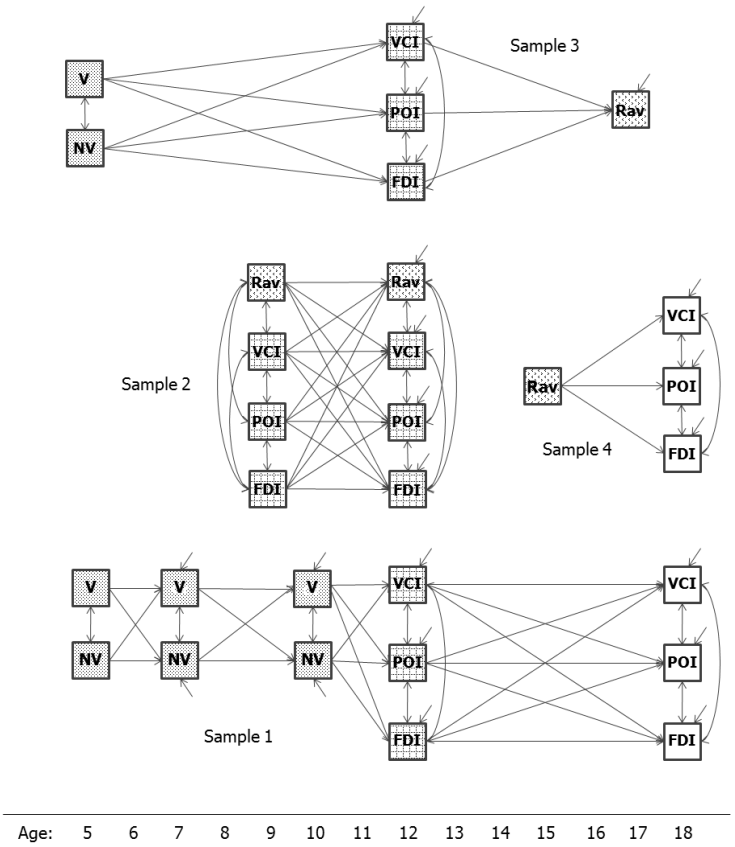

Figure 3. Phenotypic simplex models fitted to the four samples. The spatial distance between the measurement points within a model corresponds to their temporal distance.

Phenotypic simplex models with a g *factor*. In addition to assessing the phenotypic stability at the subscale level (Figure 3), a series of models assessing the phenotypic stability at the level of general cognitive ability (*g*) was fitted. Here, *g* was defined as a first-order latent factor underlying overall subscale performance at a given age, and autoregressions were specified to account for the stability and change at the level of *g* (upper right panel Figure 4). In addition to the autoregressions at the level of *g*, simplex models were also specified to account for the stability and change at the level of subscale-specific abilities, i.e., the residuals in the model: the Verbal scores at *t*-1 predict the Verbal scores at *t*, the Nonverbal scores at *t*-1 predict the Nonverbal scores at *t*, and the VCI, POI, and FDI factors at *t*-1 predict the VCI, POI, and FDI factors at *t*, respectively (upper right panel Figure 4).

ACE simplex models. In addition to assessing phenotypic stability, a series of ACE simplex models was fitted to the subscale scores in order to assess the contributions of genes and the environment to the observed stability. An example model (fitted to Sample 1) is depicted in the lower left panel of Figure 4. To avoid clutter in the Figure, the A, C, and E components are depicted separately; however, the three components are part of the same model, in which the subscale scores are modeled as a function of genetic and environmental latent series.

ACE models with a g *factor*. In addition to modeling subscale scores, the contributions of genes and the environment to the observed stability in the *g* factor were assessed. This was achieved by modeling the subscale scores at each time point as a function of underlying genetic and environmental *g* factors (A_g, C_g, and E_g), and specifying simplex models to account for the stability and change at the level of these genetic and environmental factors, as depicted in the lower right panel of Figure 4. Although the A, C, and E components of the model are depicted separately, they are part of the same model. In addition to modeling the genetic and environmental *g* factors, we assessed the extent to which any possible subscale-specific stability is due to genetic/environmental factors, by fitting an ACE simplex model to the subscale residuals (lower right panel Figure 4).

Results

For concision, the results pertaining to Sample 1 are presented in detail, while the results pertaining to the other three samples are summarized and discussed in view of their compatibility with the results in Sample 1. The full list of results (i.e., the parameter estimates obtained for all the four samples) is given in Appendix 4A. Figure 4 displays the results obtained for Sample 1. For ease of interpretation, the results we present are fully standardized, i.e., the variance of each (observed and latent) variable is 1. Stability is expressed as the proportion of variance of a variable at age t explained by the variables at t – 1; this proportion is easily obtainable by subtracting the magnitude of innovation variance from the total variance, i.e., as $1-\zeta_t$. A different standardization (allowing for a comparison of the relative magnitude of the A, C, and E variance components) is presented at the end of the Results section. The stability of the different subscales at a given age was largely comparable; thus, whenever possible, we describe general trends. When this is not warranted, we address the stability of the subscales separately.

Phenotypic simplex model. The temporal stability of intelligence subscales, as assessed using a phenotypic simplex model (upper left panel Figure 4), is in the intermediate range, varying from 34% to 66% in Sample 1. Averaging over the subscale stabilities at each given age gives the mean stabilities of 38%, 43%, 43%, and 54% at the age intervals 5-7, 7-10, 10-12, and 12-18, respectively, indicating that the phenotypic stability of intelligence increases with age. This is especially evident if one considers that the time interval between the last two measurement points (ages 12 to 18) is more than twice the average time interval between the remaining consecutive measurement points, and that the correlation between measurement points is expected to decrease as an exponential function of their temporal distance, given equal stability over time. Thus, with stability being constant over age, one would expect a drop in the stability estimate from the observed 43% in the 10-12 interval to around 3.5% in the 12-18 interval; however, the actual stability estimate in the 12-18 interval is a high 54%, indicating a sharp increase in stability over this period. The cross-lag regression coefficients (e.g., RAKIT V to RAKIT NV) were generally small in magnitude compared to the main regression coefficients (e.g., RAKIT V to RAKIT V); estimates of variance explained by any single cross-lagged relationship ranged from 0.2% to 5.8% (see Figure 4 for estimates). Notably, the stability remained moderate to high despite the use of different tests (RAKIT, WISC, and WAIS).

In Sample 2, the average subscale stability at age 12 was 40%; an estimate comparable to the 43% stability at the same age interval in Sample 1. In Sample 3, the average subscale stabilities between the ages of 5-12 and 12-17 were 18% and 44%, respectively. An estimate of 18% in the 7-year interval prior to age 12 implies that, were the time intervals equal to

those in Sample 1 (an average of 2.3 years prior to age 12), the stability would be estimated at 42%; highly consistent with the estimate obtained in Sample 1. The estimate of 44% in the 5-year interval between the ages of 12 and 17 implies that the stability would equal 58%, given a test-retest interval comparable to that of Sample 1 (i.e., 2.5 years). Thus, the temporal stability of intelligence as estimated in Sample 3 increases with age, and is consistent in both its magnitude and its observed increase with that estimated in Samples 1 and 2. In Sample 4, the mean subscale stability between the ages 15 and 18 is estimated at 30%. This is lower than the estimates obtained for the other samples; however, in Sample 4 the Raven sum score alone is used as a predictor of the three WAIS subscales (Figure 3). Thus, while the 30% estimate may reflect a lower temporal stability, it is also attributable to the relatively low correlation between the WISC subscales and the Raven.

Overall, the phenotypic subscale analyses indicate moderate to high stability of individual differences in intelligence across childhood and adolescence. The stability increases with age; i.e., the individual differences in intelligence become increasingly stable as individuals transition from childhood to adolescence. Notably, the stability remains in the intermediate to high range despite the variation in the instruments used to assess intelligence, and the results replicate well despite the differences in tests and measurement intervals across the four different samples.

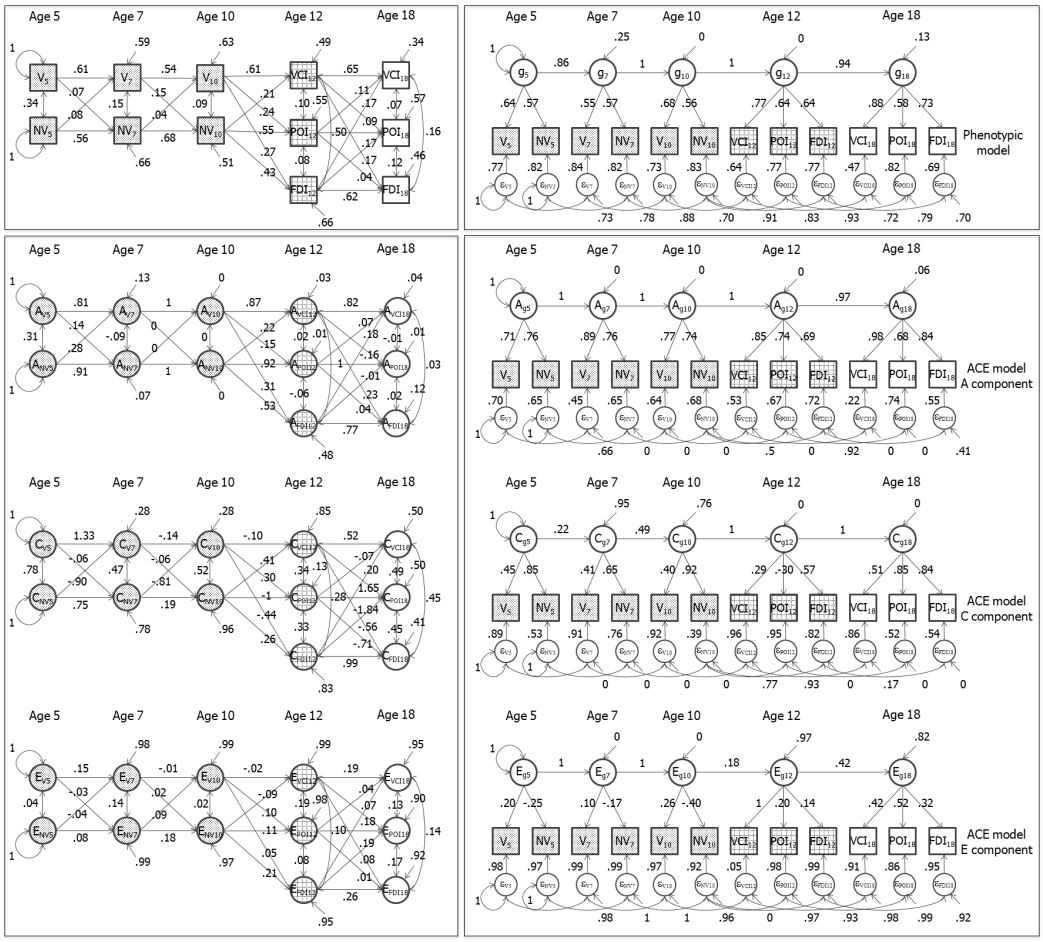

Figure 4. Parameter estimates obtained for Sample 1. Top left: phenotypic simplex model; top right: phenotypic simplex model with a *g* factor; bottom left: ACE simplex model; bottom right: ACE simplex model with a *g* factor. The results are completely standardized, i.e., the total variance of each (latent and observed) variable in the model is 1. In the right panels, the numbers in the bottom of the figures denote residual innovation variance (ζ), rather than residual regression coefficients (βs). The residual bs are not depicted, but may be inferred from the residual variances (i.e., ζs). To minimize clutter in the figure, residual covariances are depicted as double-headed arrows connecting the observed variables (or the genetic/environmental components thereof), rather than the residuals.

Phenotypic simplex model with a g *factor*. The upper right panel of Figure 4 shows the phenotypic simplex model with a *g* factor fitted to Sample 1. On average, the *g* factor explained around 37%, 31%, 38%, 47%, and 55% of subscale variance at ages 5, 7, 10, 12, and 18, respectively (possibly indicating an increasing role of g in intelligence over time, but also possibly reflecting the differences in the tests used). The temporal stability of the *g* factor is remarkably high: nearly the entire inter-individual variation at a given age can be predicted by the variation at the preceding age. The residual, subscale-specific variation displays a modest degree stability over time: 20% on average. It should, however, be noted that this is

a lower estimate of residual stability, as the estimates of the subscale-specific variation also include measurement error.

In Sample 2, the *g* factor explained an average of 47% and 37% of subscale variance at ages 9 and 12, respectively. The stability of *g* from age 9 to age 12 was around 80%, and the stability of the residual scores was modest (~15%), as in Sample 1. In Sample 3, the stability estimates were somewhat lower (42% and 65% at intervals 5-12 and 12-17, respectively). Note, however, that stability estimates of 42% and 65% over 7 and 5 years, respectively, imply that the stability would have been estimated at around 75% and 80%, respectively, had the time intervals been comparable to those of Samples 1 and 2 (2-2.5 years). In Sample 4, the stability was estimated at 60%. Again, it should be borne in mind that in Sample 4, Raven alone was used as a predictor of all three WAIS subscales; therefore, the lower stability estimate may reflect a lower temporal stability in Sample 4, but may also be due to the relatively low correlation between the Raven sum score and the WAIS subscale scores.

In summary, the phenotypic stability of *g* over childhood and adolescence is high, and exceeds the stability of individual subscales. The *g* factor explains around 30%-55% of subscale variance, regardless of the test used. Across all four samples, the stability of the subscale-specific scores is modest (around 15%-20%).

ACE simplex model. The lower left panel of Figure 4 shows the parameter estimates obtained for the ACE simplex model fitted to Sample 1. As evident from the Figure, the additive genetic influences on intelligence are highly stable; the stability estimates range from approximately 90% to 100%, and display an increase with age. Therefore, the genes that influence intelligence in early childhood overlap largely, if not entirely, with those that affect it throughout childhood and adolescence. Cross-lag regressions across measurement points and residual correlations within measurement points are fairly low, indicating that the genetic factors affecting e.g. verbal abilities are largely distinct from those affecting e.g. non-verbal abilities, both within and across measurement points. The stability of common environmental influences is in the intermediate range, and differs per subscale: the common environmental stability of verbal abilities before the age 10 is considerably higher than the common environmental stability of non-verbal abilities in this period; at later ages, however, the difference in stability between the subscales appears to disappear. However, as the magnitude of C component is small and decreases over time (see end of the Results section), the apparent differences in subscale stabilities are likely attributable to the unreliability of the relevant parameters. The unique environmental influences display virtually no stability over time; the stability estimates are close to zero at all time points.

In Sample 2, the additive genetic influences are highly stable (over 90% on average), the common environmental stability is high (~85% on average), and unique environmental stability is low (12% on average). Similarly, in Sample 3, the additive genetic stability is high (close to 100% except for the FOI subscale at age 12, the stability of which is estimated at 38%), the common environmental stability is estimated at 47% in the 5-12 interval and 100% in the 12-17 interval, and the E stability is virtually zero. In Sample 4, the A influences are estimated to be around 80% stable, the C influences around 60% stable, while the E influences display virtually zero stability.

Overall, the results indicate a high additive genetic stability (largely 90%-100%), a moderate to high common environmental stability, and a complete absence of unique environmental stability for both verbal and nonverbal abilities. The cross-subscale (e.g., verbal-nonverbal) stability is consistently low.

ACE simplex model with a g factor. The ACE simplex model with a *g* factor fitted to Sample 1 is shown in the lower right panel of Figure 4. In this sample, the additive genetic *g* factor explains around 60% of the additive genetic subscale variance and displays nearly perfect stability; 100% at most time points. Similarly, the additive genetic subscale residuals generally display a high temporal stability. The common environmental *g* factor explains around 40% of the common environmental subscale variance. The stability of common environmental influences appears to increase after age 10: the stability estimates are 5%, 24%, 100%, and 100% at ages 7, 10, 12, and 18, respectively. However, the magnitude of the C (and the E) variance is relatively small, and thus the reliability of the C (and the E) stability parameters is likely low. Overall, the residual C stability is estimated to be high. The unique environmental component displays an opposite pattern to the common environmental component: the stability of *g* before the age 10 is high, and declines substantially thereafter. However, the unique environmental *g* factor explains only 16% of the unique environmental subscale variance on average; the rest is explained by the subscale-specific E factors, which display virtually no stability.

In Sample 2, the additive genetic *g* factor explains 52% of the additive genetic subscale variance, and is 70% stable on average. Similarly, the residuals are highly stable (85%). The C_g factor displays complete stability, and explains 90% of the C subscale variance. The E_g factor explains only 24% of the unique environmental subscale variance, and is 30% stable on average, with highly unstable residuals. In Sample 3, the additive genetic *g* factor explains 70% of the subscale variance and is 93%-100% stable. The C_g factor explains around 60% of the C subscale variance, and declines in stability from 100% at ages 5-12 to 34% at ages 12-17. Again, however, the variance in the C stability estimates is likely due to the small magnitude of C. The E subscale variance was only modestly explained by E_g (~25%), and displayed stability neither at the *g* level, nor at the residual level. In Sample 4, the A_g, C_g, and E_g factors explained around 76%, 52%, and 11% of their respective variance, and were 100%, 100%, and 16% stable, respectively.

In summary, the A_g, C_g, and E_g factors explained an average of ~65%, ~60%, and ~20% of the A, C, and E variance, respectively. The A_g factor was highly stable over time (mostly close to 100%), with highly stable residuals. The C_g factor was generally highly stable (close to 100%), with some exceptions (ages 5-10 in Sample 1 and ages 12-17 in Sample 3; however, considering the small magnitude of the C variance component, these exceptions likely reflect the unreliability of the estimates). The E_g factor displayed modest stability (around 35% on average), but explained only around 30% of the E variance, the remainder of the variance being entirely unstable (close to 0%) across all samples.

Magnitude of variance components. The relative magnitude of the A, C, and E variance components, as estimated in the ACE simplex models and averaged over subscales at each age, is depicted in Figure 2 in Appendix 4. An age-related increase in heritability accompanied by a relative decline in common environmental variance, expected based on the literature (e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & Van Baal, 1998; Deary et al., 2006; Haworth et al., 2009a; R. A. Hoekstra et al., 2007; McGue et al., 1993a; Petrill et et al., 2004; Plomin, 1986; Polderman et al., 2006), is evident in Samples 1, 3, and 4. In Sample 2, where only two measurement points were available (ages 9 and 12), this trend was not apparent. This lack of trend can presumably be attributed to the brevity of test-retest time interval.

Integrated results. Figures 5 and 6 depict estimates of standardized variance components and A, C, and E stabilities, respectively, obtained across all four samples and shown for verbal

and nonverbal abilities separately. Unlike data in Figure 6, the data in Figure 5 did not appear to show considerable deviations from linearity; therefore the general trends are represented by linear regression lines weighted by sample size in Figure 5 and by a smoothing function (lowess function as implemented in R; R Core Team, 2013) in Figure 6. Consistently with Figure 2 in Appendix 4, an increase in the relative magnitude of additive genetic variance accompanied by a decrease in common (and, to some extent, unique) environmental variance is evident from Figure 5. Figure 6 indicates an increase in stability of all three components over time, and suggests that the observed phenotypic stability is driven primarily by additive genetic factors, with unique environment contributing primarily to change. Note that, for comparability, Figure 5 re-expresses the stability estimates on a scale on which all measurement points are equidistant (6 years). As explained earlier, the stability estimates are dependent on the time interval one uses for estimation and therefore the absolute magnitude of stability estimates is not interpretable in itself. The choice of time interval used to re-express the estimates is therefore arbitrary; the reason a 6-year period was chosen in this case is the fact that, with smaller (e.g., 1-year) time intervals the stability estimates reach an upper bound, making it impossible to distinguish between the stability of the different variance components (i.e., the C and E stability estimates increase, whereas the A stability estimates readily hit the upper bound of 1). Finally, Table 1 gives all available estimates of the phenotypic, genetic and environmental correlations obtained under an ACE simplex model, for verbal and nonverbal abilities separately. Again, it is evident that the observed stability of intelligence is driven primarily by additive genetic factors, with common environment contributing both to stability and change, and the unique environment predominantly generating change.

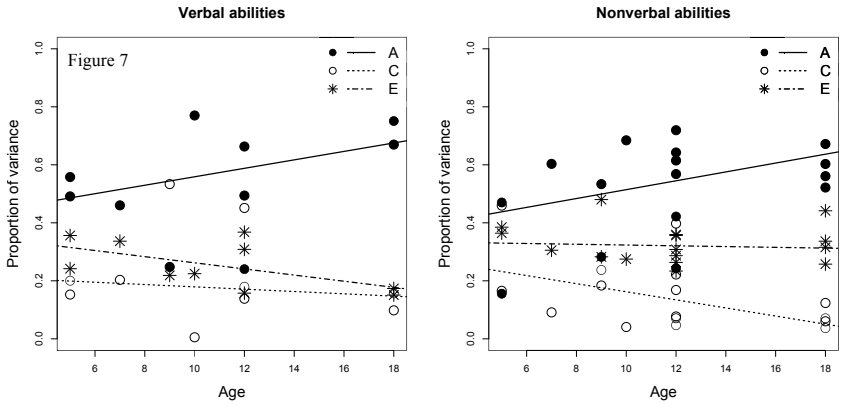

Figure 5. The relative magnitude of the A, C, and E variance components (y-axis) as a function of age, for verbal (left) and nonverbal (right) abilities. All available estimates from the four samples are included. Regression lines (weighted by sample size) represent the general trends.

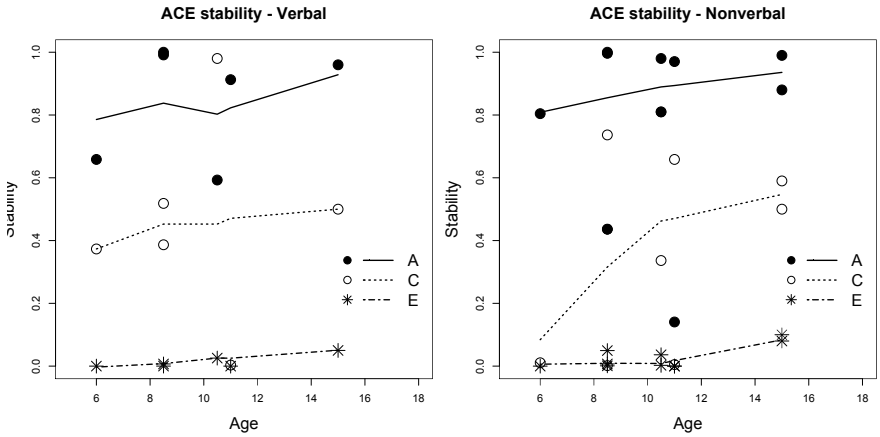

Figure 6. The ACE stability of verbal (left) and nonverbal (right) abilities. All available estimates from the four samples are included, and re-expressed on a scale on which all measurement points are equidistant (6 years). Lines (locally weighted scatterplot smoothing functions) represent general trends.

Discussion

The present study examined the stability of verbal abilities, nonverbal abilities, and general intelligence across childhood and adolescence, and assessed the genetic and environmental etiology of this stability. Other questions included the feasibility of combining results on multiple types of intelligence tests administered in a longitudinal design with the aim of utilizing the combined score in the context of gene-finding studies, and the relationship between different types of intellectual abilities over time (and the genetic/environmental etiology thereof).

The results indicate an intermediate to high phenotypic stability of individual differences in intelligence across the developmental period under study, with an increase in stability as individuals transition from childhood to adolescence. General intelligence, defined as a first first-order latent factor underlying subscale performance at a given age, explained around 30-55% of variance in subscale performance and displayed high temporal stability, exceeding that of individual subscales. The phenotypic stability appears to be driven primarily by genetic factors: the additive genetic influences were highly to entirely stable. The environment shared by family members appeared to contribute to stability to a moderate degree, while environmental factors unique to family members contributed mainly to innovation (i.e., to temporal instability). Similarly, the observed stability in the *g* factor was driven primarily by genetic factors: the additive genetic *g* factor displayed near complete stability, the common environmental *g* factor was generally stable but explained less of the phenotypic variance than the A_g factor, while the unique environmental *g* factor was modestly stable but explained only a minor fraction of the phenotypic variance in *g*. An age-related increase in heritability accompanied by a relative decline in common environmental variance, expected based on the literature (e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & Van Baal, 1998; Deary et al., 2006; Haworth et al., 2009a; R. A. Hoekstra et al., 2007; McGue et al., 1993a; Petrill et et al., 2004; Plomin, 1986; Polderman et al., 2006), was observed. In addition, the cross-subscale stability was consistently low, indicating a small to non-existent contribution of one domain of intelligence to another over time.

Table 1

The phenotypic, genetic and environmental correlations obtained in the four samples under an ACE simplex model. The correlations are given for verbal (below diagonal) and nonverbal (above diagonal) abilities separately

Σ_A	5	7	9	10	12	15	17	18
5		.96		.95	.79			.82
7	.89			1	.84			.81
9					.94			
10	.90	1			.83			.84
12	.91	.97	.75	.97				.95
15								
17								
18	.83	.94		.93	.96			

Σ_C	5	7	9	10	12	15	17	18
5		.28		.04	0			0
7	.63			.18	-.01			0
9					.59			
10	-.07	-.29			-.40			.36
12	-.29	.02	.95	.06				.20
15								
17								
18	00	.04		-.03	.70			

Σ_E	5	7	9	10	12	15	17	18
5		.08		.01	.04			0
7	.15			.18	.04			.01
9					.09			
10	0	0			.17			.05
12	.04	0	.14	-.03				.21
15								
17								
18	0	0		0	.21			

Σ	5	7	9	10	12	15	17	18
5		.58		.42	.33			.20
7	.64			.70	.43			.33
9					.58			
10	.38	.59			.58			.44
12	.34	.42	.69	.69				.68
15								
17								
18	.23	.36		.57	.79			

The stability of intelligence remained in the intermediate to high range despite the variation in the instruments used to assess it, and the results replicated well across the four samples despite the variation in tests and the time intervals used for estimation. The former relates to a common situation in data registries (e.g., twin registries), where data are often collected using a number of different psychometric instruments, the choice of test often being dependent on participants' age. Given the increased accessibility of genotyping and sequencing technologies and the consequent increase in the use of twin registry data in gene-finding studies, the question of how to optimally combine the existing longitudinal data in defining the phenotype for such studies is gaining in relevance. In this context, there are two prominent issues: 1) the actual modeling of a measured genetic variant in multivariate data; and 2) the accommodation of family members in the analysis. The latter does not pose a problem as the methods and software for family-based gene finding studies are well developed (e.g., Chen & Abecasis, 2007; Lippert et al., 2011; Minică et al., 2013; Minică, Dolan, Kampert, Boomsma, & Vink, 2014; Purcell et al., 2007). The former is potentially more problematic as full multivariate phenotypic modeling of family data is not computationally feasible, or perhaps even desirable. There are many possible loci of a genetic variant effect in a multivariate model, and therefore many possible models to consider. The present results, as pertaining to the longitudinal genetic covariance structure, suggest that a simple phenotypic sum score based on the repeated measures within a cognitive domain (e.g. verbal) should not result in any appreciable loss of information in a genetic association study (see Minică et al., 2010). Whether one should sum over cognitive domains is a different question. The genetic *g* factor accounted for about 60% of the genetic variance of the subtest scores. Summing over domains will only improve the power to detect a genetic variant if it contributes to this common genetic variance. Rather than running the risk of missing genetic variants, it is advisable to carry out gene-finding studies for each domain separately. One can still arrive at an omnibus test of the genetic variant (i.e., address the question of whether the genetic effect generalizes over domains) by combining the statistical results (van der Sluis, Posthuma, & Dolan, 2013).

While we believe that the high genetic stability provides a reasonable justification for summing over repeated measures within an individual, we note that this recommendation is limited in two ways. First, it applies to the present longitudinal results as obtained in the repeated measures design. From the point of view of power, a cross-sectional design may be preferable (and is certainly more efficient and cheaper to implement). However, the exact relationship between power and design is beyond the present scope (Minică et al., 2010 do consider different multivariate designs). Second, the recommendation is based strictly on the present choice to model the covariance structure by means of autoregressive and cross-lagged modeling. This approach is informative with respect to stability, but does not consider developmental change from the point of view of individual growth curves (Ramsden et al., 2011). We did not consider growth curve modeling as our IQ test scores were age-corrected, meaning that the present data were not informative with respect to individual developmental growth curves. Finally, the results present here were based on the standard genetic simplex model, in which A, C and E are assumed to be uncorrelated sources of individual differences. Whether this assumption (e.g., the absence of genotype-environment covariance) is valid to a reasonable approximation is an open question. Any genotype-environment covariance is unlikely to undermine our recommendations concerning data summarization in gene finding studies. However, a representation involving phenotype to environment transmission, typically envisaged as smart children contributing to their own "smart" environment (a.k.a. niche picking; Eaves, Last, Martin, &

Jinks, 1977) is possible (Dolan, de Kort, Kan, et al., 2014; Dolan, de Kort, van Beijsterveldt, Bartels, & Boomsma, 2014).

Chapter 8

Intelligence: Shared Genetic Basis between Mendelian Disorders and a Polygenic Trait

Abstract

Multiple inquiries into the genetic etiology of human traits indicated an overlap between genes underlying monogenic disorders (e.g., skeletal growth defects) and those affecting continuous variability of related quantitative traits (e.g., height). Extending the idea of a shared genetic basis between a Mendelian disorder and a classic polygenic trait, we performed an association study to examine the effect of 43 genes implicated in autosomal recessive intellectual disability on intelligence in an unselected Dutch population (N=1316). Using both single nucleotide polymorphism (SNP)- and gene-based association testing, we showed support for an association between intelligence and the genes of interest, with genes *ELP2*, *TMEM135*, *PRMT10*, and *RGS7* showing the strongest associations. This is the first demonstration of the functional relevance of genes implicated in monogenic disorders of intelligence on normal-range intellectual ability, and a corroboration of the utility of employing knowledge on monogenic disorders in identifying the genetic variability underlying complex traits.

Appendices can be obtained at http://sanjafranic.com/dissertation.

Based on: Franić, S., Groen-Blokhuis, M.M., Dolan, C.V., Kattenberg, M.V., Xiao, X., Scheet, P.A., Ehli, E.A., Davies, G.E., van der Sluis, S., Abdellaoui, A., Hansell, N.K., Martin, N.G., Hudziak, J.J., van Beijsterveldt, C.E.M., Swagerman, S., Hulshoff Pol, H.E., de Geus, E.J.C., Bartels, M., Ropers, H.H., Hottenga, J.J., & Boomsma, D.I. (2014) IQ: Shared Genetic Basis between Mendelian Disorders and a Polygenic Trait. *Under review.*

Multiple inquiries into the genetic etiology of complex human traits have shown that, for a number of phenotypes, the genetic variants affecting continuous, polygenic phenotypic variation may be concentrated in the same genes as those giving rise to monogenic (i.e., Mendelian) disorders. For instance, 180 loci associated with normal variation in the classic polygenic trait of adult height were shown to be enriched in genes underlying skeletal growth disorders (Allen et al., 2010). Many rare genetic variants in three candidate genes (*ABCA1*, *APOA1*, and *LCAT*), which give rise to pathogenically low levels of HDL-cholesterol in plasma, are also found in individuals with the common, complex version of the low-HDL-cholesterol trait (Antonarakis & Beckmann, 2006; Cohen et al., 2004; Frikke-Schmidt, Nordestgaard, Jensen, & Tybjærg-Hansen, 2004). Genes underlying Mendelian disorders of lipid levels and those affecting their normal concentration overlap almost entirely (Hirschhorn & Gajdos, 2011). Other examples include hemoglobin F levels (Hirschhorn & Gajdos, 2011), fat mass (Loos et al., 2008), type 2 diabetes (Hirschhorn & Gajdos, 2011; Sandhu et al., 2007; Winckler et al., 2007), and Parkinson's disease (Gasser, 2009; Lesage & Brice, 2009).

Genes underlying Mendelian disorders, in which protein functioning is severely altered, may therefore provide an opportunity to localize and understand the genetic variability that underlies susceptibility to a similar common polygenic phenotype (Antonarakis & Beckmann, 2006). In the present study, we utilize this idea to examine the effects of 43 genes implicated in autosomal recessive intellectual disability (Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, & Jamali, 2011) on intelligence in a Dutch sample from the general population (N=1316; see Methods and Figures 1-2 in Appendix 5A). Despite its being one of the most heritable human traits (with heritability estimates ranging from .6 to .8 in adolescence and adulthood; Deary, Johnson, & Houlihan, 2009; Plomin et al., 2008), no loci consistently associated with normal-range variation in intelligence have thus far been reported (Butcher, Davis, Craig, & Plomin, 2008; Chabris et al., 2012; Deary et al., 2009). The two largest genome-wide association studies (GWAS) to date failed to find replicable genome-wide association in SNP-based analyses in adults and children, respectively (Benyamin, Pourcain, et al., 2013b; Davies et al., 2011). On the other hand, hundreds of genes underlying monogenic disorders of intelligence have been identified (Inlow & Restifo, 2004; Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, & Jamali, 2011; Najmabadi et al., 2007).

The 43 genes used in the present study are a subset of the genes identified in a recent study that used homozygosity mapping, exon enrichment and next-generation sequencing in consanguineous families with autosomal-recessive intellectual disabilities to identify single, presumably disease-causing variants in 50 novel candidate genes (Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, & Jamali, 2011). The genome-wide dataset of the Netherlands Twin Register (NTR; van Beijsterveldt, Groen-Blokhuis, Hottenga, Franić, et al., 2013), used in the present study, contains SNP data on 43 out of these 50 genes (Table 1), including 1227 genotyped SNPs in total (see Appendix 5C).

Method

Sample

The data were obtained from the Netherlands Twin Register (NTR; Boomsma et al., 2006; van Beijsterveldt, Groen-Blokhuis, Hottenga, Franić, et al., 2013). The NTR is a population-based register of Dutch twins born after 1986, recruited at birth and measured longitudinally at ages 1 through 18. The sample consisted of 1316 individuals from 662

families (978 twins, 231 siblings, and 107 of their parents). To keep the genetic within-family covariance matrix approximately compound symmetric (i.e., to keep the genetic covariances between each type or relatives approximately equal), the data were selected so as to contain no complete MZ twin pairs and no more than one parent per family. Thus, each family consisted of individuals who are genetically either siblings or parent-offspring, i.e., the expected genetic correlation between any given pair of family members was .5. The observed intraclass correlation between the family members was .57 (s.e.=.025). 45.8% of the sample were males. The mean ages of children and parents were 12.7 (sd=4.1) and 43.9 (sd=4.1), respectively. The age distribution (showing each participant's mean age across measurement occasions) is given in Figure 1 in Appendix 5A.

Phenotype data

Intelligence was assessed longitudinally using the Revised Amsterdam Children Intelligence Test (RAKIT; Bleichrodt et al., 1984), Wechsler Intelligence Scale for Children (WISC; Sattler, 1992; Van Haasen et al., 1986; Wechsler et al., 2002), Raven's Standard and Advanced Progressive Matrices (SPM, APM; Raven et al., 1998; Raven, 1960), and the Wechsler Adult Intelligence Scale (WAIS; Stinissen et al., 1970; Wechsler, 1997), the choice of test being largely dependent on participants' age. A previous study employing the same dataset found high genetic stability of intelligence scores as assessed by the different tests (the autoregressive coefficients between the additive genetic factors at consecutive measurement occasions ranging from .8 to 1; Franić et al., 2014). Therefore, the individuals' mean scores across the different ages were used as measure of the phenotype. The IQ scores were derived based on the age- and sex-appropriate norms for the RAKIT, WISC, or WAIS, and subsequently converted to z-scale within each measurement occasion and averaged over measurement occasions. For 154 participants, we used z-transformed scores on Raven's Matrices. The distribution of intelligence scores is given in Figure 2 in Appendix 5A.

Genotype data

Blood and/or buccal samples for DNA extraction were collected as part of several projects within the NTR. Genotyping was performed using the Affymetrix Human SNP Array 6.0. Genotypes were called using the BIRDSEED V2 algorithm. SNPs in Hardy-Weinberg equilibrium ($p>.00001$) with a minor allele frequency exceeding .01 and a missingness rate below 5% were included in the analyses. Samples were selected if their call rate exceeded 95% and were checked for Mendelian errors, excessive heterozygosity ($-.1<F<.1$), and discrepancies in relatedness (Scheet et al., 2012). Genotypes displaying Mendelian inheritance errors were excluded from the analyses.

For the present study we selected all genotyped SNPs from the 50 genes of interest, including a 5 kb border around each gene. 7 out of the 50 genes contained no genotyped SNPs. The distribution of the SNPs (1227 in total) over the remaining 43 genes is shown in Table 1. The full list of SNPs is given in Appendix 5C.

SNP-based analyses

As a first step, we tested for an association between the phenotype and each of the 1227 SNPs. As the observations were clustered in families, the analyses were performed using a multilevel regression model with random intercepts to account for the within-family covariance structure. Specifically, the model for phenotype of person *i* in family *j* was: ph_{ij} =

$b_{0j} + b_1*SNP_{ij} + res_{ij}$, where *ph* denotes phenotype, b_{0j} is intercept in family *j*, b_1 is a (fixed) slope parameter, and res_{ij} denotes an individual-specific residual term. The intercept term can be further decomposed as $b_{0j} = g_0 + k_{0j}$, where g_0 is a fixed component and k_{0j} is a component that is random over families. Using random intercepts prevents the inflation of type I error associated with applying a standard (fixed-effects) regression model to family-clustered data. The within-family genetic covariance structure was approximately compound symmetric (i.e., the expected genetic correlation between any given type of relatives was .5). The analyses were implemented using the "nlme" package in R (R Core Team, 2013). The code used to carry out the analyses is given in Script 1 in Appendix 5B.

Additionally, we performed association testing using the Plink software package (Purcell et al., 2007). Here the association between the phenotype and each of the 1227 SNPs was examined using the Huber-White sandwich variance estimator to account for the family structure in the data. The results were compared to those obtained using the multilevel regression model in R. A high degree of correspondence between the results obtained using the multilevel regression model (which effectively assumes an AE background covariance structure among first degree relatives) and those obtained using the Huber-White sandwich estimator (which corrects for relatedness without assuming a background model) would imply that any background misspecification in the random effects model has not affected the conclusions. A high degree of correspondence is expected, because the test of a fixed effect in the multilevel regression model is fairly robust to possible background misspecification (Minică et al., 2013).

To empirically evaluate the results obtained for the 1227 SNPs, we drew a number of random samples of: a) 1227 SNPs from the entire genome, b) 1227 SNPs from intragenic regions of the genome, and c) 43 genes (including all SNPs on those genes) from the entire autosomal genome. All samples excluded the 1227 SNPs of interest. Each of the random samples was subjected to the analyses described above. The resulting QQ plots and genomic inflation factors (λ) were compared to those obtained for the 1227 SNPs of interest.

As additional verification of the results, permutation testing was employed to generate an empirical distribution of λ values under the null hypothesis of no association. The genotypes (i.e., the 1227 SNPs) were randomly reallocated over the phenotypes 1000 times, and each of the 1000 permuted datasets was analyzed using the random intercept multilevel regression model described above. To account for the background covariance structure arising from the clustering of data in families, family data were relocated jointly: the genotypes of any 2-member family were reassigned to phenotypes of another randomly selected 2-member family, and the same was done for 3- and 4-member families. Thus, the family structure in the permuted datasets remained intact. As in the original analysis, the family structure was subsequently corrected for using a multilevel model. The null distribution of λ values generated using the permuted datasets was compared to the λ obtained for the 1227 SNPs of interest.

Finally, a genome-wide association study was performed. Here, the phenotype was regressed on each of the available genotyped SNPs (538652 SNPs) using the Plink software package (Purcell et al., 2007).

All analyses were performed using an additive model and included 8 genotypic principal components (Abdellaoui et al., 2013a) as covariates to account for the possible effects of population stratification. All l values were estimated as regression coefficients of the observed on the expected $-\log_{10}$ of the p-values, using the GenABEL package in R (R Core Team, 2013).

Gene-based analyses

Next, the SNP-based p-values obtained using the multilevel regression model were used as input for gene-based analysis. A gene-based association test that employs the extended Simes procedure (GATES) was used (M.-X. Li, Gui, Kwan, & Sham, 2011). GATES involves jointly analyzing all available SNPs in a gene to obtain a single p-value associated with the gene. The method assumes that an association test between the phenotype and all available SNPs on the gene has been carried out, and that the resulting p-values and pair-wise allelic correlation coefficients *r* for all SNPs are available. In the present case, we used the p-values obtained in the SNP-based multilevel regression analysis, and pair-wise allelic correlation coefficients obtained using the *--r* option in Plink. Given *m* SNPs on a gene, a gene-based p-value is obtained through an iterative procedure by combining the ascendingly ordered *m* p-values in the following way: $P_G = \min (m_e p_{(j)} / m_{e(j)})$, where m_e is the effective number of independent p-values among the *m* SNPs, $m_{e(j)}$ is the effective number of independent p-values among the top *j* SNPs (j = 1, …, m), and p_j is the *j*-th lowest p-value (i.e., the p-value associated with the *j*-th top SNP). The null hypothesis of this gene-based test is that none of the SNPs are associated with the phenotype; the alternative is that at least one SNP is associated. The effective number of independent p-values among the *m* SNPs, m_e, is estimated as $m_e = m - \sum_{i=1}^{m} [I(\lambda_i>1)(\lambda_i-1)]$ $\lambda_i>0$, where $I(x)$ is an indicator function and λ_i is the *i*-th eigenvalue of the *m* x *m* correlation matrix (ρ) of the p-values obtained in the SNP-based association test. The pair-wise p-value correlation coefficient, ρ_{ij}, can be approximated by a 6th order polynomial function of the allelic correlation coefficient r_{ij}: $\rho_{ij} = .2982r_{ij}^6 - .0127r_{ij}^5 + .0588r_{ij}^4 + .0099r_{ij}^3 + .6281r_{ij}^2 - .0009r_{ij}$, where ρ_{ij} and r_{ij} are the *ij*-th elements of the SNP p-value correlation matrix ρ, and of the allelic correlation matrix *r*, respectively. For a full overview of the method, we refer the reader to the original publication (M.-X. Li et al., 2011) and to Script 2 in Appendix 5B, which contains our implementation of GATES in R. The R script performs the test *k* times given *k* genes in the input file.

Additionally, we performed a gene-based association test using the Versatile Gene-Based Test for Genome-wide Association Studies (VEGAS; J. Z. Liu et al., 2010), and compared the results to those obtained using GATES. VEGAS is a simulation-based method that uses information from the full set of SNPs within a gene and accounts for linkage disequilibrium (LD) by using simulations from the multivariate normal distribution. The analyses were performed using the VEGAS web-interface (J. Z. Liu et al., 2010).

Results

SNP-based analyses

Association between intelligence scores and each of the 1227 SNPs (see Methods) was examined using an additive model and 8 principal components (Abdellaoui et al., 2013a) to account for the possible effects of population stratification (Script 1 in Appendix 5B). The left panel of Figure 1 shows a quantile-quantile (QQ)-plot, including 95% confidence intervals (CIs)[31], of the association p-values (also see Figures 3 and 4 and Table 1 in Appendix 5A). As evident from the Figure, the distribution of the observed p-values differs markedly from that expected under the null hypothesis of no effect, indicating an enrichment of the 43 candidate genes for polymorphisms associated with intelligence. Note that the significant

[31] The CI estimates were obtained while taking into account the linkage disequilibrium (LD) structure between the SNPs: instead of N=1227, we used an estimate of the effective number of independent p-values (N=625).[22] This approach produces relatively broader CIs; we thus adopt a more stringent approach to evaluating the significance of the difference between the expected and the observed distributions.

inflation along nearly the entire length of the QQ plot (genomic inflation factor λ=1.26) is not necessarily indicative of population stratification, particularly in the context of a candidate SNP study. Here, the observed inflation is expected under the alternative hypothesis of (polygenic) effects of a relatively large number of the candidate SNPs tested (Yang, Weedon, et al., 2011). As the analyses were performed while adjusting for 8 principal components (seven of which were highly correlated with geographic latitude and longitude in the present sample, thereby feasibly representing differences in ancestry), population stratification does not appear to be a likely cause of the inflation. This was empirically verified in subsequent analyses.

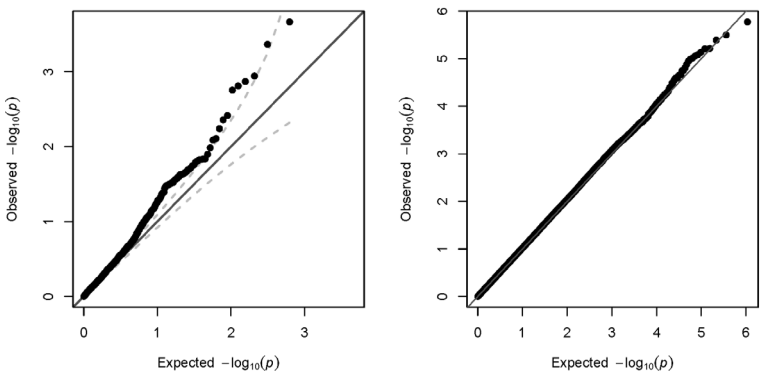

Figure 1. Left: QQ-plot based on the 1227 candidate SNPs. Right: Genome-wide QQ-plot based on 538652 SNPs. Dashed lines: 95% confidence intervals (CIs).

To empirically verify the present finding and confirm the absence of population stratification, we performed SNP-based association testing on samples of SNPs drawn randomly from the genome. In particular, we drew 1000 random samples of: a) 1227 SNPs from the entire genome, b) 1227 SNPs from intragenic regions of the genome, and c) 43 genes (including all SNPs on those genes) approximately matched for size with the 43 candidate genes and sampled from the entire autosomal genome. All random samples excluded the 1227 SNPs of interest. The distributions of the λ values obtained for each set of random samples, along with the λ obtained for the 1227 SNPs of interest (marked by a horizontal line), are depicted in Figure 2. As evident from the Figure (Panels a and b), the effect found for the SNPs of interest did not replicate in any of the 2000 random samples obtained by sampling SNPs from the entire genome or from the intragenic regions of the genome. For SNPs residing on randomly sampled sets of 43 genes (Panel c Figure 2), only 3.6% of λ values exceed the λ obtained for the candidate SNPs.[32]

As further empirical verification, we performed permutation testing to obtain an empirical distribution of λ values under the null hypothesis of no association: the genotypes (i.e., the 1227 SNPs of interest) were randomly reallocated over the phenotypes 1000 times, and each of the 1000 permuted datasets was analyzed using SNP-based association testing. The resulting distribution of λ values and the λ obtained for the non-permuted dataset

[32] Note that the higher variance in Panel c of Figure 2 relative to that in Panels a and b is expected given that the degree of non-independence of SNPs (i.e., linkage disequilibrium (LD)) is considerably higher in SNPs sampled from the same gene relative to those sampled from the entire genome. A reduced effective number of independent SNPs is expected to result in a less precise estimate of λ, i.e., in a higher dispersion around the mean λ value.

(λ=1.26) are shown in Panel d of Figure 2. Here, only 2.9% of the λ values exceed the λ value of interest; an empirical p-value highly consistent with that obtained from random sampling.

Finally, a genome-wide association analysis was performed. Here, the phenotype was regressed on each of the available genotyped SNPs (538652 SNPs). The resulting QQ-plot is depicted in the right panel of Figure 1. As evident from the Figure, the genome-wide p-values in the right panel show no notable inflation (λ=1.03), in contrast to the left panel (λ=1.26).

The present results thus unequivocally indicate an enrichment of the candidate set of genes for polymorphisms associated with intelligence, while plausibly ruling out population structure as the cause of the observed effect. The former is demonstrated by the significant inflation of the association p-values for the candidate set of SNPs as compared to random subsets of SNPs (empirical p-value p=.036) and to a permutation-based null-distribution (empirical p=.029). The latter is established by a) the inclusion of genetic principal components into the association study, b) the near absence of comparable p-value inflation in randomly selected sets of SNPs, and c) the absence of genome-wide p-value inflation.

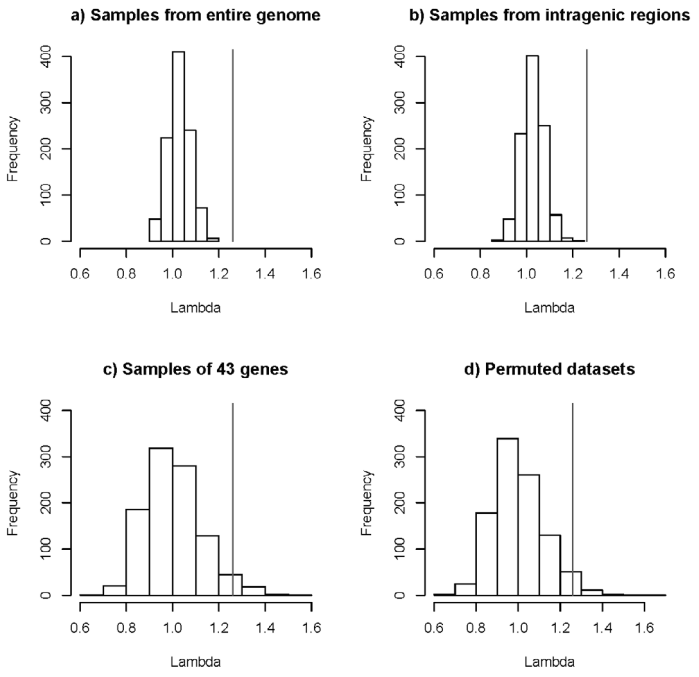

Figure 2. Distribution of genomic inflation factors (λ) obtained for 1000 a) random samples of 1227 SNPs from the entire genome, b) random samples of 1227 SNPs from intragenic regions of the genome, c) random samples of 43 genes from the entire genome, d) permuted datasets. Horizontal line: λ obtained for in the non-permuted dataset for the 1227 SNPs of interest.

Gene-based analyses

Next, gene-based testing was carried out (see Methods and Script 2 in Appendix 5B). The full list of gene-based results is given in Table 2 in Appendix 5A. Genes *ELP2* (p=.007), *TMEM135* (p=.007), *PRMT10* (p=.019), and *RGS7* (p=.044) displayed the strongest associations. Notably, two out of the 50 genes from the Najmabadi et al. (2011) study harbor more than one mutation associated with intellectual disability; one of those genes is ELP2, which, in the present study, shows the strongest evidence of association.

Focusing our interest on the top four genes, we examined the positions of the most strongly associated SNPs in these genes relative to the mutations in Najmabadi et al. (Figure 5, Appendix 5A). As evident from the Figure, both mutations in *ELP2*, as well as the mutations in *TMEME135* and *PRMT10*, are relatively close to our top SNP for their respective genes; the distances range from 4.8 kb to 31.4 kb. On *RGS7*, the distance between the mutation and the top SNP is relatively large (535.7 kb). Note that any distance between the disease-causing mutation and our top SNP is consistent with the logic of the present study however, as the gene is viewed as a functional unit with regard to its etiological relevance to intelligence, regardless of the distribution of the functionally relevant polymorphisms along the gene.

For validation, both the gene-based analyses and the SNP-based analyses were performed using several different methods, as implemented in different software packages (see Methods). The results obtained using the different methods converged highly: the $\log_{10}$ of the p-values obtained using two methods of SNP-based testing correlated .88, and the p-values obtained using two different gene-based tests correlated .89 (Table 2 in Appendix 5A).

Discussion

The present study demonstrated an enrichment of 43 genes implicated in autosomal recessive intellectual disability in consanguineous Iranian families (Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, & Jamali, 2011) for polymorphisms associated with normal-range intelligence in a Dutch population-based sample. This is the first demonstration of the functional relevance of genes implicated in monogenic disorders of cognitive ability on continuous variability in intelligence. Despite the high heritability of intelligence (Deary et al., 2009; Franić, Dolan, Borsboom, van Beijsterveldt, & Boomsma, 2010; Haworth et al., 2009a; Plomin et al., 2008), the progress in the identification of loci consistently associated with variation in its normal range has thus far been limited (Benyamin, Pourcain, et al., 2013b; Butcher et al., 2008; Chabris et al., 2012; Davies et al., 2011; Deary et al., 2009; Need et al., 2009). Exceptions are the apolipoprotein E (*APOE*) gene at older ages (Deary et al., 2002) and formin binding protein 1-like (*FNBP1L*), the latter having recently been shown to be associated with both childhood and adulthood intelligence (Davies, et al., 2011; Benyamin, et al., 2013). The present approach utilizes the idea that the differentially sized effects of individual mutations located within a gene functionally relevant to the phenotype may range from severe disruptions of protein functioning resulting in a Mendelian disorder, to smaller effects underlying polygenic variation. Utilizing pre-existing knowledge on genetics of Mendelian disorders may therefore prove a valuable approach to the identification of genetic variability underlying polygenic traits, with the advantage of requiring considerably smaller sample sizes than GWAS. This may prove especially useful in the study of phenotypes for which large samples are difficult to obtain, for instance because the phenotype is difficult or costly to measure (e.g., neuropsychological or fMRI measures), and/or in detection of genetic

variants characterized by small effect sizes. For instance, in the present study we clearly demonstrate enrichment, although none of the p-values for individual SNPs fall below a Bonferroni-corrected significance threshold (a=.05/1227=.00004, or a=.05/625=.00008 if one corrects by the number of independent SNPs (M.-X. Li et al., 2011)), indicating that the magnitudes of individual SNP effects are too small to be detected in regular GWAS.

Although larger sample sizes are needed to identify the exact genes and genetic variants driving the association in the present study, we tentatively focus on the top four genes that reach nominal significance. The most strongly associated gene, *ELP2* (elongator complex protein 2), encodes a subunit of the RNA polymerase II elongator complex, involved in acetylation of histones H3 and probably H4 and possibly in chromatin remodeling (Fellows, Erdjument-Bromage, Tempst, & Svejstrup, 2000). *TMEM135* (transmembrane protein 135) is involved in fat metabolism and energy expenditure (Exil et al., 2010). *PRMT10* (protein arginine methyltransferase 10) affects chromatin remodeling leading to transcriptional regulation, RNA processing, DNA repair and cell signaling (Fisk et al., 2010). *RGS7* (regulator of G-protein signaling 7) interacts with 14-3-3 protein, tau and snapin (a component of the SNARE complex required for synaptic vesicle docking and fusion) (Ilardi, Mochida, & Sheng, 1999).

The utilization of knowledge on monogenic disorders to identify polymorphisms that affect the variability of continuous phenotypes is a cost-efficient approach to understanding the genetic variability underlying polygenic traits. At present, the causal variants for a large number of monogenic disorders have been identified (over 3,000 disorders; Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nlm.nih.gov/omim), and recent developments in sequencing technologies have made it increasingly possible to employ exome sequencing or whole-genome sequencing, possibly in combination with homozygosity mapping, as an efficient approach to identifying novel causal variants underlying Mendelian disorders (Ku, Naidoo, & Pawitan, 2011). The National Human Genome Research Institute has opened Centers for Mendelian Genomics (NHGRI Genome Sequencing Program, http://www.genome.gov/), whose primary goal is the discovery of as yet unknown variation underlying Mendelian disorders. Thus, at present, the utilization of existing and impending knowledge on variants underlying Mendelian disorders to identify the variation underlying polygenic traits may prove a viable, efficient and cost-effective complement to standard approaches such as GWAS. The present finding highlights the importance of continuing the efforts directed at studying monogenic diseases (Ku et al., 2011; Ropers, 2010b) at a time when focus has shifted away from them, as they can advance our understanding of multifactorial traits.

Table 1
Chromosomal position (hg19), length, and number of genotyped SNPs for the 43 genes

Gene	Chr	Start (bp)	End (bp)	Length (kb)	N of SNPs
PARP1	1	224097741	226600780	2503.039	385
RGS7	1	240926554	241525530	598.976	177
TMEM135	11	86743886	87039756	295.870	83
LAMA1	18	6936743	7122813	186.070	68
FRY	13	32600437	32875794	275.357	67
ADK	10	75905960	76474061	568.101	48
SCAPER	15	76635526	77202785	567.259	48
ASCC3	6	100951070	101334248	383.178	31
PECR	2	216856052	216952678	96.626	27
POLR3B	12	106746436	106908976	162.540	24
ENTPD1	10	97449774	97642023	192.249	23
ACBD6	1	180239515	180477089	237.574	23
NDST1	5	149860381	149942773	82.392	18
ZBTB40	1	22773344	22862650	89.306	17
INPP4A	2	99056317	99212496	156.179	17
ELP2	18	33704407	33762909	58.502	16
TAF2	8	120738015	120850103	112.088	14
LINS	15	101094574	101148435	53.861	14
KDM5A	12	384223	503620	119.397	13
CACNA1G	17	48633429	48709835	76.406	13
SLC31A1	9	115978842	116034217	55.375	11
CAPN10	2	241521133	241562122	40.989	9
KIF7	15	90147020	90203682	56.662	8
RALGDS	9	135968107	136044301	76.194	8
WDR45L	17	80567438	80611411	43.973	8
GON4L	1	155714508	155834191	11.9683	8
C9orf86	9	139689818	139740639	50.821	6
TTI2	8	33325904	33376119	50.215	5
UBR7	14	93668401	93700561	32.160	4
ZCCHC8	12	122952417	122990518	38.101	4
CCNA2	4	122732599	122750087	17.488	3
C11orf46	11	30339598	30364774	25.176	3
FASN	17	80031215	80061106	29.891	3
PRMT10	4	148553936	148610381	56.445	3
MAN1B1	9	139976379	140008635	32.256	3
CNKSR1	1	26498894	26521377	22.483	3
HIST1H4B	6	26022124	26032480	10.356	2
HIST3H3	1	228607546	228618026	10.480	2
EEF1B2	2	207019309	207032652	13.343	2
CASP2	7	142980308	143009789	29.481	2
ASCL1	12	103346464	103359294	12.830	2
KDM6B	17	7738222	7763106	24.884	1
ERLIN2	8	37589117	37621619	32.502	1
				Total	1227

Chapter 9

Mendelian and polygenic inheritance of intelligence: a common set of causal genes?

Abstract

Despite twin and family studies having demonstrated a substantial heritability of individual differences in intelligence, no genetic variants have been robustly associated with normal-range intelligence to date. This is largely ascribed to the high polygenicity of intelligence, i.e., to its being subject to the effects of a large number of genes of individually small effect. Intellectual disability, on the other hand, frequently involves large effects of single genetic mutations, many of which have been identified. The present paper aims to 1) introduce the reader to the current state of genetic intelligence research, including next-generation sequencing and the analysis of rare genetic variants, and 2) examine the possible effects of known disability genes on normal-range intelligence. The rationale for the latter rests on the fact that genetic variants affecting continuous, polygenic traits are often concentrated in the same areas of the genome as those underlying related monogenic phenotypes. Using an existing pool of known intellectual disability genes, we constructed a set of 168 candidate genes for normal-range intelligence, and tested their association with intelligence in 191 individuals sampled from the high and low ends of the IQ distribution. In particular, we 1) employed exon sequencing to examine the possible effects of rare genetic variants in the 168 genes, and 2) used polygenic prediction to examine the overall effect of common genetic variants in the candidate gene set in a larger sample (N=2125). No significant association between the candidate gene set and intelligence was detected.

Based on: Franić, S., Dolan, C.V., Broxholme, J., Hu, H., Zemojtel, T., Davies, G.E., Nelson, K., Ehli, E.A., the Childhood Intelligence Consortium, Ropers, H.-H., & Boomsma, D.I. (2014) Mendelian and polygenic inheritance of intelligence: a common set of causal genes? Using Next-Generation Sequencing to examine the effects of 168 cognitive disability genes on normal-range intelligence. *Under review.*

Intelligence is one of the most frequently studied human behavioral traits and one of the strongest known determinants of major life outcomes such as educational attainment, occupational success, health and longevity (Deary, Johnson, & Houlihan, 2009; Deary et al., 2004; Gottfredson, 1997b; Gottfredson & Deary, 2004; Neisser et al., 1996; Schmidt & Hunter, 2004). Over the past century it has motivated research across a diverse range of fields including not only the behavioral sciences, but also neurosciences, molecular biology, economics, and genetics. Interestingly, behavior genetic studies of intelligence frequently converge on two seemingly incompatible findings. On the one hand, twin and family studies have demonstrated 1) a substantial genetic component of individual differences in intelligence (e.g., Bouchard & McGue, 1981; Deary et al., 2006; Plomin et al., 2008; Plomin & Spinath, 2004), and 2) an increase in the relative magnitude of this component across development (from around 20% in infancy, to ~40-50% in middle childhood and ~60-80% in adulthood (e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & Van Baal, 1998; Deary et al., 2006; Haworth et al., 2009b; R. A. Hoekstra et al., 2007; McGue, Bouchard Jr, Iacono, & Lykken, 1993b; Petrill et al., 2004; Plomin, 1986; Polderman et al., 2006). On the other hand, genetic association studies aiming to identify the genetic variation driving the observed individual differences have cumulatively identified genetic variants that explain less than 1% of the observed variability (Benyamin, Pourcain, Davis, Davies, Hansell, Brion, Kirkpatrick, Cents, et al., 2013; Chabris et al., 2012; Davies et al., 2011). This gap between the estimated heritability and the variance explained by known variants, frequently termed the 'missing heritability' (Maher, 2008), has been assigned a multitude of explanations, including the insufficient statistical power to detect genetic variants of small effect size, the potential overestimation of heritability by twin studies, problems pertaining to the measurement and operationalization of intelligence, and the possibility of genetic variants not tagged on present genotyping platforms (including rare and structural variation) underlying the heritability (see, e.g., Dickson, Wang, Krantz, Hakonarson, & Goldstein, 2010; Eichler et al., 2010; Goldstein et al., 2013b; Manolio et al., 2009; van der Sluis et al., 2010; Zuk, Hechter, Sunyaev, & Lander, 2012). The largest genome-wide association (GWA) studies to date identified no genetic variants robustly associated with intelligence, and only one gene, FNBP1L, has been tentatively implicated in the etiology of normal-range intelligence to date (Benyamin, Pourcain, Davis, Davies, Hansell, Brion, Kirkpatrick, Cents, et al., 2013; Davies, et al., 2011).

Recent years have seen an increase in the use of several additional approaches to addressing the missing heritability issue. Firstly, the development of the methodology for the estimation of heritability using measured genetic information, implemented in the genome-wide complex trait analysis tool (GCTA; Yang, Lee, Goddard, & Visscher, 2011), has enabled the estimation of the proportion of the variance in intelligence explained by the total additive effects of common genetic variants tagged on the present genotyping platforms. Ranging from ~22-46% in children and adolescents (Benyamin, Pourcain, et al., 2013a; Plomin, Haworth, Meaburn, Price, & Davis, 2013; Trzaskowski et al., 2014; Trzaskowski, Yang, Visscher, & Plomin, 2013; Trzaskowski, Shakeshaft, & Plomin, 2013) to ~29-51% in adults (Davies et al., 2011; Marioni, Davies, et al., 2014), the estimates of this proportion are substantially larger than the variance presently explained by the presently underpowered GWA studies. However, they remain lower than the estimates of the total genetic variance of intelligence. In addition, while demonstrating that a substantial proportion of the variance in intelligence is attributable to the additive effects of common genetic variation, GCTA estimates provide no information on the specific genetic variants associated with intelligence. Secondly, the recent advent of the large-scale use of sequencing technologies, which enable the measurement of the complete nucleotide sequence of a genome, has

opened a wealth of possibilities for the study of intellectual disability (much of which is monogenic, i.e., caused by a single genetic mutation), leading to the discoveries of many previously unknown genetic causes of cognitive impairment (e.g., Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, Jamali, et al., 2011, Najmabadi et al., 2007). For instance, DNA sequencing has enabled the identification of genes underlying a range of sporadic, syndromic conditions involving intellectual disability (e.g. Schinzel–Giedion syndrome, Kabuki syndrome; Hoischen et al., 2010; Ng et al., 2010), and many sporadic and familial causes of non-syndromic intellectual disability (see Topper, Ober, & Das, 2011). However, sequencing technologies are seldom employed to study the genetics of normal-range intelligence. This is partly due to the highly polygenic nature of intelligence (i.e., its being subject to a large number of very small genetic effects), and the consequent need for (often prohibitively) large samples to achieve sufficient statistical power for the detection of individual causal variants.

In the present study, we utilize the existing knowledge on the genetics of monogenic (i.e., Mendelian) disorders to construct a plausible set of candidate genes for normal-range intelligence. The study utilizes a simple rationale, namely the idea that the genetic variants giving rise to monogenic disorders may be localized in the same areas of the genome as those affecting continuous variation in related polygenic traits. Previous research has amply demonstrated the plausibility of this with respect to several other phenotypes. For instance, several genes causing monogenic forms of Parkinson's disease have been associated with the common, polygenic form of the disease (Gasser, 2009). Rare genetic variants in three candidate genes (ABCA1, APOA1, and LCAT), giving rise to pathogenically low levels of HDL-cholesterol in plasma, are also found in individuals with the common, polygenic version of the low-HDL-cholesterol trait (Cohen et al., 2004; Frikke-Schmidt et al., 2004). Other examples include height (Allen et al., 2010), body mass index (Loos et al., 2008), lipid levels (Hirschhorn & Gajdos, 2011), hemoglobin F levels (Hirschhorn & Gajdos, 2011), and type 2 diabetes (Sandhu et al., 2007). Genes underlying monogenic disorders, in which protein functioning is severely altered, may therefore provide an opportunity to localize the genetic variation underlying a similar, polygenic phenotype. Utilizing this idea, we sequenced the exons (i.e., expressed regions) of 168 genes known to underlie intellectual disability, and examined their association with intelligence in a sample of 191 individuals. By design, we focused on the detection of the possible effects of rare genetic variation. This is in line with the assumption of inter-individual variability in intelligence being maintained by low-frequency, disruptive mutations of small effect size (e.g., Hsu, 2012; Marioni, Penke, et al., 2014). Because selection on fitness-related traits, including intelligence, is expected to a) prevent mutations with large negative effects from becoming common in the population, and b) lead to an accumulation of mutations with large positive effects, resulting in their uniform presence in the population (as monomorphisms, i.e., non-variable DNA sites), the genetic architecture of intelligence is expected to be marked by the absence of genetic variants of large effect sizes. Mutations of small negative effects, however, are expected to linger at a low frequency (e.g., Hsu, 2012), and the genetic architecture of high intelligence may potentially be conceptualized as the absence of a large number of these disruptive mutations (e.g., Hsu, 2012; Marioni, Penke, et al., 2014). While the above argues in favor of rare deleterious variants, we also consider common variants, as these may be present in the form of effectively neutral mutations (subject to genetic drift), or as relatively positive mutations (subject to positive selection), which have yet to become fixed in the population. The possible effects of common genetic variants were tested for by examining whether polygenic scores (Purcell et al., 2007) summarizing the effects of common single-nucleotide polymorphisms (SNPs) in the 168 genes of interest were predictive of intelligence, in a larger

random sample of 2125 individuals. More details on the polygenic score prediction, next-generation sequencing and the analysis of rare variants can be found in the *Methods* section.

Method

Phenotype data

Data on psychometric intelligence were obtained from the Young Netherlands Twin Register (YNTR, Boomsma et al., 2006; van Beijsterveldt, Groen-Blokhuis, Hottenga, Franić, et al., 2013). YNTR is a population-based register of Dutch twins born after 1986, recruited at birth and measured longitudinally at ages 1 through 18. The sequenced sample consisted of 191 unrelated children and adolescents of Dutch ancestry (Abdellaoui et al., 2013b), aged 5-18 at the time of measurement. 46% of the participants were male. Intelligence was assessed longitudinally, using the Revised Amsterdam Children Intelligence Test (RAKIT; Bleichrodt et al., 1984), the Wechsler Intelligence Scale for Children (WISC; Sattler, 1992; Van Haasen et al., 1986; Wechsler et al., 2002), and the Wechsler Adult Intelligence Scale (WAIS; Stinissen et al., 1970; Wechsler, 1997), the choice of instrument being partly dependent on the participants' age. IQ scores were derived based on the respective age- and sex-appropriate norms for RAKIT, WISC, or WAIS, and subsequently converted to z-scale within each measurement occasion (i.e., within each time point used for assessment), and averaged over measurement occasions (i.e., across the different time points, within each participant). A previous study employing the same dataset found a high temporal stability of the additive genetic effects on intelligence (with the correlations between the additive genetic factors at consecutive measurement occasions ranging from .8 to 1; Franić et al., 2014), implying that the same genetic factors are relevant to intelligence over the developmental period under study (5-18 years of age). In situations of high genetic stability, averaging over the measurement occasions has been shown to be a sensible approach from the perspective of statistical power (Minică et al., 2010).

The scores of the 191 individuals belonged to the tails of the IQ distribution: individuals were included into the study if their IQ z-score exceeded .8 (~112 IQ points) or was below -.8 (~88 IQ points), but above -1.33 (~80 IQ points). The rationale for excluding the individuals with an IQ below 80 is the focus of the present study on non-monogenic inheritance, i.e., the fact that the genetic architecture underlying their intellectual (dis)ability may differ from that of individuals from the rest of the distribution. Several additional exclusion criteria were applied during sample selection. Participants were not included into the study if their IQ scores displayed excessive variation across the different measurement points (SD>1 on a z-scale) or differed excessively from the IQ scores of their family members ('excessive' being defined as a difference of ~18 and ~11 IQ points for monozygotic (MZ) and dizygotic (DZ) twins, respectively; these numbers correspond to a difference at least one standard deviation greater than the average twin difference in our sample). Additional exclusion criteria included low birth weight (under 1000g), known genetic defects, and discordance between IQ and educational attainment scores (individuals in the low IQ group were not included into the study if their educational attainment score on the Dutch national test of educational attainment (CITO, 2002) exceeded 539, i.e., belonged to the top 40% of the distribution). The IQ scores were dichotomized ('high' and 'low'; N=104 and N=87, respectively) for the first set of the analyses (gene-based testing).

In the second set of the analyses (polygenic prediction), all individuals from the Netherlands Twin Register (NTR; Boomsma et al., 2002; Boomsma et al., 2006; Willemsen et al., 2013) with psychometric intelligence and SNP microarray data were included into the

sample (N=2125, 45.4% male). The age distribution of the participants at the time of measurement is given in Appendix 6 (mean=20.4, SD=14.1). The testing and the computation of IQ scores were performed in the same way as above, with the exception of participants for whom only the scores on Raven's Progressive Matrices (Raven et al., 1998; Raven, 1960) were available; for these participants, a z-transformed number of correct answers, rather than a z-transformed IQ score, was analyzed. Unlike the sample used for exon sequencing (N=191), the larger sample was unselected on phenotype, i.e., the intelligence scores followed a normal distribution.

Next-generation sequencing

Nucleic acid sequencing is a set of methods used in the determination of the precise order of nucleotides in a nucleic acid molecule (see e.g. Grada & Weinbrecht, 2013 for a nontechnical overview). Initially accomplished through chain-termination methods (i.e., so-called Sanger or first-generation sequencing; Sanger et al., 1977), DNA sequencing is presently performed using a set of methodologies commonly denoted next-generation sequencing (e.g., Metzker, 2010; Rusk & Kiermer, 2008; Shendure & Ji, 2008). Next-generation sequencing is an umbrella term denoting a set of technologies (e.g., Illumina (Solexa) sequencing, Roche 454 sequencing, Ion torrent: Proton / PGM sequencing, SOLiD sequencing) that perform sequencing in a massively parallel fashion, sequencing millions of DNA fragments simultaneously. Unlike SNP microarrays that only measure common genetic variation (i.e., variants whose population frequency exceeds ~1%), sequencing technologies enable the interrogation of the entire nucleotide sequence of the genome, including rare and structural variation. The development of next-generation sequencing technologies was accompanied by a rapid decline in the cost of DNA sequencing, resulting in a sharp increase in the accessibility of sequence data over the past decade. In addition, next-generation sequencing coupled with efficient DNA capture (i.e., the isolation of specific DNA targets) has facilitated the emergence of exome sequencing as a novel approach to the identification of rare variants underlying polygenic phenotypes, and a cost-efficient alternative to whole-genome sequencing (see Kiezun et al., 2012). Exome sequencing (i.e., targeted exome capture) denotes the sequencing of the entire set of expressed regions of the genome (i.e., exome, comprising of all exons or 'EXpressed regiONs' of the genome), while exon sequencing refers to the sequencing of a particular exon or a set of exons. The present study employed the latter approach.

Genotype data

The genes examined in the present study were selected from the pool of genes presently known to underlie various forms of syndromic and non-syndromic intellectual disability (Najmabadi, Hu, Garshasbi, Zemojtel, Abedini, Chen, Hosseini, Behjati, Haas, Jamali, et al., 2011; Ropers, 2008, 2010a). The selection of the genes was partially guided by practical considerations, e.g., by the limited target size allowed by the HaloPlex G9901B exon enrichment kit, which was used to selectively capture the genomic regions of interest from DNA samples prior to sequencing. 107 of the 168 genes were autosomal. Table 1 and Appendix 6A provide an overview of the genes and their function. Exon sequencing was performed using an Illumina HiSeq2000 sequencer with 100bp paired-end reads. The raw reads were aligned to the NCBI37 human reference genome using the Stampy package (Lunter & Goodson, 2011). Variants were called using Platypus (Rimmer, Phan, Mathieson, Lunter, & McVean, 2013). The information on quality control and filtering of the genotype

data can be found in Supplementary Methods. Mean sequencing depth (i.e., the mean number of times each nucleotide base was sequenced) was ~212x.

The second set of analyses (i.e., polygenic prediction) employed all common SNPs (i.e., SNPs with a minor allele frequency exceeding 1%) in the 168 genes that were both a) measured in the 2125 individuals, and b) analyzed in a recent meta-analysis of childhood intelligence (Benyamin, Pourcain, et al., 2013a). The reason for applying the latter criterion is our subsequent use of the effect size estimates from the meta-analysis as weights in the construction of the polygenic predictor (see *Analyses*). In total, this resulted in 8559 SNPs from 99 autosomal genes being used in the polygenic prediction (8 out of the 107 autosomal genes were not present in the meta-analysis dataset). Information on imputation and quality control of the SNP data can be found in Supplementary Methods.

Analyses

Association testing. To examine the association between intelligence and the rare genetic variants in the genes of interest, we applied a series of gene-based association tests implemented in the PLINK/SEQ tool (https://atgu.mgh.harvard.edu/plinkseq/), using the data on 2900 variants as input. Gene-based as opposed to single-locus testing was used as a means of increasing statistical power (Kiezun et al., 2012; Purcell, Cherny, & Sham, 2003), seeing as the inherently small number of observations of rare variants limits the statistical power for their individual detection. Six gene-based association tests were employed: a burden test using adaptive permutation to test for excess of rare alleles in cases relative to controls (--*assoc* keyword in PLINK/SEQ), a test based on the count of case-unique rare alleles (--*uniq* command in PLINK/SEQ), a frequency-weighted test (see Madsen & Browning, 2009; --*fw* command), the variable threshold test (Price et al., 2010; --*vt* command in PLINK/SEQ), the c-alpha test (B. M. Neale et al., 2011; --*calpha*), and a sum of single-site statistics (--*sumstat*). Overall, the tests aim to assess the genetic burden due to the effects of rare genetic variants, working on the assumption that the phenotypic variation may be explained by the overall burden of rare deleterious mutations, while the individual causal variants may be heterogeneous and interchangeable.

The first test uses adaptive permutation to test for excess of rare alleles in the individuals in the low IQ group relative to those in the high IQ group. Permutation entails random re-allocation of genotypes over the phenotypes to generate an empirical distribution of p-values under the null-hypothesis, against which the p-value of interest can be compared. The permutation is adaptive in the sense that the variants that are highly unlikely to achieve statistical significance are dropped from the procedure. The second test is based on the count of alleles exclusive to the low end of the phenotypic distribution (i.e., low IQ). This strategy effectively eliminates common alleles from the test, because they would be present in individuals at both extremes unless they have a very large effect. The frequency-weighted test (similar to Madsen & Browning, 2009) scores each individual by a weighted sum of mutation counts within each gene. The weighing scheme assigns higher weights to variants that are rare in individuals from the high end of the phenotypic distribution (and thus presumably detrimental), effectively preventing common variants from dominating the test. Group counts (i.e., weighted sums in cases and controls) are compared, and permutation is used to evaluate the significance of the result. The variable threshold test (Price et al., 2010) is based on the regression of the phenotype on the genotype. The test assumes that there is an unknown threshold T, such that variants with a minor allele frequency below T are substantially more likely to have a functional effect than variants with a minor allele frequency above T. The test consists of computing a test statistic

using only the variants that fall below a certain minor allele frequency cutoff, for the full range of cutoffs. The final test statistic is subsequently defined as the maximum of the test statistics across all the cutoffs. By optimizing the test statistic in this way, the test effectively gives higher weight to variants predicted to be functionally significant (i.e., to variants that fall below a minor allele frequency cutoff that resulted in the best test statistic). All of the aforementioned tests entail the assumption of rare variants within a given gene acting in the same direction (either increasing or decreasing intelligence). The c-alpha test (B. M. Neale et al., 2011) does not involve this assumption, i.e., it accommodates possible differences in the direction of effect across the measured variants. The test assesses the imbalance in the distribution of alleles over cases and controls, such that, e.g., the risk variants are more present in cases and protective variants more present in controls. Under the null hypothesis of no effect, the risk and the protective variants are expected to be distributed randomly over the cases and controls. An excess of, for instance, a risk allele in the cases, would result in an overdispersion in the distribution of this allele. C-alpha assesses this overdispersion, regardless of its origin (risk or protective), and is ideally suited for detecting a mixture of effects, such that some variants confer risk while others are neutral or protective. As evident, all of the six tests focus on the detection of the possible effects of rare genetic variants. This is consistent with our expectation of rare variants being enriched for functional alleles, and exhibiting stronger effect sizes than common genetic variants (e.g., Frazer, Murray, Schork, & Topol, 2009; Kryukov, Pennacchio, & Sunyaev, 2007; Pritchard, 2001).

Correction for multiple testing was performed by dividing the desired significance threshold (.05) by the total effective number of independent tests in the study. The estimate of the number of independent tests was based on the number of genes for which PLINK/SEQ's I-statistic (i.e., estimate of the minimal achievable p-value for a gene) was smaller than .05, as genes with an I-statistic greater than .05 are considered insufficiently powered and thus necessitate no correction (Kiezun et al., 2012). Bonferroni correction would be too stringent in the present context, as it assumes that each gene displays sufficient variation to achieve the asymptotic properties for the test statistic (Kiezun et al., 2012); an assumption that is not necessarily realistic in the context of rare variant data and the present sample size. For genes on the X chromosome, in addition to being performed on the entire sample, the analyses were performed for the males and the females separately.

Polygenic prediction. Subsequently, we examined whether continuous intelligence scores in the larger (N=2125) sample can be predicted from a polygenic score constructed on the basis of the common SNPs in the candidate gene set. Here, the polygenic score is used as a means of summarization of genetic effects across the relevant genes: it is obtained as a weighted sum of the number of effect alleles within an individual, across all common SNPs in the candidate gene set. The weighing of the SNPs, and the determination of 'effect allele', were informed by prior knowledge: the weights were the effect size estimates for individual SNPs obtained in a large meta-analysis of GWA studies on childhood intelligence (Benyamin, Pourcain, Davis, Davies, Hansell, Brion, Kirkpatrick, RAM Cents, et al., 2013). The continuous intelligence scores were subsequently regressed on the polygenic scores. A significant regression coefficient would imply a genetic signal amongst the variants (see, e.g., Dudbridge, 2013).

The meta-analysis results were based on an analysis of six independent cohorts (combined N=12,441): the Avon Longitudinal Study of Parents and Children, the Lothian Birth Cohorts, the Brisbane Adolescent Twin Study, the Western Australian Pregnancy Cohort Study, and the Twins Early Development Study (Benyamin, Pourcain, Davis, Davies, Hansell, Brion, Kirkpatrick, RAM Cents, et al., 2013). The polygenic scores were constructed

by multiplying the number of effect alleles (0, 1, or 2) at a given locus in the present dataset by the meta-analysis regression coefficient for that locus, and summing the resulting scores over all relevant loci within an individual. The subsequent regression of intelligence on the polygenic scores was performed using generalized estimating equations ('gee' package in R; V. J. Carey, Lumley, & Ripley, 2012; Minică et al., 2014; R Core Team, 2013) to control for the dependency in the data arising from the fact that some individuals in the sample are closely genetically related (e.g., twins, parents). To control for possible spurious association arising from population stratification (i.e., from any possible systematic differences in allele frequencies between the high and low IQ groups due to differences in ancestry; see, e.g., Cardon & Palmer, 2003; Freedman et al., 2004; Price et al., 2006) and to remove any phenotypic variance associated with sex, sex and nine principal components reflective of the Dutch population structure (Abdellaoui et al., 2013b) were included into the regression as covariates. As different populations frequently exhibit systematic differences in allele frequencies, principal components of a genome-wide covariance matrix of the individuals' allelic values frequently reflect variation in ancestry, and are known to efficiently control for population stratification (e.g., Price, et al., 2006).

Results

The application of deep sequencing to the 168 genes of interest revealed 2900 point-mutations that passed quality control filters and differed from the reference dataset (Genomes Project Consortium, 2012) in at least one of the 191 DNA samples. Of these 2900 variants, 972 and 61 were observed only once and twice in the 191 samples, respectively. The frequency distribution of the 2900 variants is displayed in Figure 1. As evident from the Figure, ~70% and ~76% of the variants were observed in less than 10% and 20% of the samples, respectively. Around 50% of the exonic variants were synonymous (i.e., base substitutions did not affect the produced amino acid sequence), with the remainder being non-synonymous. The distribution of the variants over the 168 genes is displayed in Figure 2.

The QQ plots of the gene-based p-values obtained using the six association tests in Plink/SEQ are shown in Figure 3. An inflation of the QQ plot, i.e., an excess of low p-values relative to the uniform expectation, would indicate a possible genetic signal in the candidate set of genes. As visible in the Figure, no inflation was observed for any of the six gene-based tests. After correction for multiple testing, none of the individual genes displayed a significant association with intelligence. Sex-stratified analysis confirmed the absence of a detectable association for the genes on the X chromosome. In addition, the polygenic score was not predictive of intelligence (p=.69).

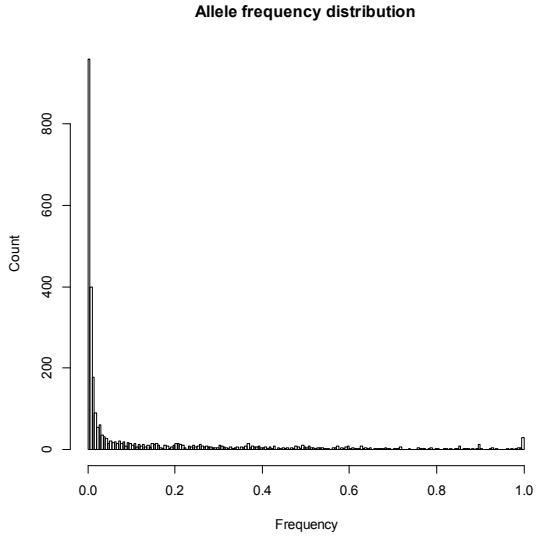

Figure 1. Distribution of non-reference allele frequencies[33] at diallelic sites (97.25% sites in the present dataset).

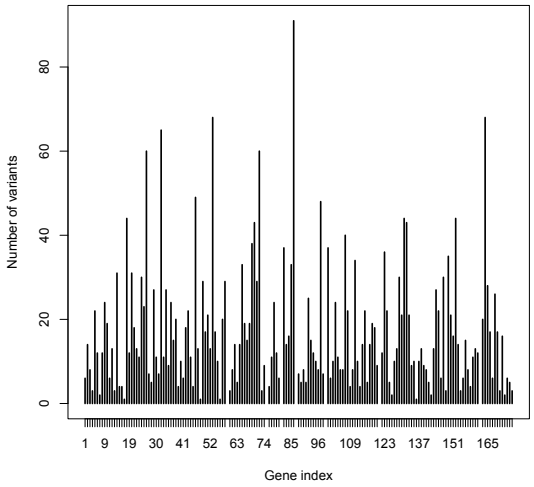

Figure 2. Distribution of the 2900 variants over the 168 genes. x-axis: gene index, y-axis: number of variants on the gene.

[33] Polymorphic sites in a DNA sequence can be described by the frequency of one of their alleles. The figure shows the distribution of these frequencies for all diallelic sites (i.e., all sites displaying two alternate forms) in the present dataset.

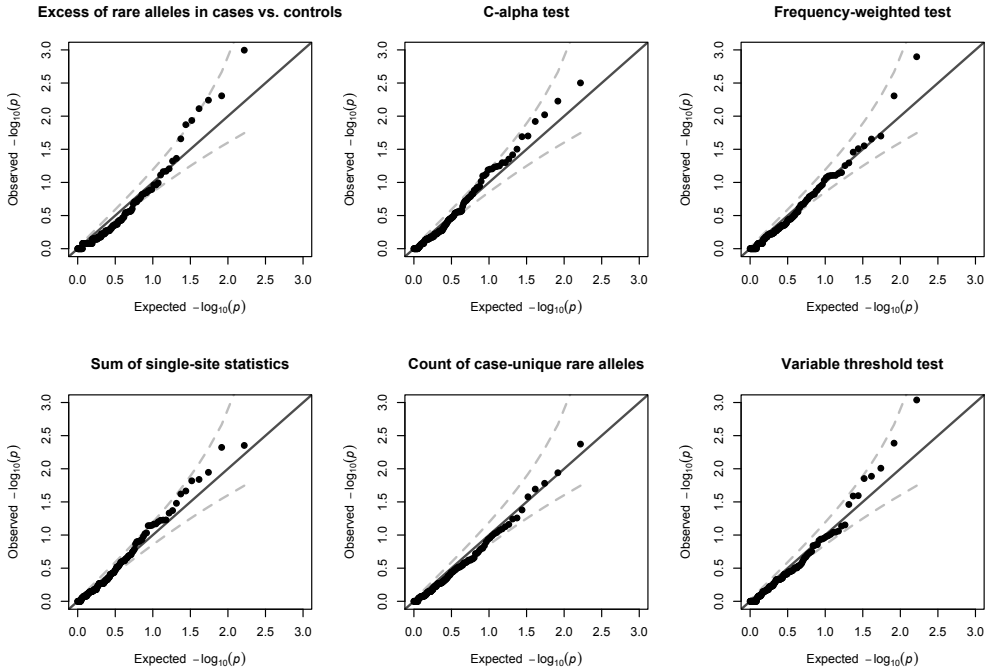

Figure 3. QQ plots of the gene-based p-values obtained using six different association tests in PLINK/SEQ. Grey lines represent 95% confidence intervals.

Discussion

Utilizing existing knowledge on the genetics of monogenic disorders, the present study sought to examine the association of 168 genes implicated in genetics of intellectual disability with normal-range intelligence. Using exon sequencing and focusing primarily on rare genetic variation, we addressed this question in a sample of 191 individuals sampled from the ends of the IQ distribution (>112, <88 and >80). Several different methods of gene-based testing, implemented in the PLINK/SEQ tool, indicated the absence of a detectable association at the present sample size. Additionally, we employed polygenic prediction to examine the overall effect of common genetic variation in the candidate gene set, and found no significant prediction.

The first set of analyses focused on the detection of the possible effects of rare genetic variation, in line with the assumption of the inter-individual variability in intelligence being maintained by low-frequency, disruptive mutations of small effect size. The consistent positive associations between intelligence and fitness components across the life span (Arden, Gottfredson, Miller, & Pierce, 2009; Banks, Batchelor, & McDaniel, 2010; Batty, Deary, & Gottfredson, 2007; Deary, Strand, Smith, & Fernandes, 2007; Deary, Whalley, Batty, & Starr, 2006; Gale, Batty, Tynelius, Deary, & Rasmussen, 2010; Silventoinen, Posthuma, Van Beijsterveldt, Bartels, & Boomsma, 2006; Strenze, 2007; S. Van Dongen & Gangestad, 2011) in combination with the absence of the consequently expected depletion of the underlying genetic variation, indicates that the existing genetic variation is likely to be retained through mutation-selection balance, i.e., a balance between the rate of occurrence of new, mostly deleterious mutations and the rate of their removal by selection (Falconer & Mackay, 1996;

Marioni, Penke, et al., 2014). Because selection will quickly eliminate mutations of strong deleterious effect on fitness-related traits, this mechanism suggests a genetic architecture that lacks common genetic variants of large effect size (Gibson, 2012; Marioni, Penke, et al., 2014); an expectation consistent with the lack of replicable findings from candidate gene and GWA studies (e.g., Benyamin, Pourcain, et al., 2013a; Chabris et al., 2012; Davies et al., 2011). Utilizing the same rationale, Marioni et al. (Marioni, Penke, et al., 2014) recently examined the relationship between the genome-wide count of rare exonic variants and cognitive ability in childhood and old age (N=1596), and detected no significant association. Yeo et al. (Yeo, Gangestad, Liu, Calhoun, & Hutchison, 2011) found a negative association between the genome-wide burden of rare copy number variants and psychometric intelligence in a sample of 74 individuals. However, subsequent studies using larger samples were unable to replicate this finding (Bagshaw et al., 2013; MacLeod et al., 2012; McRae, Wright, Hansell, Montgomery, & Martin, 2013). The present study focused on a smaller part of the genome, in line with the hypothesis that the genetic variation affecting continuous variation in quantitative traits may be concentrated in the same areas of the genome as that underlying similar monogenic phenotypes. Although the lack of detectable association is consistent with the aforementioned studies, a larger study may still be advisable to minimize the probability of the finding reflecting a power issue. Considering the diverse nature of intellectual abilities, as well as the pervasive disagreement between intelligence researchers on the existence and causal relevance of general intelligence (e.g., Gottfredson, 1997a; Neisser et al., 1996), future studies may also employ a finer-grained definition of the phenotype (e.g., verbal and nonverbal intelligence, specific subscale scores, or additive genetic factor(s) derived through the application of genetic covariance structure modeling to twin data; e.g., Franić et al., 2012; M. C. Neale & Cardon, 1992). A substantial heterogeneity in the genetic etiology of intelligence has been demonstrated by previous studies (e.g., Johnson et al., 2007; Luo, Petrill, & Thompson, 1994; Rijsdijk et al., 2002), which typically show significant additive genetic influences specific to cognitive abilities (e.g., verbal, special, perceptual, arithmetic, etc.), in addition to a genetic *g* factor. Provided that such subscale-specific influences are a significant contributor to the genetic etiology of intelligence, future studies may consider their explicit modeling (using, for instance, a multivariate approach). More fundamentally, the question of the ontological and biological reality of the *g* factor has been debated widely (Jensen, 1998; van der Maas et al., 2006; van der Maas, Kan, & Borsboom, 2014); if *g* is a causal entity generating the observed covariation between distinct cognitive abilities, as assumed throughout much of the literature, the approach of attempting to identify genes for *g* is sensible both from a substantive perspective and the perspective of statistical power. However, if *g* is simply an index variable summarizing the covariation between different cognitive abilities without playing a causal role, then seeking genetic influences at the level of *g* will diminish the statistical power to detect the effects of measured genetic variants, relative to seeking genetic influences at the level of its constituent abilities. Genetically informed item-level analyses that assess the mediatory role of intelligence with respect to genetic and environmental effects (Franić, Dolan, Borsboom, Hudziak, et al., 2013) can be used to address some of the above issues.

The rationale behind the present study, namely the supposition of the relevance of genes involved in intellectual disability to normal-range intelligence, is based on ample similar examples in the literature, including height (Allen et al., 2010), body mass index (Loos et al., 2008), lipid levels (Hirschhorn & Gajdos, 2011 review), hemoglobin F levels (Hirschhorn & Gajdos, 2011), type 2 diabetes (Sandhu et al., 2007), Parkinson's disease (Gasser, 2009; Lesage & Brice, 2009), and others. A recent study by Blair et al. (Blair et al.,

2013) identified nearly 3,000 comorbidities between Mendelian disorders and complex diseases present in the electronic medical records in the United States and Denmark. Importantly, each complex disease displayed an association with a unique set of Mendelian disorders, implying a sharing of the causal pathways between the Mendelian and the polygenic phenotypes. Presumably, a monogenic disorder stems from a severely damaging mutation in a gene that normally affects healthy variation in the same phenotype. Consistently with the above findings, we recently demonstrated an enrichment of 43 genes underlying Mendelian disorders of intellectual functioning (39 of which were included in the present study) for common polymorphisms associated with intelligence (Franić, Groen-Blokhuis, et al., 2013). The present study aimed to extend this work to a larger set of genes (i.e., all of the presently identified intellectual disability genes that could be accommodated by our methodology) and examine the effect of rare variants, in addition to common genetic variation. The absence of a detectable association at the present sample size may be considered a (partial) non-replication, although, as mentioned, minimizing the probability of a power issue by employing a larger sample size may be advisable. Other improvements to the present study may include the examination of structural variation (including, for instance, copy-number variants; Redon et al., 2006), gene-by-gene interactions, heterogeneity of genetic effects across different environments, or the intronic regions of the genome. The increasing availability of next-generation sequencing technologies and the rising number of collaborative projects are expected to facilitate a more detailed study into some of the above issues.

Supplementary Methods

Exon sequence data. Quality control and filtering of the genotype data were performed using the VCFtools software (Danecek et al., 2011) and R (R Core Team, 2013). Sites were included into the study based on read quality (Phred-scaled quality > 20), the quality of the bases surrounding the variant (median minimum base quality > 0), strand bias (binomial p-value for strand bias test > .1), mapping quality (root mean square of mapping qualities of reads at the variant position > 20), average variant quality (variant quality / read depth for the variant > 5), probability that the variant segregates in the data (Phred-scaled posterior probability > 20), and proportion of missing data per site. Sites were retained if the sample missingness per site did not exceed 5% (i.e. if at least 95% of samples had at least one read at the site), and sites with sample missingness between 5% and 20% were retained only if their mean depth per sample exceeded 20x. The application of these criteria reduced the number of variants to 2900 (including 972 singletons and 61 doubletons) and the number of genes to 168 (from an initial pool of 175). The frequency distribution of the 2900 variants, and their distribution over the 168 genes, are given in Figures 1 and 2. The mean depth was ~212x, with an average of ~2% sites missing per individual.

SNP data. The SNP data for the polygenic risk prediction were obtained from a larger NTR dataset (N=14,003; see Lin et al., 2014; Nivard et al., 2013). Buccal or blood samples for DNA extraction were collected as part of multiple projects within NTR. DNA extraction and purification were performed at various points in time, following several manufacturer-specific protocols. Genotyping of several partly overlapping subsets was performed on multiple platforms. Chronologically, the following platforms were used: Affymetrix Perlegen 5.0, Illumina 370, Illumina 660, Illumina Omni Express 1M, and Affymetrix 6.0. Genotype calls were made using platform-specific software (APT Genotyper, Beadstudio (Illumina)).

Quality control was performed within and between platforms and subsets. For each platform, the individual SNPs were lifted over to build 37 (HG19) of the Human reference genome using the LiftOver tool (Kuhn et al., 2007). SNPs that did not map at all, had ambiguous locations, or did not have matching (or strand-opposite) alleles were removed. Subsequently, the data were strand-aligned with the 1000 Genomes phase 1 release v3 panel. SNPs were removed based on mismatches of alleles with those in the reference set, differences of allele frequencies (larger than .2) from those in the reference set, minor allele frequency (<1%), absence of Hardy-Weinberg equilibrium (p<.00001), and call rate (<95%). Samples were excluded from the dataset based on mismatch of expected sex and the sex derived from the genotype data, genotype missing rate (>10%), and the coefficient of inbreeding (F>0.10 or < -0.10).

Following these steps, the data from the individual arrays were merged into a single dataset using PLINK 1.07 (Purcell et al., 2007). Within the merged set, identity by state (IBS) sharing was calculated between all possible pairs of individuals; subsequently, IBS was compared to the known family structure within the NTR. Samples that did not display the expected IBS sharing were removed. DNA samples that were typed on multiple platforms and displayed discordance between the overlapping SNPs (concordance rate <99%) were removed. Subsequently, a single DNA sample was selected from each MZ twin pair, resulting in a total of 12,240 unique DNA samples to be imputed. Hardy-Weinberg equilibrium, minor allele frequency, and allele frequency (>.2) filters were re-applied to the merged data. To avoid erroneous strand alignment, SNPs with the allele combinations C/G and A/T and a minor allele frequency between .35 and .5 were removed.

Phasing of the samples and the imputation of the SNPs missing across some platforms were performed using MACH 1.0 (Y. Li & Abecasis, 2006). The phased data were imputed using Minimac (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012), in batches of ~500 individuals for 561 chromosome chunks obtained using the CHUNKCHROMOSOME software (E. Y. Liu, Li, Wang, & Li, 2013). After imputation, the data on DNA-confirmed MZ twins were duplicated back into the dataset, resulting in a dataset containing a total of 14,003 individuals. The mean imputation R^2 was 0.38. The imputed dataset contained 30,051,533 autosomal SNPs. Post-imputation, SNPs were filtered based on Mendelian error rate in families (>2%), Hardy-Weinberg equilibrium (<0.00001), imputation quality R^2 (<.3), minor allele frequency (<.005), and a discrepancy in allele frequencies of the imputed SNPs and the 1000 Genomes reference panel (>.15). This resulted in a final dataset containing 7,981,681 autosomal SNPs with a mean R^2 of .86.

Chapter 10

Summary and Discussion

The present dissertation has focused broadly on the ontology of latent psychometric variables, and the genetics of intelligence. Below I provide a summary of the preceding eight chapters, followed by a general discussion.

Chapter 2 introduces the basics of structural equation modeling, as applied in the classical twin design. After introducing the basic method of exploiting familial relationships to infer the effects of unmeasured genetic and environmental factors, the chapter reviews the implementation of models from the structural equation modeling literature into genetically informative designs, and structural equation models developed specifically within genetics. The former include simplex and latent growth curve models, and the latter include common and independent genetic factor models, genotype-environment interaction models, sex-limitation models, and direction of causation models. The chapter concludes with a discussion of the incorporation of measured genetic variables into structural equation modeling-based association analysis.

Chapter 3 discusses the application of genetically informed item-level analyses in addressing questions regarding the ontology of latent behavioral phenotypes (e.g., depression, general cognitive ability), via the study of their mediatory role with respect to genetic and environmental influences. The presence of genetically informative item-level data allows one to 1) test an empirical implication of the realist interpretation of latent psychological traits, namely its mediation of genetic and environmental influences on the observed item covariation, and 2) study the (possibly different) dimensionalities of the latent genetic and environmental covariance structures giving rise to the observed item covariation. I note that the frequently encountered problems in psychometric dimensionality assessment may be viewed as a function of the differences between these genetic and environmental covariance structures, and propose using genetically informative item-level analyses as a tool in improving phenotypic dimensionality assessment.

Chapter 4 employs the methodology discussed in Chapter 3 to examine the ontology and the genetic and environmental etiology of the Internalizing syndrome dimensions of the Child Behavior Checklist (CBCL; Achenbach, 1991; Verhulst et al., 1996). The results 1) suggest that the syndrome dimensions may be better understood as a composite of unconstrained genetic and environmental influences than as causally relevant entities generating the observed symptom covariation, and 2) indicate a common genetic basis for anxiety, depression, and withdrawn behavior, with the distinction between these syndromes being driven by the individual-specific environment. The finding is discussed in the context of the frequently encountered difficulties in phenotypic delineation between different diagnostic categories, e.g., anxiety and depression.

Chapter 5 employs the same methodology to examine 1) the tenability of the realist interpretation of the Big Five personality dimensions (McCrae & Costa, 2008), and 2) the structure of the genetic and environmental covariance matrices underlying the observed covariation of NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) personality items. Interestingly, and unlike the case of the CBCL, the genetic and the environmental covariance matrices underlying NEO-FFI item covariation exhibit similar (5-factor) structures. However, the latent personality dimensions do not appear to fully mediate the genetic and environmental effects on the items, as would be expected under the realist

interpretation of the Big Five. Implications for the substantive interpretation of the Big Five are discussed.

Chapter 6 provides an overview of the genetic covariance structure modeling-based methodology for the study of childhood anxiety and depression, and a cross-section of the relevant findings. The review focuses on questions that go beyond the relatively simple task of assessing the contributions of genetic and environmental factors to anxiety and depression. The review presents relatively consistent evidence for: a) small to negligible sex differences in the genetic etiology of these disorders, b) a substantial role of genetic factors in accounting for their temporal stability, c) a contribution of genetic factors to the comorbidity between them, d) a possible role of genotype-environment interaction in affecting their liability, e) a role of genotype-environment correlation, and f) a minor, if any, etiological role of sibling interaction.

Chapters 7-9 focus on the genetics of intelligence. Chapter 7 reports on a combined analysis of all longitudinal measures of verbal, nonverbal, and general intelligence present in the Young Netherlands Twin Register (Bartels, Beijsterveldt, et al., 2007) in 2009. Simplex modeling was used to examine the genetic and environmental etiology of the temporal stability of the measures. Given the information on stability, I subsequently address the question of how to optimally utilize the existing longitudinal data in the context of gene-finding studies. The high stability of the additive genetic factors indicates that a single set of genes underlies the variation in intelligence throughout the developmental period under study, justifying the use of a linear combination of scores across the different ages in the context of genetic association studies.

The results obtained in Chapter 7 were used to inform the modeling of the phenotype in the association studies reported in Chapters 8 and 9. Chapter 8 reports on a study testing for an association between normal-range intelligence and common single-nucleotide polymorphisms (SNPs) in 43 known cognitive disability genes. The study utilized a simple rationale, namely the fact that the genetic variation affecting continuous, polygenic traits (e.g., normal-range intelligence) may be concentrated in the same areas of the genome as that underlying similar, monogenic phenotypes (e.g., intellectual disability). Although no individual single-nucleotide polymorphism (SNP) reached statistical significance, SNP-based analyses indicated an enrichment of the candidate gene set for polymorphisms associated with intelligence. The study is the first demonstration of the relevance of genes implicated in monogenic disorders of intellectual functioning to normal-range intelligence.

Chapter 9 extends the work reported in Chapter 8 to 168 known intellectual disability genes, but, unlike Chapter 8, uses next-generation exon sequencing and focuses on the detection of the possible effects of rare genetic variation. Consistently with the literature to date, no enrichment of the candidate gene set for mutations associated with normal-range intelligence was detected at the present sample size. The finding is discussed in the context of literature.

Discussion

The present dissertation has focused on a) the use of genetically informed item-level analyses in psychometric dimensionality assessment and the study of the ontology and the genetic and environmental etiology of latent traits, with application to childhood internalizing problems and the Big Five personality dimensions in adults, and b) the genetics of intelligence. A variety of techniques were used to address these topics and various issues therein, discussed in turn below.

The role of genetics in psychometric dimensionality assessment

The past several decades have seen major developments in the methodology for the assessment of psychometric dimensionality, i.e., the determination of the number of latent attributes underlying a set of indicators (e.g., item responses, symptoms). The standard toolkit for dimensionality assessment, including exploratory factor analysis and related models (e.g., principal components analysis), has been expanded to include confirmatory methods, e.g., item response theory modeling and confirmatory factor analysis. A good deal of work has gone into the development of heuristics to facilitate this process, resulting in an impressive statistical toolbox of methods (including, e.g., the scree plot, the "eigenvalue-greater-than-one" rule, the minimum average partial correlation, the Chi-square test, and parallel analysis). A variety of fit indices developed in structural equation modeling (e.g., RMSEA, ECVI, incremental fit indices, and information criteria) found widespread application in both the exploratory and confirmatory approach, and IRT-based methods have given rise to specialized software for dimensionality assessment (e.g., DIMTEST; Stout, 1987).

Notwithstanding the availability of these tools, the assessment of dimensionality has remained difficult. One only needs to look at fields of intelligence, psychopathology and personality assessment, where substantial controversy still exists regarding the origin of covariation between different symptoms/behaviors/questionnaire items. For instance, there is presently a lack of consensus on whether the general intelligence (*g*) factor can be equated with some of the more specific intellectual abilities, such as working memory or fluid reasoning (Ackerman et al., 2005). In internalizing psychopathology research, the covariation of symptoms of anxiety and depression has given rise to a host of theories, ranging from those that view the two disorders as separate entities with overlapping features, to those that view them as different points along a single continuum (Clark, 1989).

The present dissertation has inquired why dimensionality assessment is so difficult, and proposed that one of the reasons lies in the fact that the genetic and environmental influences, of which the observed covariation is a function, differ from each other in structure and dimensionality. Employing item-level analyses on genetically informative data enables the explicit study of the dimensionality of these genetic and environmental influences, thereby moving the question of dimensionality from the observed to the genetic and environmental level. As demonstrated, the increased resolution afforded by this approach may further the understanding of the nature of problems arising in dimensionality assessment, elucidate the origin of the phenotypic dimensionality of observed symptoms/behaviors/item responses, and help improve the definition of phenotype in genetic association studies. On a conceptual level, the approach can inform the discussion on the ontology of the latent variables obtained in psychological research. The present dissertation has laid out the tools that may be used to this end, and examined the feasibility of the analyses proposed. Can genetics thus help psychometrics? The present dissertation has argued that the answer is yes. Applications to childhood internalizing problems and personality dimensions illustrate this point in practice.

Despite decades of research into the origin of covariation between psychometric items (including both internalizing symptoms and personality indicators), the development of noncontroversial taxonomies has proven challenging. In internalizing psychopathology research this has given rise to a host of questions, ranging from those that inquire, e.g., whether anxiety and depression are different manifestations of a single entity (Clark, 1989), to those inquiring whether they are entities at all. In personality research similar questions

arise: can the fundamental structure of personality be uncovered by the application of factor analysis to personality items, what is the number and the interpretation of the relevant latent factors, and should the structure of personality be conceptualized as entailing latent factors at all (J. Block, 1995)? For both of the empirical datasets analyzed in the present dissertation, the answer with respect to the ontology of latent variables has proven negative: neither personality factors nor internalizing dimensions appeared to fully mediate genetic and environmental effects on the items, implying that one cannot interpret them as behavior-generating entities in the realist sense, at least not as they are currently defined. In the case of personality dimensions the answer is somewhat more complicated as the structures of genetic and environmental covariance matrices, seemingly paradoxically with respect to our conclusion regarding their ontology, both display highly similar, five-factor structures. Is this a result of the careful pre-selection of items during the decades of psychometric construction and refinement of the item set, or a finding reflecting a fact of nature, namely a five-factor structure of personality? Could one think of a defendable way to accept diluted versions of realist latent constructs that only partially mediate genetic and environmental influences? What would the theory of such partial mediation be? Are there other reasonable hypotheses that one could construct about the finding? These and similar questions may motivate future research, examining for instance whether the misfit of the common pathway model was due to local as opposed to global violations, or whether the same analyses on a different set of personality items would produce similar results.

How do the results of this type of analyses relate to genetic association studies? If, for instance, the Anxious/Depressed dimension of the CBCL is not a unitary construct, should attempts be made to identify genes that predispose individuals for a high standing on this dimension? Is this comparable to deriving a factor score from items measuring, say, shoe size, cholesterol levels, and income (which may well display a positive manifold of correlations unless one controls for age), and attempting a search for genes that predispose individuals for a high standing on this trait? Yes, and no. The situations are comparable in that neither the Anxious/Depressed variable, nor (presumably) the shoe size/cholesterol/income variable, would mediate all the genetic and environmental effects on their indicators, as neither appear to be entities in the realist sense. Importantly, however, they are not comparable in that, unlike the shoe size/cholesterol/income indicators, the Anxious/Depressed indicators appear to be genetically unidimensional, i.e., affected by a single set of genes. The pertinent question for gene-finding purposes is that of genetic unidimensionality: a genetically unitary construct (such as the Anxious/Depressed dimension) need not necessarily be problematic in the context of gene finding, regardless of its phenotypic complexity. A related question is that of genetic and environmental unidimensionality over time. In the presence of longitudinal data, one may inquire how to construct a phenotype that optimally indexes genetic effects. For instance, intelligence measures collected in late adolescence display a larger heritability than those collected in childhood, but the use of those collected at earlier ages may imply a larger sample size. Using data from a single age may be inefficient in terms of discarding other data (the addition of which could increase the measure's reliability), while using all measures simultaneously may dilute the genetic signal if different sets of genes affect the measure across development. If one opts to use all available data, should one employ a multivariate model, or can a summary statistic (e.g., a mean across ages) adequately represent the phenotype? Finally, how do the above choices affect the statistical power to detect genetic effects? The above issues were addressed in Chapter 7 with respect to intelligence. The results, and the subsequent gene-finding efforts that used them, are discussed below.

Intelligence: temporal stability and the search for genes

The 'missing heritability' problem (i.e., the discrepancy between heritability estimates yielded by twin and family studies and the proportion of variance explained by significantly associated variants; Maher, 2008) appears pervasive in the genetic study of complex traits, with examples ranging from anthropometric traits (e.g., height, body mass index), metabolic traits (e.g., fasting glucose and insulin levels), and common diseases (e.g., cardiovascular, metabolic, neurological, or immune system disease), to behavioral traits (e.g., neuroticism, extraversion) and psychiatric disorders (e.g., depression, schizophrenia, autism, personality disorders). As evident in the present dissertation, the situation is not dissimilar with respect to intelligence: despite major efforts by large consortia, no significantly associated single-nucleotide polymorphisms (SNPs) have been identified, and only one gene (FNBP1L) has been tentatively implicated in the etiology of normal-range intelligence to date (Benyamin, Pourcain, et al., 2013b; Davies et al., 2011). A plethora of explanations have been put forward to account for the missing heritability phenomenon; these include the presently insufficient statistical power of genome-wide association (GWA) studies to detect genetic variants of small effect size, the potential overestimation of heritability by twin studies, problems pertaining to the measurement and operationalization of the phenotype, and the possibility of genetic variants not tagged on the present genotyping platforms (including rare and structural variation) underlying the heritability (e.g., Dickson et al., 2010; Eichler et al., 2010; Goldstein et al., 2013a; Teri A Manolio et al., 2009; van der Sluis et al., 2010; Zuk et al., 2012).

With respect to statistical power, the consensus view is clear: larger samples are preferable, and with respect to intelligence it appears that large enough sample sizes in GWA studies are yet to be reached (the largest to date GWA (meta-)analyses comprised N≈18,000 and N≈3,500 in children and adults, respectively; Benyamin, Pourcain, et al., 2013b; Davies et al., 2011). The potential overestimation of heritability by twin studies remains a looming issue in the study of many phenotypes, seeing as the estimation of epistatic interactions (i.e., interactions of alleles across different genetic loci), on which the degree of potential overestimation of heritability in the classical twin design will depend (Keller & Coventry, 2005), is a difficult issue to tackle empirically. Indeed, while fixing certain parameters (including the non-additive variance component) to zero is expedient to circumvent parameter indeterminacy inherent to the classical twin design, there is no a priori reason not to expect additive and non-additive genetic, and common and unique environmental factors to all jointly affect the phenotype. With respect to intelligence, previous analyses have indicated the empirical data to be consistent with non-additivity (Devlin, Daniels, & Roeder, 1997; Lindon J Eaves, 1973), although non-additivity alone is unlikely to explain the entire missing heritability gap, seeing as a) its estimated magnitude is small, and b) an estimated ~22-46% and ~29-51% of the variance in intelligence in children and adults, respectively, have been shown to be explained by the additive effects of common genetic variants measured on the present SNP microarrays (Benyamin, Pourcain, et al., 2013b; Davies et al., 2011). A number of other phenomena that may inflate heritability estimates, including the possible interaction between the additive genetic and common environmental factors, sibling competition effects, and systematic differences in the treatment of MZ and DZ twins, have been examined with respect to many phenotypes and appear to not pose problems for the interpretation of variance components obtained in the twin design (e.g., Borkenau, Riemann, Angleitner, & Spinath, 2002; Kendler, Neale, Kessler, Heath, & Eaves, 1993; Molenaar, van der Sluis, Boomsma, & Dolan, 2012).

A major issue in the analysis of intelligence data concerns the definition and the modeling of the phenotype. What exactly are we looking for genes for? Chapter 7 dealt with this question in view of optimally utilizing longitudinal data, i.e., establishing whether data summarization over ages is likely to diminish the power to detect genetic effects. More generally, one can think about the resolution of the phenotype – would modeling individual items or subscales be more beneficial than modeling general intelligence? In this light, the study of the genetic dimensionality of intelligence items and the mediatory role of general intelligence, or that of more specific abilities (e.g., verbal, nonverbal), a proposed in Chapters 3-5, would be highly informative. The present capacities of computing resources and the development of the relevant software (e.g., openMx; Boker et al., 2010) would likely make this a feasible task, despite the typically large number of items included in intelligence tests.

Finally, part of the variation in intelligence may potentially be explained by the effects of variants not tagged on the present genotyping platforms, including rare and structural variation. For instance, it has recently been demonstrated that individual dinucleotide short tandem repeats (STRs) may explain over six times more phenotypic variance than individual diallelic SNPs (Willems, Gymrek, Highnam, Mittelman, & Erlich, 2014). This potential of STRs to contribute to phenotypic variance, in combination with their poor tagging on SNP arrays, suggests that they may be a significant contributor to the missing heritability phenomenon. Another issue concerns the possible role of rare variants. While there is no a priori reason to exclude variants from any part of the allelic frequency spectrum as potentially relevant to intelligence, most of the research to date has dealt with the estimation of the possible effects of common genetic variation. Arguments can be made in favor of both rare and common variants, however. As proposed by, e.g., Hsu (Hsu, 2012; Marioni, Penke, et al., 2014), part of the genetic variability in intelligence can be maintained by rare deleterious mutations of small effect size, whose modest effects make their elimination by selection unlikely. Common variants, on the other hand, may be present in the form of effectively neutral mutations (subject to genetic drift), or as relatively positive mutations (subject to positive selection), which have yet to become fixed in the population.

The role of rare genetic variation in the etiology of normal-range intelligence is still a largely unexplored issue, although the declining costs of exome- and whole-genome sequencing will enable more extensive investigations into this issue in the near future. The present dissertation has already taken a step in this direction, albeit with a limited sample size. The study design, utilizing knowledge on Mendelian disorders to study a related polygenic phenotype, may be a useful tool in the identification of genomic areas harboring causal variants, as has been exemplified both by the enrichment reported in Chapter 8, and by a recent study that revealed nearly 3,000 comorbidities between Mendelian disorders and complex diseases present in the electronic medical records in the United States and Denmark (Blair et al., 2013). Importantly, the study reported each complex disease to be associated with a unique set of Mendelian disorders, implying shared causal pathways between the Mendelian and the polygenic phenotypes. In combination with the widely observed enrichment of associations for complex traits in genes known to underlie related monogenic conditions (e.g. body mass index – monogenic obesity, height – skeletal growth disorders; Allen et al., 2010; Loos et al., 2008), this finding suggests that Mendelian disorders will provide a guiding light in mapping the normal variability underlying complex traits. Ultimately, theoretically informed approaches in tandem with a better understanding of the phenotype and the consequent improvements in its modeling, along with the increasing accessibility of larger amounts of whole-genome sequence data, will help make significant strides in this direction.

The broader context: the twin design in the 21st century

As evident from the diversity in the methodology employed in the present dissertation - ranging from genetic covariance structure modeling to next-generation sequencing - the field of behavior genetics has undergone radical development over the past half a decade. One of the questions arising in this context concerns the relevance of the methodologies outlined at the beginning of this dissertation in the present era of genomics: are twin designs still relevant, or do the new technologies render them obsolete? Does the era of next-generation sequencing leave any questions that may be uniquely addressed by the study of twins? Related to that, what is the practical applicability of the results obtained by the twin method? The following sections will address these issues in turn.

The utility of twins in the genomics era

Some of the traditional uses of the twin design and its various extensions have been outlined in the introduction of this dissertation. Beyond the estimation of heritability, twin designs have enabled the study of a range of issues including the genetic and environmental etiology of developmental stability and change in behavioral phenotypes, the dependency of polygenic effects on measured environmental exposures, the etiology of inter-individual variation in age-related growth and decline, the direction of phenotypic causality between traits, rater bias, sibling imitation and contrast effects, and the ontology of latent psychological traits. Now that many of the aforementioned issues have been settled with respect to many phenotypes and genetics has well entered the age of the widespread availability of measured genetic information, one may pose the obvious question of whether there is further utility in the study of twins. The present section will attempt to address this question by reviewing some of the main areas of application of twin designs that go beyond the standard applications outlined in this dissertation, and are poised to address novel issues arising in the context of the recent technological advances in the biomedical sciences (J. P. van Dongen et al., 2012). The issues include the timing of *de novo* mutagenesis, the role of epigenetic changes and gene expression in disease pathogenesis, disease-associated changes in metabolite levels, and the identification of microbial signatures associated with disease. In addition, I discuss how the recent advances in sequencing technologies can be employed to verify fundamental assumptions of the twin design concerning the degree of genetic and environmental sharing between monozygotic (MZ) and dizygotic (DZ) twins.

Although a number of twin designs may be employed to address the issues listed above, the continuing utility of twins is perhaps most discernibly exemplified by the discordant MZ twin design. By comparing the biological feature of interest (e.g., the genome or the metabolome) in MZ twins discordant for a given phenotype, the application of this design can provide insight into disease pathogenesis and aid in the detection of biomarker profiles for medical conditions (J. P. van Dongen et al., 2012). For instance, a comparison of gene expression in subcutaneous fat of MZ twins discordant for obesity has demonstrated differential expression in a range of genes, including those involved in inflammatory pathways (upregulated in obese twins) and in mitochondrial branched-chain amino acid catabolism (downregulated in obese twins) (Pietiläinen et al., 2008). Similar designs employing metabolomics data have detected differences in serum and fat tissue lipid profiles of discordant MZ twin pair members; this work prompted subsequent simulation of lipid bilayer dynamics using lipidomics and gene expression data, which provided novel functional insights into the biological pathways underlying adipocyte expansion (Pietiläinen et al., 2011; Pietiläinen et al., 2007). Twin studies of obesity have also been carried out using

the microbiome; for instance, a comparison of faecal microbial communities in obese and lean MZ twins have indicated that obesity is associated with a reduced bacterial diversity and differential representation of specific bacterial genes and metabolic pathways (Turnbaugh et al., 2009). Interestingly, a study of ulcerative colitis (a form of inflammatory bowel disease) indicated that the condition may be associated with a loss of interaction between the mucosal transcriptional profile and the colonic microbiota, based on a comparison of discordant twins MZ that showed that fewer RNA transcripts correlate with bacterial genera in the affected than in the unaffected twins (Lepage et al., 2011). Another area of application concerns the study of the role of epigenetic variation (i.e., changes in gene activity that are not caused by changes in the DNA sequence) in disease pathogenesis. For instance, differential regulation of miRNA transcripts in lymphoblastoid cell lines in twins discordant for autism aided in the subsequent identification of ID3 and PLK2 genes (the target genes for two of the differentially expressed miRNAs) as candidate genes for autism (Sarachana, Zhou, Chen, Manji, & Hu, 2010). The analysis of DNA methylation patterns of MZ twins discordant for systemic lupus erythematosus (a chronic autoimmune inflammatory disease) identified several genomic regions in which DNA methylation was associated with the disease (Javierre et al., 2010).

Aside from the study of disease etiology, the discordant MZ twin design can be used to study the timing of the occurrence of *de novo* mutations (i.e., mutations that arise in the offspring without being present in either parent; Veltman & Brunner, 2012). For instance, a *de novo* mutation present in a single MZ twin pair member only would indicate post-twinning mutagenesis; a *de novo* mutation present in both MZ twin pair members indicates a pre-twinning mutation event. The presence of a mutation in the sodium channel α1 subunit gene (SCN1A) in multiple embryonic tissue lines in a twin affected by Dravet's syndrome coupled with its absence in the unaffected twin, for instance, indicated that the mutation had likely occurred at the two-cell stage in the pre-morula embryo (Vadlamudi et al., 2010). Information on the timing of mutagenesis is of crucial importance in genetic counseling, as a mutation that occurred in the parental gamete is associated with a negligible risk of recurrence in additional offspring.

Uni- and multivariate implementations of the classical twin design, as presented in Chapter 2, remain of utility too. These are increasingly employed to study a host of newly emerging phenotypes, including the epigenome, the transcriptome, the metabolome, the proteome, and the microbiome (J. P. van Dongen et al., 2012). The application of the classical twin design to gene expression data, for instance, has demonstrated the importance of both genetic and environmental factors in genome-wide expression levels, with the relevance of genetic and environmental influences varying over different genes and tissues (Mcrae et al., 2007; York et al., 2005). Multivariate analyses can be employed to quantify the extent to which genetic and environmental factors that are shared across different genomic regions affect epigenetic regulation and gene expression, or biological variation across different cells and tissues (J. P. van Dongen et al., 2012). In addition, the quantification of the effect of genetic factors on epigenetic changes may be accomplished using the classical twin design. Thus far, such applications have demonstrated a low overall heritability of epigenetic changes across all loci, although substantial genetic influences on some loci (e.g., the imprinted *IGF2–H19* locus) have been detected (Heijmans, Kremer, Tobi, Boomsma, & Slagboom, 2007).

A number of other twin designs can be employed in the study of newly emerging biomedical phenotypes. For instance, the offspring-of-twins design can be used to study transgenerational inheritance of epigenetic regulation and the role of maternal effects on epigenetic marks (J. P. van Dongen et al., 2012), and longitudinal twin studies can be

employed to identify biomarkers associated with ageing (e.g., telomere length in relation to longevity; e.g., Bakaysa et al., 2007). Longitudinal twin designs may also be employed to resolve the direction of causation with respect to epigenetic changes, i.e., to distinguish between a situation in which an epigenetic change brought about a phenotypic condition from one in which an underlying cause brought about both the epigenetic change and the phenotypic condition.

As evident from the above examples, the twin design remains a useful tool in present-day biomedical and psychiatric research. In return, modern technologies have aided the twin design, by enabling the explicit verification of some of its fundamental assumptions. For instance, next-generation sequencing has demonstrated that genetic sharing between MZ twins is, as expected, nearly 100% (although minor differences are sometimes detected), and genome-wide microsatellite data have indicated that the proportion of genetic sharing between DZ twins mostly lies between 42% and 58%, with an average close to 50% (Visscher et al., 2007). Aside from the evident value of disease-discordant MZ twins in the study of a range of newly emerging phenotypes, the existence of discordant MZ twins also has implications for the prospect of genomic risk prediction and the ethical concerns that have been raised in this context. Namely, the existence of phenotype-discordant MZ twins indicates that the genome does not necessarily fully predict the phenotypic outcome of individuals. Thus, barring the case of fully penetrant traits, precise individual risk prediction based only on the DNA sequence is likely to remain unfeasible, even if all the genetic variation contributing to disease risk is identified. Moreover, this is true regardless of the trait's heritability – for instance, despite 80% of individual differences in liability to schizophrenia being explained by genetic factors, MZ twin concordance for this disease is only ~40-50%. Similarly to various other phenomena related to the concept of heritability, this may sound somewhat counterintuitive. To place this in a broader context and outline what the definition of heritability does – and does not – entail, in the next section I review some of the other (mis)conceptions related to this concept, and their implications for prevention and treatment.

Heritability – (mis)interpretations and implications for prevention and intervention

As outlined in the present dissertation, research has highlighted the relevance of both genetic and environmental factors to observed individual differences in behavioral traits. For instance, the heritability of internalizing behaviors in 10-12 year old children is ~30% (Chapters 3 and 4), the heritabilities of NEO-FFI personality indicators in adults range from ~60% to ~80% (Chapter 5), and the heritability of general cognitive ability increases from ~40% in early childhood to ~70% in adolescence (Chapter 7). However, while the importance of environmental factors is usually interpreted as implying modifiability, heritability is frequently taken to imply immutability. In the present section I consider this issue (i.e., the implications of heritability for the prospects of modifying the level of a phenotype - e.g., is an 80% heritable trait easier to modify than a 20% heritable trait?), and, related to this, address several common misconceptions about the concept of heritability.

Being a proportion of variance, heritability quantifies inter-individual differences, i.e., the proportion of individual differences in a phenotype explained by the variation in genetic polymorphisms relevant to the trait. This implies, amongst other things, that 1) heritability only gauges the relative contribution of the genetic loci that segregate in the population, i.e., it ignores the contributions of genetic loci that are monomorphic (although such loci contribute to many crucial aspects of human development, including those that are

prerequisites for the phenotype of interest to develop), 2) related to this, heritability pertains to variability, not to the absolute level of a trait, and 3) a heritability estimate cannot be interpreted on an individual level (e.g., a heritability estimate of 30% does not imply that 30% of a child's internalizing problems are due to his or her genes, with the remaining 70% being due to environmental factors). Related to this is the perceived immutability of the degree of importance of genetic factors in the etiology of a trait, as represented by a heritability estimate. Is heritability an immutable intrinsic property of a trait? For several reasons, no. Firstly, per definition, heritability depends on the population in which it is estimated, as both genetic and environmental variation are population-specific. The genetic variance depends on the segregation of alleles relevant to the trait, allele frequencies, and their effect sizes and mode of action, all of which may differ across populations. Similarly, the variance of environmental factors relevant to the trait can differ across populations. Think of an environment with a low degree of relevant non-genetic variability in which most of the variation in the phenotype is accounted for by genetic factors, in contrast to an environment with a high degree of non-genetic variability, in which the same phenotype is consequently less heritable. In relation to this, heritability may also be influenced by the level of the environment, as it has been shown that some environments facilitate the genetic expression of a trait, while others may suppress it. For instance, an intellectually stimulating environment might facilitate the (genetically influenced) differentiation between children in terms of their cognitive abilities, relative to a less stimulating environment in which there is nothing to elicit the bright children's potential, thereby fostering a more uniform development. Another feature of heritability that highlights its dynamic nature is its age-dependency: the heritabilities of many traits, including intelligence and internalizing problems, display an age-related increase (Bergen et al., 2007). This phenomenon may be partly due to active gene-environment correlation, i.e., to individuals selecting environments compatible with their genetic propensities, which in turn reinforces the expression of those propensities (e.g., Haworth & Davis, 2014; Plomin et al., 2008).

Importantly, as mentioned, because it pertains only to individual differences, heritability is inherently uninformative on, and independent of, the absolute value of the phenotype. For instance, the steady increase in intelligence test scores over the past decades has not been accompanied by a change in heritability (Flynn, 1987; Kan, Wicherts, Dolan, & van der Maas, 2013; Sundet, Tambs, Magnus, & Berg, 1988). For similar reasons, heritability estimates are not necessarily informative on how modifiable the value of a trait is and, conversely, the success in changing the value of a trait is not necessarily informative on the importance of genes in explaining its variation (Haworth & Davis, 2014). To illustrate some of the above points, think of interpreting a statistic such as the mean number of bicycles per person in The Netherlands (presently .98) as immutable. Similarly as a change in the number of bicycles per person would lead to a change in the mean statistic, changes in trait values due to an environmental intervention may, depending on their pattern of influence, change the statistic describing its heritability. In this sense heritability is a descriptive; a statistic describing the state of affairs as it is, given various contextual factors that directly or indirectly enter the equation (e.g., the conduciveness of the environment to the genetic expression of a trait, the age of the population measured, the presence and magnitude of relevant environmental variation, etc.). The heritability estimate does not provide any information on what might be, were those factors different (in this sense, the use of the word 'estimate' may perhaps be questioned as it implies the assessment of an intrinsic property of a trait, and a term along the lines of 'descriptive' may be more appropriate).

Now that I have laid out several issues with respect to which the heritability estimate is uninformative (e.g., the absolute value of a phenotype, its intra-individual etiology, and

the potential to modify its value), the question of what heritability estimation and, more broadly, the findings of genetic research, *can* inform us on emerges as relevant. As mentioned, the heritability of a phenotype need not have implications for the potential success of environmental interventions. The intuition that the opposite is the case seems to stem from the idea that underlying biology is difficult to change. Indeed, the prospect of genetic engineering for complex traits is presently slim. However, genetic research can aid interventions insofar as it may provide information on mechanisms and causal pathways involved in the genetic etiology of disease (Haworth & Davis, 2014). Subsequent environmental interventions, which may target any level of physiology and behavior – from biological causal pathways to behavioral endpoints of interest – can significantly benefit from such information. While interventions of this type are, strictly speaking, not 'genetic' (as they do not alter the DNA sequence), they can utilize mechanistic knowledge on the phenotype's genetic etiology to modify the connection between the genotype and the phenotype. A classic example is Phenylketonuria (a congenital condition characterized by a defective gene for the enzyme that breaks down phenylalanine, leading to abnormal brain development), whose heritability dropped from 100% to 0% due to an entirely environmental intervention, namely the elimination of phenylalanine from the affected children's diet. In this case the connection between the phenotype and the genotype was effectively broken, resulting in the recessive homozygotes no longer developing the phenotype despite possessing the relevant alleles. Similar examples are found amongst complex traits, the genetic risk for many of which is commonly mitigated via environmental interventions (e.g., diabetes, obesity). A related, and presently underexplored issue is the genetic etiology of individual differences in treatment response and, in particular, the question of whether the same or different genes are involved in baseline phenotype and treatment response. Treated as error term in traditional intervention designs that focus on mean changes, individual differences in treatment response may potentially provide mechanistic insight into the efficacy of interventions (i.e., understanding why treatment works better for some people may help understand why it works at all), and inform future efforts on the potential value of personalizing treatment. Twin and family designs can make a significant contribution to addressing this and related questions (for instance, why individuals often rebound to their pre-intervention state, whether there are critical periods in which intervention is most effective, and how interventions exert their influence (e.g., epigenetic processes); Haworth & Davis, 2014). Ultimately, a better understanding of the genetic and environmental etiology of individual differences may result in better-informed and more efficient intervention and prevention designs.

Conclusion

As evident from the present dissertation, behavior genetic research has undergone radical development over the past half a decade. Numerous and varied features of the genetic etiology of behavioral traits, including internalizing psychopathology, personality, and intelligence, have been studied extensively and successfully using the existing methods. Presumably, the coming five years will entail improvements in the continuing efforts towards the identification of the relevant genetic variation, and the enhancement of the prospects of genetically informed prevention and intervention. The future developments in the relevant methodology in combination with the existing approaches (e.g., the use of twin designs in the study of newly emerging biomedical phenotypes, functional studies, and research on treatment response) should greatly increase the feasibility of this.

Appendices

Appendices 2-5 can be obtained at http://sanjafranic.com/dissertation.

Appendix 1. Introduction to Quantitative Genetic Theory

Genetics: basic terms and concepts

The human genome consists of DNA (deoxyribonucleic acid), a polymeric molecule comprised of a chain of monomeric subunits termed nucleotides. The bulk of the human genome - around 3,200,000,000 DNA nucleotides - is contained in the cell's nucleus. The nucleotides are organized into 24 types of linear molecules, each contained in a different chromosome (22 autosomes, i.e., non-sex chromosomes, and 2 sex chromosomes). Vast majority of cells are diploid, i.e., they contain two copies of each autosome and two sex chromosomes (XX for females and XY for males), 46 chromosomes in total. These are somatic cells, in contrast to haploid sex cells or gametes, which contain a single copy of each autosome and a single sex chromosome.

A gene is traditionally thought of as a sequence of DNA nucleotides coding for an RNA and/or polypeptide molecule (ref);[34] for the current purpose we will adopt the definition of a gene as a region of the genomic sequence corresponding to a unit of inheritance. Genes may correspond to regulatory regions, transcribing regions and/or other functional sequence regions typically associated with the production of proteins; via this association, genes influence observable traits, also denoted phenotypes. Locus is the site of the gene on the chromosome. Alternative forms of a gene occupying a locus are denoted alleles. Genes with only one allele present in the population are termed monomorphic; those with two or more alleles are termed polymorphic (or segregating) genes. Despite the large number of alleles associated with certain genes (e.g., the HLA-B gene has over 400 alleles, ref), the descriptions that follow will be illustrated using the simplest instance, i.e., a diallelic gene (gene with only 2 alleles).

In diploid cells, each chromosomal locus consists of a pair of alleles, each allele contained in one of the chromosomes of a chromosomal pair. The two corresponding alleles constitute the genotype for that particular locus. For instance, denoting alleles at a single diallelic locus A_1 and A_2, the possible genotypes for this locus are A_1A_1, A_1A_2, and A_2A_2. Considering two diallelic loci, the possible genotypes are $A_1A_1B_1B_1$, $A_1A_1B_1B_2$, $A_1A_1B_2B_2$, $A_1A_2B_1B_1$, $A_1A_2B_1B_2$, $A_1A_2B_2B_2$, $A_2A_2B_1B_1$, $A_2A_2B_1B_2$, and $A_2A_2B_2B_2$. Homozygosity refers to the presence of identical alleles at both corresponding loci of a chromosomal pair; e.g. A_1A_1 for a single locus, or $A_1A_1B_2B_2$ for two loci. Heterozygosity denotes the opposite (e.g., A_1A_2 or $A_1A_2B_1B_2$). Polygenicity is the phenomenon of a trait being affected by genes at multiple loci. Pleiotropy is the phenomenon of a gene affecting the expression of more than one trait.

Quantitative genetic theory: means, variances, and resemblance between relatives

Genetic applications of SEM are rooted in quantitative genetic theory. In contrast to classical Mendelian genetics, which deals with the inheritance of inter-individual differences in traits

[34] The definition of a gene is becoming increasingly replete with controversy. See e.g. Pearson, 2006.

along which individuals can be divided into distinct categories (e.g., eye color), quantitative or biometrical genetics is concerned with inheritance of inter-individual differences in traits which vary continuously (i.e., quantitative traits, e.g. height). The fact that the intrinsically discontinuous variation in the type of alleles present at the genome may yield continuous variation in observed traits is explained by a) the supposition of polygenic inheritance (quantitative traits are assumed to be affected by genes at multiple genetic loci, whose contribution to the variation in the phenotype is small in comparison to effects of other sources of variation), and b) non-genetic variation, which is truly continuous, being superimposed on the genetic effects on the phenotype. Given the intrinsic reliance of genetic structural equation models on the quantitative genetic theory-based predictions of genetic and environmental covariation between individuals of differing degrees of genetic relatedness, before addressing genetic applications of SEM, we will first review how those predictions are derived (Falconer & Mackay, 1996).

Values and means

The measured value of a trait, or its *phenotypic value* (P), is typically conceptualized as a sum of two components, one attributable to the particular assemblage of segregating genes relevant to the phenotype in question (the *genotypic value*, G), and the other to all of the non-genetic factors affecting the phenotype (*environmental deviation*, E).[35] The mean environmental deviation in the population is typically scaled at zero; thus the mean phenotypic value equals the mean genotypic value. The aim of succeeding sections will be to demonstrate the derivation of the average degree of genetic resemblance between relatives; in this light, we focus primarily on the genotypic value.

Consider, for instance, a single locus with two alleles, A_1 and A_2. The genotypic values of the two homozygotes (A_1A_1 and A_2A_2) and that of the heterozygote (A_1A_2) may be denoted +a, -a, and d, respectively (Figure 1). The point of zero genotypic value is defined as the midpoint between the two homozygotes. The value of d reflects the degree of *genetic dominance*: in the absence of dominance, $d = 0$; if A_1 is dominant over A_2, $d > 0$; if A_2 is dominant over A_1, $d < 0$; in case of overdominance, $d > a$ or $d < -a$.

Figure 1. Arbitrarily assigned genotypic values in a system with 2 alleles (Falconer & Mackay, 1996).

If the relative frequencies with which the A_1 and A_2 alleles occur in the population of interest are denoted p and q, respectively ($p + q = 1$), the frequencies of the genotypes arising from the process of random mating[36] between individuals within this population are

[35] In the text that follows we refer only to the component of P that varies in the population; thus the effects of monomorphic genes, as well as the non-variable aspects of the environment, are ignored (but may be modeled by addition of appropriate constants).

[36] Mating is random if an individual has an equal chance of mating with any other individual in the population. For effects of non-random mating see e.g. Falconer & Mackay (1996).

Table 1
Genotype frequencies in the offspring generation as a function of allele frequencies in the parental generation

		Maternal gametes and their frequencies	
		A *(p)*	A *(q)*
Paternal gametes and their frequencies	A *(p)*	AA (p^2)	Aa *(pq)*
	A *(q)*	Aa *(pq)*	Aa (q^2)

Table 2
Derivation of mean genotypic values (G), breeding values (A) and dominance deviations (D)

		Genotype				
		A_1A_1	A_1A_2	A_2A_2		
Genotypic value (g_i)		a	d	-a		
Genotype frequency(f_i)		p^2	2pq	q^2		
Mean genotypic value ($\mu_{gi} = g_i f_i$)		p^2a	2pqd	$-q^2a$	Mean gen. value across genotypes: $\mu_G = \sum\mu_{gi} = a(p - q) + 2dpq$	
		Frequencies of genotypes produced			Mean values of genotypes	Average effect of allele
		A_1A_1	A_1A_2	A_2A_2	produced (μ_{Gj})	($\alpha_j = \mu_{Gj} - \mu_G$)
Parental gametes	A_1	p	q		pa + qd	$\alpha_1 =$ q[a + d(q – p)]
	A_2		p	q	pd – qa	$\alpha_2 =$ – p[a + d(q – p)]
					Average effect of allele substitution: $\alpha = \alpha_1 - \alpha_2 = a + d(q - p)$	
Breeding value (A_i)		$2\alpha_1 = 2q\alpha$	$\alpha_1 + \alpha_2 = (q - p)\alpha$	$2\alpha_2 = -2p\alpha$	$E[A] = \sum A_i f_i = 2p^2q\alpha + 2pq(q - p)\alpha - 2pq^2\alpha = 2pq\alpha(p + q - p - q) = 0$	
Genotypic value ($G_i = \mu_{gi} - \mu_G$)		$2q(\alpha - qd)$	$(q - p)\alpha + 2pqd$	$-2p(\alpha + pd)$	$E[G] = \sum G_i f_i = 2p^2q(\alpha - qd) + 2pq[(q - p)\alpha + 2pqd] - 2pq^2(\alpha + pd) = 0$	
Dominance deviation ($D_i = G_i - A_i$)		$-2q^2d$	2pqd	$-2p^2d$	$E[D] = \sum D_i f_i = -2p^2q^2d + 4p^2q^2d - 2p^2q^2d = 0$	

given by the binomial expansion $(p + q)^2 = p^2 + 2pq + q^2$, as shown in Table 1.[37] The mean genotypic value of this locus may be obtained by multiplying the value of each genotype by its frequency and summing over the three genotypes:

$$\mu_G = p^2a + 2pqd - q^2a = a(p - q)(p + q) + 2dpq = a(p - q) + 2dpq.$$

The allelic equivalent of the genotypic value is *average effect* of the allele. The average effect of an allele is the mean deviation from the population mean of individuals who received that allele from one parent, the other allele having come at random from the population (Table 2). For instance, if a number of gametes carrying the A_1 allele unite at random with gametes from the population (where p is the frequency of the A_1 allele and q of the A_2 allele), the frequencies of the genotypes produced will be p of A_1A_1 and q of A_1A_2. Taking into account these frequencies and the genotypic values associated with each genotype, the mean genotypic value of the locus may be expressed as pa + qd. Subtracting the population mean from this expression yields the expression for the average effect of the A_1 allele: $\alpha_1 = q[a + d(q - p)]$. Correspondingly, the average effect of A_2 is $\alpha_2 = -p[a + d(q - p)]$. The average effect may also be expressed in terms of the *average effect of gene substitution*, which is simply the difference between the average effects of the two alleles: $\alpha = \alpha_1 - \alpha_2 = a + d(q - p)$. The average effects of the two alleles may be conveyed in terms of the average effect of gene substitution: $\alpha_1 = q\alpha$, and $\alpha_2 = -p\alpha$ (Table 2).

Summing the average effects over the two alleles at a locus yields a component of the genotypic value termed the *breeding value* or the *additive genotype* (A). The remainder of the genotypic value is the *dominance deviation* (D). Dominance deviation arises in the presence of *genetic dominance* (i.e., within-locus interaction between alleles), and reflects the possible non-additive effects arising from the allelic pairing at a locus. It may be derived as D = G – A (Table 2). If the genotypic value refers to an aggregate value of genotypes across more than one locus, the expression for G takes on the form: G = A + D + I, where I stands for *interaction deviation*[38] arising from possible non-additive gene co-action across loci (i.e., *epistasis*).

Variance

Questions formulated within the context of genetic study of quantitative traits pertain predominantly to variation, the basic idea being the decomposition of phenotypic variance into components attributable to different causes. These variance components correspond to components of value described in the last section, so that e.g. the genotypic variance is the variance of the genotypic values. Assuming that the genotypic values and the environmental deviations are not correlated and do not interact, the variance decomposition is:

$$\begin{aligned} V_P &= V_G + V_E \\ &= V_A + V_D + V_I + V_E, \end{aligned}$$

the more general expression being $V_P = V_G + V_E + 2cov_{GE} + V_{GE}$, where cov_{GE} is the covariance between genotypic values and the environmental deviations, and V_{GE} the

[37] p^2, 2pq, and q^2 adequately describe the proportions of genotypes in the offspring generation in populations with no migration, mutation or selection. For effects of migration, mutation and selection see e.g. Falconer & Mackay (1996).

[38] Alleles may interact in pairs or threes or higher numbers. In the expression above, aggregate interactions of all sorts are treated together as a single interaction deviation.

variance due to interaction between genotypes and the environment. The ratio V_A/V_P represents the degree to which the variation in the phenotype is due to variation in the breeding values, and is known as the *heritability coefficient*.

Covariance

The phenotypic covariance between individuals may generally be expressed in terms of the aforementioned variance components. In terms of the covariation between genotypic values, for instance, the resemblance between offspring and parent may be represented as the covariance of the parents' genotypic values of with the mean genotypic values of their offspring. Since the mean value of the offspring is by definition half the breeding value of the parent, the covariance to be deduced is that of the parent's genotypic value (G = A + D) with half of their breeding value (½A): $\text{cov}_{G\_OP} = [\sum ½A(A + D)]/N = (½\sum A^2 + ½\sum AD)/N = ½\sum A^2/N + ½\sum AD/N = ½VA + ½\text{cov}AD$. The covariance between the breeding values and dominance deviations (cov_{AD}) is zero, as can be verified by multiplying each of the breeding values by the corresponding dominance deviation and frequency (given in Table 2) and summing over the three genotypes: $–4p^2q^3\alpha d + 4p^2q^2(q – p)\alpha d + 4p^3q^2\alpha d = 4p^2q^2\alpha d(– q + q – p + p) = 0$. Thus the genetic component of the parent-offspring covariance equals $\mathbf{\text{cov}_{G\_OP} = ½V_A}$.

In full siblings, the mean additive genotypic value of a group of siblings is the mean breeding value of the two parents. Denoting the breeding values of the two parents A and A', the covariance between the additive genotypic values of offspring is $\text{cov}_{A\_FS} = \sum ½(A + A')½(A + A')/N = \sum ¼(A + A')^2/N = \sum ¼(A^2 + 2AA' + A'^2)/N = ¼V_A + ¼V_{A'} + ½\text{cov}_{AA'}$. The assumption of random mating implies that the covariance between the parents' breeding values ($\text{cov}_{AA'}$) is zero. Thus the covariance of the breeding values of full siblings reduces to $\text{cov}_{A\_FS} = ¼V_A + ¼V_{A'}$. If the additive genetic variance is equal in the two sexes, this expression becomes $\text{cov}_{A\_FS} = ¼V_A + ¼V_A = ½V_A$. In addition, if parental genotypes at a single locus are A_1A_2 and A_3A_4, the offspring may have one of the four possible genotypes: A_1A_2, A_1A_4, A_2A_3, and A_2A_4. If the first sibling has any of these genotypes, the probability that the second sibling has the same genotype is ¼. Thus, one quarter of full siblings have the same genotype for this locus, and consequently the same dominance deviation. For these pairs, the covariance due to dominance deviations is $\text{cov} = \sum D^2/N = V_D$. In other pairs the covariance due to dominance deviations is zero. Thus, over all pairs of siblings, the covariance due to dominance deviations is $¼V_D$. The total genotypic covariance between full siblings is therefore $\mathbf{\text{cov}_{G\_FS} = ½V_A + ¼V_D}$.

The same expression holds true for dizygotic (fraternal, DZ) twins, whose degree of genetic relatedness is the same as that of full siblings. Monozygotic (identical, MZ) twins have identical genotypes and therefore share their entire genotypic variance, thus $\mathbf{\text{cov}_{G\_MZ} = V_A + V_D}$.

References

Abdellaoui, A., Hottenga, J.-J. de Knijff, P., Nivard, M.G., Xiao, X., Scheet, P., . . . Davies, G.E. (2013). Population structure, migration, and diversifying selection in the Netherlands. *European Journal of Human Genetics, 21*(11), 1277-1285.

Achenbach, T. M. (1966). Classification of childrens psychiatric symptoms - a factor-analytic study. *Psychological Monographs, 80*(7), 1-37.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. (2001). *ASEBA School-Age Forms & Profiles*: Aseba Burlington.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30-60.

Allen Lango, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., . . . Raychaudhuri, S. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*(7317), 832-838.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*(1).

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental health disorders* (4th ed.). Washington DC: American Psychiatric Association

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(1), 57-87.

Antonarakis, S. E., & Beckmann, J. S. (2006). Opinion - Mendelian disorders deserve more attention. *Nature Reviews Genetics, 7*(4), 277-282.

Arden, R., Gottfredson, L.S., Miller, G., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence, 37*(3), 277-282.

Bagshaw, A.T.M., Horwood, L.J., Liu, Y., Fergusson, D.M., Sullivan, P.F., & Kennedy, M.A. (2013). No effect of genome-wide copy number variation on measures of intelligence in a New Zealand birth cohort. *PLoS ONE, 8*(1), e55208.

Bakaysa, S.L., Mucci, L.A., Slagboom, P.E., Boomsma, D.I., McClearn, G. E., Johansson, B., & Pedersen, N.L. (2007). Telomere length predicts survival independent of genetic influences. *Aging cell, 6*(6), 769-774.

Banaschewski, T., Becker, K., Scherag, S., Franke, B., & Coghill, D. (2010). Molecular genetics of attention-deficit/hyperactivity disorder: an overview. *European Child & Adolescent Psychiatry, 19*(3), 237-257.

Banks, G.C., Batchelor, J.H., & McDaniel, M.A. (2010). Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry. *Intelligence, 38*(4), 393-401.

Bartels, M., Boomsma, D.I., Hudziak, J. J., van Beijsterveldt, Tcem, & van den Oord, E.J.C.G. (2007). Twins and the study of rater (dis)agreement. *Psychological Methods, 12*(4), 451-466.

Bartels, M., van Beijsterveldt, C.E.M., Derks, E.M., Stroet, T.M., Polderman, T.J.C., Hudziak, J.J., & Boomsma, D.I. (2007). Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10*(1), 3-11.

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D.I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics, 32*(4), 237-249.

Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D.I. (2009). Breastfeeding, Maternal Education and Cognitive Function: A Prospective Study in Twins. *Behavior Genetics, 39*(6), 616-622.

Bartels, M., van Beijsterveldt, C. E. M., Derks, E. M., Stroet, T. M., Polderman, T. J. C., Hudziak, J. J., & Boomsma, D.I. (2007). Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10*(1), 3-11.

Batty, G.D., Deary, I.J., & Gottfredson, L.S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Annals of epidemiology, 17*(4), 278-288.

Bauer, D. J., & Hussong, A. M. (2009). Psychometric Approaches for Developing Commensurate Measures Across Independent Studies: Traditional and New Models. *Psychological Methods, 14*(2), 101-125.

Benyamin, B., Pourcain, B.S., Davis, O.S., Davies, G., Hansell, N.K., Brion, M.-J.A., . . . Visscher, P.M. (2013). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry, 19*, 253–258.

Benyamin, B, Pourcain, BSt, Davis, OS, Davies, G, Hansell, NK, Brion, M-JA, . . . Haworth, CMA. (2013). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular psychiatry*.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics, 10*(3), 423-433.

Bishop, E.G., Cherny, S.S., Corley, R., Plomin, R., DeFries, J.C., & Hewitt, J.K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. *Intelligence, 31*(1), 31-49.

Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., . . . Brunak, S. (2013). A nondegenerate code of deleterious variants in mendelian Loci contributes to complex disease risk. *Cell, 155*(1), 70-80.

Bleichrodt, N., Drenth, P.J.D., Zaal, J.N., & Resing, W.C.M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test, RAKIT*: Lisse: Swets & Zeitlinger.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117*(2), 187.

Block, J.H., & Block, J. (1980). *The role of ego-control and ego-resiliency in the organization of behavior.* Paper presented at the Minnesota symposium on child psychology.

Boker, S., Neale, M.C., Maes, H., Metah, P., Kenny, S., Bates, T., . . . Spiegel, M. (2010). OpenMx: The OpenMx Statistical Modeling Package. Retrieved from http://openmx.psyc.virginia.edu

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Boomsma, D. I., & van Baal, G.C.M. (1998). Genetic influences on childhood IQ in 5-and 7-year-old Dutch twins. *Developmental Neuropsychology, 14*(1), 115-126.

Boomsma, D. I., Vink, J. M., van Beijsterveldt, Tcem, de Geus, E. J. C., Beem, A. L., Mulder, Ejcm, . . . van Baal, G. C. M. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research, 5*(5), 401-406.

Boomsma, D.I. (1998). Twin registers in Europe: an overview. *Twin Res, 1*(1), 34-51.

Boomsma, D.I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*(11), 872-882.

Boomsma, D.I., de Geus, E. J. C., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., . . . Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9*(6), 849-857.

Boomsma, D.I., de Geus, E. J., van Baal, G. C., & Koopmans, J. R. (1999). A religious upbringing reduces the influence of genetic factors on disinhibition: evidence for interaction between genotype and environment on personality. *Twin Res, 2*(2), 115-125.

Boomsma, D.I., Martin, N. G., & Molenaar, P. C. M. (1989). Factor and Simplex Models for Repeated Measures: Application to Two Psychomotor Measures of Alcohol Sensitivity in Twins. *Behavior Genetics, 19*(1), 79-96.

Boomsma, D.I., Martin, N. G., & Neale, M.C. (1989). Genetic analysis of twin and family data: Structural modeling using LISREL. *Behavior Genetics, 19*(1), 5-7.

Boomsma, D.I., & Molenaar, P. C. M. (1986). Using lisrel to analyze genetic and environmental covariance structure. *Behavior Genetics, 16*(2), 237-250.

Boomsma, D.I., & Molenaar, P. C. M. (1987). The genetic-analysis of repeated measures. I. Simplex models. *Behavior Genetics, 17*(2), 111-123.

Boomsma, D.I., & van Baal, G.C.M. (1998). Genetic influences on childhood IQ in 5 and 7 year old Dutch twins. *Developmental Neuropsychology, 14*(1), 115-126.

Boomsma, D.I., Van Beijsterveldt, C.E.M., Bartels, M., & Hudziak, J.J. (2008). Genetic and environmental influences on anxious/depression: A longitudinal study in 3- to 12-year-old children. In J. J. Hudziak (Ed.), *Developmental Psychopathology and Wellness*. Washington DC: American Psychiatric Publishing.

Boomsma, D.I., Van Beijsterveldt, C.E.M., & Hudziak, J.J. (2005a). Genetic and environmental influences on Anxious/Depression during childhood: a study from the Netherlands Twin Register. *Genes, Brain and Behavior, 4*(8), 466-481.

Boomsma, D.I., van Beijsterveldt, C.E.M., & Hudziak, J.J. (2005b). Genetic and environmental influences on Anxious/Depression during childhood: a study from the Netherlands Twin Register. *Genes Brain and Behavior, 4*(8), 466-481.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F.M. (2002). Similarity of childhood experiences and personality resemblance in monozygotic and dizygotic twins: A test of the equal environments assumption. *Personality and Individual Differences, 33*(2), 261-269.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64*(9), 1089-1108.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219.

Bouchard T.J. Jr, , & Loehlin, J.C. (2001). Genes, evolution, and personality. *Behavior Genetics, 31*(3), 243-273.

Bouchard, T.J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212*(4498), 1055-1059.

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin, 111*(2), 244-255.

Brendgen, M., Vitaro, F., Boivin, M., Girard, A., Bukowski, W. M., Dionne, G., . . . Perusse, D. (2009). Gene-environment interplay between peer rejection and depressive behavior in children. *Journal of Child Psychology and Psychiatry, 50*(8), 1009-1017.

Brown, T. A. . (1996). Validity of the DSM-III-R and DSM-IV classification systems for anxiety disorders. In R. M. Rapee (Ed.), *Current controversies in the anxiety disorders* (pp. 21-45). New York: Guilford Press.

Burt, S. A., McGue, M., Carter, L. A., & Iacono, W. G. (2007). The different origins of stability and change in antisocial personality disorder symptoms. *Psychological Medicine, 37*(1), 27-38.

Busjahn, A. (2002). Twin registers as a global resource for genetic research [Special Issue]. *Twin Research, 5*, 317–506.

Butcher, L.M., Davis, O.S.P., Craig, I.W., & Plomin, R. (2008). Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes, Brain and Behavior, 7*(4), 435-446.

Cantor, R. M. (1983). A multivariate genetic-analysis of ridge count data from the offspring of monozygotic twins. *Acta Geneticae Medicae Et Gemellologiae, 32*(3-4), 161-207.

Caprara, G.V., Barbaranelli, C., & Comrey, A.L. (1995). Factor analysis of the Neo-PI Inventory and the Comrey Personality Scales in an italian sample. *Personality and Individual Differences, 18*(2), 193-200.

Cardon, L.R., & Palmer, L.J. (2003). Population stratification and spurious allelic association. *The Lancet, 361*(9357), 598-604.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16*(3), 319-341.

Carey, G. (1992). Twin imitation for antisocial-behavior - implications for genetic and family environment research. *Journal of Abnormal Psychology, 101*(1), 18-25.

Carey, G. (2009). *The problem with c2*. Paper presented at the Annual meeting of the Behavior Genetics Association, Minneapolis, MN, USA.
http://psych.colorado.edu/~carey/Bouchard/docs/Environment in BG3.pdf

Carey, V.J., Lumley, T., & Ripley, B. (2012). Generalized Estimation Equation solver. Retrieved from http://cran.r-project.org/web/packages/gee/index.html

Carlson, M., & Mulaik, S. A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28*(1), 111-159.

Carroll, J. B. . (2003). The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-21). New York: Pergamon.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., . . . Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386-389.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bullentin, 38*, 592.

Cattell, R. B. (1943). The description of personality. I. Foundations of trait measurement. *Psychological Review, 50*(6), 559.

Cattell, R. B. (1945). The description of personality: Principles [sic] findings in a factor analysis. *American Journal of Psychology, 58,*, 69-90.
Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annu. Rev. Psychol., 56*, 423-452.
Chabris, C.F., Hebert, B.M, Benjamin, D.J., Beauchamp, J., Cesarini, D., van der Loos, M., . . . Atwood, C.S. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science, 23*(11), 1314-1323.
Chen, W.-M., & Abecasis, G.R. (2007). Family-based association tests for genomewide association scans. *The American Journal of Human Genetics, 81*(5), 913-926.
Cherny, S. S. (2008). Variance Components and Related Methods for Mapping Quantitative Trait Loci. *Sociological Methods Research, 37*, 227-250.
Ciesla, J. A., & Roberts, J. E. (2007). Rumination, negative cognition, and their interactive effects on depressed mood. *Emotion, 7*(3), 555-565.
CITO. (2002). Eindtoets Basisonderwijs. Arnhem: Citogroep.
Clark, L. A. (1989). The anxiety and depressive disorders: Descriptive psychopathology and differential diagnosis. In P. C. Kendall & D. Watson (Eds.), *Anxiety and depression: Distinctive and overlapping features* (pp. 83-129). New York: Academic Press.
Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression - psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316-336.
Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., & Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science, 305*(5685), 869-872.
Costa, P.T., & McCrae, R.R. (1985). *The NEO personality inventory: Manual, form S and form R*. Odessa: Psychological Assessment Resources.
Costa, P.T., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa: Psychological Assessment Resources.
Costa, P.T., & McCrae, R.R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). *The SAGE handbook of personality theory and assessment, 2*, 179-198. London: Sage.
Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences, 33*(2-3), 137–193.
Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., . . . Sherry, S.T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.
Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., . . . Luciano, M. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry, 16*(10), 996-1005.
de Moor, M.H.M., Boomsma, D.I., Stubbe, J. H., Willemsen, G., & de Geus, E. J. C. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry, 65*(8), 897-905.
de Moor, M.H.M., Costa, P.T., Terracciano, A., Krueger, R.F., De Geus, E.J.C., Toshiko, T., . . . Derringer, J. (2010). Meta-analysis of genome-wide association studies for personality. *Molecular Psychiatry, 17*(3), 337-349.
Deary, I.J., Johnson, W., & Houlihan, L.M. (2009). Genetic foundations of human intelligence. *Human genetics, 126*(1), 215-232.
Deary, I.J., Spinath, F.M., & Bates, T.C. (2006). Genetics of intelligence. *European Journal of Human Genetics, 14*(6), 690-700.
Deary, I.J., Whiteman, M.C., Starr, J.M., Whalley, L.J., & Fox, H.C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*(1), 130.
Deary, I.J, Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21.
Deary, I.J., Whalley, L.J., Batty, G.D., & Starr, J.M. (2006). Physical fitness and lifetime cognitive change. *Neurology, 67*(7), 1195-1200.
Deary, I.J., Whiteman, M.C., Pattie, A., Starr, J.M., Hayward, C., Wright, A.F., . . . Whalley, L.J. (2002). Ageing: Cognitive change and the APOE ϵ 4 allele. *Nature, 418*(6901), 932-932.

de Kort, J.M., Dolan, C.V., Kan, K.-J., van Beijsterveldt, C.E.M., Bartels, M., & Boomsma, D.I. (2014). Can GE-Covariance Originating in Phenotype to Environment Transmission Account for the Flynn Effect? *Journal of Intelligence, 2*(3), 82-105.

Derks, E. M., Dolan, C.V., & Boomsma, D.I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9*(3), 403-411.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature, 388*(6641), 468-471.

DeYoung, C.G., Hirsh, J.B., Shane, M.S., Papademetris, X., Rajeevan, N., & Gray, J.R. (2010). Testing Predictions From Personality Neuroscience Brain Structure and the Big Five. *Psychological Science, 21*(6), 820-828.

Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., & Goldstein, D.B. (2010). Rare variants create synthetic genome-wide associations. *PLoS biology, 8*(1), e1000294.

Dolan, C.V. (1992). *Biometric decomposition of phenotypic means in human samples.* (PhD doctoral dissertation), University of Amsterdam, Amsterdam.

Dolan, C.V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*(2), 309-326.

Dolan, C.V., Boomsma, D.I., & Neale, M.C. (1999). A note on the power provided by sibships of size 3 and 4 in genetic covariance modeling of a codominant QTL. *Behavior Genetics, 29*, 163-170.

Dolan, C.V., de Kort, J.M., van Beijsterveldt, C.E.M., Bartels, M., & Boomsma, D.I. (2014). GE Covariance through phenotype to environment transmission: An assessment in longitudinal twin data and application to childhood anxiety. *Behavior Genetics, 44*(3), 240-253.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics, 9*(3), e1003348.

Eaves, L.J. (1979). Use of twins in the analysis of assortative mating. *Heredity, 43*(Dec), 399-409.

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. (1983). Errors of inference in the detection of major gene effects on psychological test-scores. *American Journal of Human Genetics, 35*(6), 1179-1189.

Eaves, L.J. (1984). The resolution of genotype × environment interaction in segregation analysis of nuclear families. *Genetic Epidemiology, 1*, 215–228.

Eaves, L.J. (1987). Including the environment in models for genetic segregation. *Journal of Psychiatric Research, 21*(4), 639-647.

Eaves, L.J., Fulker, D. W., & Heath, A. C. (1989). The effects of social homogamy and cultural inheritance on the covariances of twins and their parents - a LISREL model. *Behavior Genetics, 19*(1), 113-122.

Eaves, L.J., Heath, A. C., Neale, M.C., Hewitt, J. K., & Martin, N. G. (1998). Sex differences and non-additivity in the effects of genes on personality. *Twin Res, 1*(3), 131-137.

Eaves, L.J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human-behavior. *Heredity, 41*(Dec), 249-320.

Eaves, L.J., Last, K., Martin, N. G., & Jinks, J. L. (1977). Progressive approach to non-additivity and genotype-environmental covariance in analysis of human differences. *British Journal of Mathematical & Statistical Psychology, 30*(May), 1-42.

Eaves, L.J., Long, J., & Heath, A. C. (1986). A theory of developmental-change in quantitative phenotypes applied to cognitive-development. *Behavior Genetics, 16*(1), 143-162.

Eaves, L.J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A., . . . Hewitt, J. K. (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. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(8), 965-980.

Eaves, L.J. (1973). Assortative mating and intelligence: An analysis of pedigree data. *Heredity, 30*(2), 199-210.

Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., & Nadeau, J.H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics, 11*(6), 446-450.

Eley, T. C., Bolton, D., O'Connor, T. G., Perrin, S., Smith, P., & Plomin, R. (2003). A twin study of anxiety-related behaviours in pre-school children. *Journal of Child Psychology and Psychiatry, 44*(7), 945-960.

Eley, T. C., Liang, H. L., Plomin, R., Sham, P., Sterne, A., Williamson, R., & Purcell, S.M. (2004). Parental familial vulnerability, family environment, and their interactions as predictors of depressive symptoms in adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(3), 298-306.

Eley, T. C., & Stevenson, J. (1999). Exploring the covariation between anxiety and depression symptoms: A genetic analysis of the effects of age and sex. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(8), 1273-1282.

Epskamp, S., Cramer, A.O. J., Waldorp, L.J., Schmittmann, V.D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software, 48*(4), 1-18.

Evans, D. M., Gillespie, N. A., & Martin, N. G. (2002). Biometrical genetics. *Biological Psychology, 61*(1-2), 33-51.

Exil, V.J., Avila, D.S., Benedetto, A., Exil, E.A., Adams, R., Au, C., & Aschner, M. (2010). Stressed-Induced TMEM135 Protein Is Part of a Conserved Genetic Network Involved in Fat Storage and Longevity Regulation in Caenorhabditis elegans. *PLoS ONE, 5*(12), e14228.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th ed.). Harlow, Essex, UK: Longmans Green.

Feinberg, M. E., Button, T. M. M., Neiderhiser, J. M., Reiss, D., & Hetherington, E. M. (2007). Parenting and adolescent antisocial behavior and depression - Evidence of genotype x parenting environment interaction. *Archives of General Psychiatry, 64*(4), 457-465.

Fellows, J., Erdjument-Bromage, H., Tempst, P., & Svejstrup, J.Q. (2000). The Elp2 subunit of elongator and elongating RNA polymerase II holoenzyme is a WD40 repeat protein. *Journal of Biological Chemistry, 275*(17), 12896-12899.

Ferreira, M. A. R., & Purcell, S.M. (2009). A multivariate test of association. *Bioinformatics, 25*(1), 132-133.

Fisk, J.C., Zurita-Lopez, C., Sayegh, J., Tomasello, D.L., Clarke, S.G., & Read, L.K. (2010). TbPRMT6 is a type I protein arginine methyltransferase that contributes to cytokinesis in Trypanosoma brucei. *Eukaryotic cell, 9*(6), 866-877.

Flint, J. (2013). GWAS. *Current biology: CB, 23*(7), R265-R266.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491.

Flynn, James R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171.

Franić, S., Dolan, C.V., Borsboom, D., & Boomsma, D.I. (2012). Structural Equation Modeling in Genetics. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 617-635). New York: Guilford Press.

Franić, S., Dolan, C.V., Borsboom, D., Hudziak, J.J., van Beijsterveldt, C.E.M., & Boomsma, D.I. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological Methods, 18*(3), 406.

Franić, S., Dolan, C.V., van Beijsterveldt, C.E.M., Pol, H.E.H., Bartels, M., & Boomsma, D.I. (2014). Genetic and Environmental Stability of Intelligence in Childhood and Adolescence. *Twin Research and Human Genetics, 17*(03), 151-163.

Franić, S., Dolan, C.V., Borsboom, D., van Beijsterveldt, C.E.M., & Boomsma, D.I. (2014). Three-and-a-Half-Factor Model? The Genetic and Environmental Structure of the CBCL/6–18 Internalizing Grouping. *Behavior Genetics, 44*(3), 254-268.

Franić, S., Groen-Blokhuis, M.M., Dolan, C.V., Kattenberg, M.V., Xiao, X., Scheet, P.A., . . . Boomsma, D.I. (2013). IQ: Shared Genetic Basis between Mendelian Disorders and a Polygenic Trait. *Under review*.

Frazer, K.A., Murray, S.S., Schork, N.J., & Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics, 10*(4), 241-251.

Freathy, R.M., Mook-Kanamori, D.O., Sovio, U., & Prokopenko, I. (2010). Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nature Genetics, 42*(5), 430-435.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., . . . Pato, C.N. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics, 36*(4), 388-393.

French, J.W. (1953). The description of personality measurements in terms of rotated factors.
Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B., & Tybjærg-Hansen, A. (2004). Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *Journal of Clinical Investigation, 114*(9), 1343-1353.
Fulker, D. W., Baker, L. A., & Bock, R. D. (1983). Estimating components of covariance using LISREL. *Data Analyst, 1*, 5-8.
Fulker, D. W., Cherny, S. S., Sham, P. C., & Hewitt, J. K. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics, 64*(1), 259-267.
Gale, C.R., Batty, G.D., Tynelius, P., Deary, I.J., & Rasmussen, F. (2010). Intelligence in early adulthood and subsequent hospitalization for mental disorders. *Epidemiology, 21*(1), 70-77.
Gasser, T. (2009). Mendelian forms of Parkinson's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1792*(7), 587-596.
Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65.
Gianola, D., & Sorensen, D. (2004). Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics, 167*(3), 1407-1424.
Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics, 13*(2), 135-145.
Gillespie, N. A., Evans, D. E., Wright, M. M., & Martin, N. G. (2004). Genetic simplex modeling of Eysenck's dimensions of personality in a sample of young Australian twins. *Twin Research, 7*(6), 637-648.
Gillespie, N. A., Kirk, K. M., Evans, D. M., Heath, A. C., Hickie, I. B., & Martin, N. G. (2004). Do the genetic or environmental determinants of anxiety and depression change with age? A longitudinal study of Australian twins. *Twin Research, 7*(1), 39-53.
Goldberg, L.R. (1977). Language and personality: Developing a taxonomy of trait descriptive terms. San Francisco.
Goldberg, L.R. (1980). Some ruminations about the structure of individual differences: Developing a common lexicon for the major characteristics of personality. Honolulu.
Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology, 2*(1), 141-165.
Goldberg, L.R. (1982). From Ace to Zombie: Some explorations in the language of personality. *Advances in personality assessment, 1*, 203-234.
Goldberg, L.R. (1983). The magical number five, plus or minus two: Some conjectures on the dimensionality of personality descriptions. Baltimore, MD.
Goldberg, L.R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.
Goldberg, L.R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment, 4*(1), 26-42.
Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-26.
Goldstein, D.B., Allen, A., Keebler, J., Margulies, E.H., Petrou, S., Petrovski, S., & Sunyaev, S. (2013). Sequencing studies in human genetics: design and interpretation. *Nature Reviews Genetics*.
Goodwin, R. D., Fergusson, D. M., & Horwood, L. J. (2004). Early anxious/withdrawn behaviours predict later internalising disorders. *Journal of Child Psychology and Psychiatry, 45*(4), 874-883.
Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: Erlbaum.
Gottfredson, L.S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13-23.
Gottfredson, L.S. (1997b). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132.
Gottfredson, L.S, & Deary, I.J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13*(1), 1-4.
Grada, A., & Weinbrecht, K. (2013). Next-Generation Sequencing: Methodology and Application. *J Invest Dermatol, 133*(8), e11. doi: 10.1038/jid.2013.248
Gregory, A.M., & Eley, T.C. (2007). Genetic influences on anxiety in children: what we've learned and where we're heading. *Clinical child and family psychology review, 10*(3), 199-212.
Guo, G., & Adkins, D. E. (2008). How Is a Statistical Link Established Between a Human Outcome and a Genetic Variant? *Sociological Methods & Research, 37*(2), 201-226.
Guttman, L. (1954). A new approach to factor analysis: the Radex.

Hahn, R., & Comrey, A.L. (1994). Factor analysis of the NEO-PI and the Comrey Personality Scales. *Psychological Reports, 75*(1), 355-365.
Haig, B. D. (2005a). An abductive theory of scientific method. *Psychological Methods, 10*(4), 371-388.
Haig, B. D. (2005b). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research, 40*(3), 303-329.
Hammen, C. (1992). Cognitive, life stress, and interpersonal approaches to a developmental psychopathology model of depression *Development and Psychopathology, 4*(1), 189-206.
Happonen, M., Pulkkinen, L., Kaprio, J., Van der Meere, J., Viken, R. J., & Rose, R. J. (2002). The heritability of depressive symptoms: multiple informants and multiple measures. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*(4), 471-479.
Harris, K. M., Halpern, C. T., Smolen, A., & Haberstick, B. C. (2006). The National Longitudinal Study of Adolescent Health (Add health) twin data. *Twin Research and Human Genetics, 9*(6), 988-997.
Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G. J., . . . Sergeant, J. A. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: a critical empirical evaluation. *Journal of Child Psychology and Psychiatry, 40*(7), 1095-1116.
Haworth, C.M.A., & Davis, O.S.P. (2014). From observational to dynamic genetics. *Frontiers in Genetics, 5*.
Haworth, C.M.A., Wright, M.J., Luciano, M., Martin, N. G., De Geus, E.J.C., Van Beijsterveldt, C.E.M., . . . Davis, O.S.P. (2009a). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*(11), 1112-1120.
Heath, A. C., & Eaves, L.J. (1985). Resolving the effects of phenotype and social background on mate selection. *Behavior Genetics, 15*(1), 15-30.
Heath, A. C., Eaves, L.J., & Martin, N. G. (1989). The genetic-structure of personality .3. multivariate genetic item analysis of the epq scales. *Personality and Individual Differences, 10*(8), 877-888.
Heath, A. C., Jardine, R., Eaves, L.J., & Martin, N. G. (1989). The genetic-structure of personality .2. genetic item analysis of the epq. *Personality and Individual Differences, 10*(6), 615-624.
Heath, A. C., Kessler, R. C., Neale, M.C., Hewitt, J. K., Eaves, L.J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics, 23*(1), 29-50.
Heijmans, B.T., Kremer, Dennis, T., Elmar, W., Boomsma, D.I., & Slagboom, P.E. (2007). Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Human molecular genetics, 16*(5), 547-554.
Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research - Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416.
Hettema, J. M., Neale, M.C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry, 158*(10), 1568-1578.
Hettema, J. M., Prescott, C. A., & Kendler, K. S. (2004). Genetic and environmental sources of covariation between generalized anxiety disorder and neuroticism. *American Journal of Psychiatry, 161*(9), 1581-1587.
Hewitt, J. K., Eaves, L.J., Neale, M.C., & Meyer, J. M. (1988). Resolving Causes of Developmental Continuity or "Tracking." I. Longitudinal Twin Studies During Growth. *Behavior Genetics, 18*(2), 133-151.
Hicks, B. M., DiRago, A. C., Iacono, W. G., & McGue, M. (2009). Gene-environment interplay in internalizing disorders: consistent findings across six environmental risk factors. *Journal of Child Psychology and Psychiatry, 50*(10), 1309-1317.
Hirschhorn, J.N., & Gajdos, Z.K.Z. (2011). Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annual Review of Medicine, 62*, 11-24.
Hjelmborg, J. V. B., Fagnani, C., Silventoinen, K., McGue, M., Korkeila, M., Christensen, K., . . . Kaprio, J. (2008). Genetic influences on growth traits of BMI: A longitudinal study of adult twins. *Obesity, 16*(4), 847-852.
Hoekstra, H.A., Ormel, J., & De Fruyt, F. (1996). *NEO Personality Questionnaires NEO-PI-R, NEO-FFI: Manual*. Lisse: Swet & Zeitlinger BV.
Hoekstra, R. A., Bartels, M., & Boomsma, D.I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences, 17*(2), 97-114.

Hoekstra, R. A., Bartels, M., Hudziak, J. J., Van Beijsterveldt, C. E. M., & Boomsma, D.I. (2008). Genetic and environmental influences on the stability of withdrawn behavior in children: A longitudinal, multi-informant twin study. *Behavior Genetics, 38*(5), 447-461.

Hoischen, A., van Bon, B.W.M., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., . . . Mortier, G. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature Genetics, 42*(6), 483-485.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities*. doctoral dissertation. University of Illinois. Champaign.

Hottenga, J.J., & Boomsma, D.I. (2008). QTL detection in multivariate data from sibling pairs. In B. M. Neale, M. A. R. Ferreira, S. E. Medland & D. Posthuma (Eds.), *Statistical Genetics. Gene Mapping Through Linkage and Association* (pp. 239-258). New York: Taylor & Francis Group.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics, 44*(8), 955-959.

Hsu, S. (2012). *Genetic architecture of intelligence from SNP distance measures*. Paper presented at the Behavior Genetics Association Meeting, Edinburgh, Scotland.

Hudziak, J. J., Van Beijsterveldt, C. E. M., Althoff, R. R., Stanger, C., Rettew, D. C., Nelson, E. C., . . . Boomsma, D.I. (2004). Genetic and environmental contributions to the child behavior checklist Obsessive-Compulsive Scale - A cross-cultural twin study. *Archives of General Psychiatry, 61*(6), 608-616.

Ilardi, J.M., Mochida, S., & Sheng, Z.-H. (1999). Snapin: a SNARE–associated protein implicated in synaptic transmission. *Nature neuroscience, 2*(2), 119-124.

Inlow, J. K., & Restifo, L. L. (2004). Molecular and comparative genetics of mental retardation. *Genetics, 166*(2), 835-881.

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., . . . Rider, L.G. ((2010). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome research, 20*(2), 170-179.

Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.

Jinks, J. L., & Fulker, D. W. (1970). Comparison of biometrical genetical, MAVA, and classical approaches to analysis of human behavior. *Psychological Bulletin, 73*(5), 311-349.

Johnson, W., & Krueger, R. F. (2004). Genetic and environmental structure of adjectives describing the domains of the Big Five Model of personality: A nationwide US twin study. *Journal of Research in Personality, 38*(5), 448-472.

Johnson, W., & Krueger, R. F. (2005). Genetic effects on physical health: Lower at higher income levels. *Behavior Genetics, 35*(5), 579-590.

Johnson, W., Bouchard T.J. Jr, ., McGue, M., Segal, N.L., Tellegen, A., Keyes, M., & Gottesman, I.I. (2007). Genetic and environmental influences on the Verbal-Perceptual-Image Rotation (VPR) model of the structure of mental abilities in the Minnesota study of twins reared apart. *Intelligence, 35*(6), 542-562.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121-145.

Jöreskog, K. G. (1993). Testing Structural Equation Models. In K. A. Bollen & S. J. Long (Eds.), *Testing Structural Equation Models* (pp. 294-316). Newbury Park: Sage.

Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631-639.

Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.80 for Windows. Lincolnwood: Scientific Software International, Inc.

Kan, K.-J., Wicherts, J.M., Dolan, C.V., & van der Maas, H. L. J. (2013). On the Nature and Nurture of Intelligence and Specific Cognitive Abilities The More Heritable, the More Culture Dependent. *Psychological Science, 24*(12), 2420-2428.

Keller, M.C., & Coventry, W.L. (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics, 8*(3), 201-213.

Keller, Matthew C, Medland, Sarah E, & Duncan, Laramie E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics, 40*(3), 377-393.

Kendler, K. S. (1996). Parenting: A genetic-epidemiologic perspective. *American Journal of Psychiatry, 153*(1), 11-20.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological Medicine, 37*(5), 615-626.
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*(3), 279-289.
Kendler, K. S., Gardner, C. O., Annas, P., & Lichtenstein, P. (2008). The development of fears from early adolesence to young adulthood: a multivariate study. *Psychological Medicine, 38*(12), 1759-1769.
Kendler, K. S., Gardner, C. O., & Lichtenstein, P. (2008). A developmental twin study of symptoms of anxiety and depression: evidence for genetic innovation and attenuation. *Psychological Medicine, 38*(11), 1567-1575.
Kendler, K. S., Heath, A. C., Martin, N. G., & Eaves, L.J. (1987). Symptoms of anxiety and symptoms of depression - same genes, different environments. *Archives of General Psychiatry, 44*(5), 451-457.
Kendler, K. S., Jacobson, K. C., Gardner, C. O., Gillespie, N., Aggen, S. A., & Prescott, C. A. (2007). Creating a social world - A developmental twin study of peer-group deviance. *Archives of General Psychiatry, 64*(8), 958-965.
Kendler, K. S., & Karkowski-Shuman, L. (1997). Stressful life events and genetic liability to major depression: Genetic control of exposure to the environment. *Psychological Medicine, 27*(3), 539-547.
Kendler, K. S., Neale, M.C., Kessler, R. C., Heath, A. C., & Eaves, L.J. (1992). Major depression and generalized anxiety disorder - same genes, (partly) different environments. *Archives of General Psychiatry, 49*(9), 716-722.
Kendler, K. S., Neale, M.C., Kessler, R. C., Heath, A. C., & Eaves, L.J. (1993). A test of the equal-environment assumption in twin studies of psychiatric-illness. *Behavior Genetics, 23*(1), 21-27.
Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine, 41*(06), 1143-1150.
Kenny, D.A. (2012). Measuring Model Fit. from http://davidakenny.net/cm/fit.htm
Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., . . . Moran, J.L. (2012). Exome sequencing and the genetic basis of complex traits. *Nature Genetics, 44*(6), 623-630.
Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: Guilford Press.
Kryukov, G.V., Pennacchio, L.A., & Sunyaev, S.R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *The American Journal of Human Genetics, 80*(4), 727-739.
Ku, C.-S., Naidoo, N., & Pawitan, Y. (2011). Revisiting Mendelian disorders through exome sequencing. *Human Genetics, 129*(4), 351-370.
Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., . . . Siepel, A. (2007). The UCSC genome browser database: update 2007. *Nucleic acids research, 35*(suppl 1), D668-D673.
Lamb, D.J., Middeldorp, C.M., Van Beijsterveldt, C.E.M., Bartels, M., Polderman, T.C. , & Boomsma, D.I. (2010). Heritability of Anxious-Depressive and Withdrawn Behavior: Age-Related Changes During Adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(3), 248-255.
Lau, J. Y. F., & Eley, T. C. (2008a). Attributional Style as a Risk Marker of Genetic Effects for Adolescent Depressive Symptoms. *Journal of Abnormal Psychology, 117*(4), 849-859.
Lau, J. Y. F., & Eley, T. C. (2008b). Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. *Journal of Child Psychology and Psychiatry, 49*(2), 142-150.
Lawley, D. N., & Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method* (2nd ed.). London: Butterworths.
Legrand, L.N., McGue, M., & Iacono, W.G. (1999). A twin study of state and trait anxiety in childhood and adolescence. *Journal of Child Psychology and Psychiatry, 40*(6), 953-958.
Lenroot, R. K., Schmitt, J. E., Ordaz, S. J., Wallace, G. L., Neale, M.C., Lerch, J. P., . . . Giedd, J. N. (2009). Differences in Genetic and Environmental Influences on the Human Cerebral Cortex Associated With Development During Childhood and Adolescence. *Human Brain Mapping, 30*(1), 163-174.
Lepage, P., Häsler, R., Spehlmann, M.E., Rehman, A., Zvirbliene, A., Begun, A., . . . Raedler, A. (2011). Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology, 141*(1), 227-236.

Lesage, S., & Brice, A. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human molecular genetics, 18*(R1), R48-R59.
Li, M.-X., Gui, H.-S., Kwan, J.S.H., & Sham, P.C. (2011). GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *American Journal of Human Genetics, 88*(3), 283-293.
Li, Y., & Abecasis, G.R. (2006). Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet S, 79*(3), 2290.
Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. . (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology - Science and Practice, 1*, 71-83.
Lin, B., Mbarek, H., Willemsen, G., Dolan, C.V., Fedko, I., de Geus, E.J.C., . . . Hottenga, J.J. (2014). GWAS study, GCTA study and twins study to estimate heritability of hair color in European population. *In preparation*.
Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods, 8*(10), 833-835.
Liu, E.Y., Li, M., Wang, W., & Li, Y. (2013). MaCH-Admix: Genotype Imputation for Admixed Populations. *Genetic epidemiology, 37*(1), 25-37.
Liu, J.Z., McRae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., . . . Macgregor, S. (2010). A Versatile Gene-Based Test for Genome-wide Association Studies. *American Journal of Human Genetics, 87*(1), 139-145.
Loehlin, J.C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 44*(10), 1285 -1292.
Loehlin, J.C., & Martin, N. G. (2001). Age changes in personality traits and their heritabilities during the adult years: Evidence from Australian twin registry samples. *Personality and Individual Differences, 30*(7), 1147-1160.
Loos, R.J.F., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., . . . Beckmann, J.S. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics, 40*(6), 768-775.
Luciano, M., Posthuma, D., Wright, M. J., de Geus, E. J. C., Smith, G. A., Geffen, G. M., . . . Martin, N. G. (2005). Perceptual speed does not cause intelligence, and intelligence does not cause perceptual speed. *Biological Psychology, 70*(1), 1-8.
Lunter, G., & Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research, 21*(6), 936-939.
Luo, D., Petrill, S.A., & Thompson, L.A. (1994). An exploration of genetic g: Hierarchical factor analysis of cognitive data from the Western Reserve Twin Project. *Intelligence, 18*(3), 335-347.
Lykken, D. T., McGue, M., Bouchard, T. J., & Tellegen, A. (1990). Does contact lead to similarity or similarity to contact. *Behavior Genetics, 20*(5), 547-561.
Lykken, D. T., McGue, M., & Tellegen, A. (1987). Recruitment bias in twin research - the rule of 2/3 reconsidered. *Behavior Genetics, 17*(4), 343-362.
MacLeod, A.K., Davies, G., Payton, A., Tenesa, A., Harris, S.E., Liewald, D., . . . Gow, A.J. (2012). Genetic copy number variation and general cognitive ability. *PLoS ONE, 7*(12), e37385.
Madsen, B.E., & Browning, S.R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics, 5*(2), e1000384.
Maher, B. (2008). The case of the missing heritability. *Nature, 456*(7218), 18-21.
Manolio, T.A., & Collins, F.S. (2009). The HapMap and genome-wide association studies in diagnosis and therapy. *Annual Review of Medicine, 60*, 443-156.
Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., . . . Chakravarti, A. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753.
Marioni, R.E., Davies, G., Hayward, C., Liewald, D., Kerr, S.M., Campbell, A., . . . Hocking, L.J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence, 44*, 26-32.
Marioni, R.E., Penke, L., Davies, G., Huffman, J.E., Hayward, C., & Deary, I.J. (2014). The total burden of rare, non-synonymous exome genetic variants is not associated with childhood or late-life cognitive ability. *Proceedings of the Royal Society B: Biological Sciences, 281*(1781), 20140117.
Markus, K.A., & Borsboom, D. (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. London: Routledge.

Martin, N. G., Boomsma, D.I., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics, 17*(4), 387-392.
Martin, N. G., & Eaves, L.J. (1977). Genetic analysis of covariance structure. *Heredity, 38*(Feb), 79-95.
Martin, N. G., & Wilson, S. R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behavior Genetics, 12*(4), 467-472.
Mather, K ., & Jinks, J . L. . (1971). *Biometrical Genetics* (2nd ed.). New York: Cornell University Press.
Matzke, D., Dolan, C.V., & Molenaar, D. (2010). The issue of power in the identification of "g" with lower-order factors. *Intelligence, 38*(3), 336-344.
McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*(1), 163-200.
McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics, 20*(5), 569-608.
McArdle, J. J., Prescott, C. A., Hamagami, F., & Horn, J. L. (1998). A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology, 14*(1), 69-114.
McCarthy, M. I., Abecasis, G. R. , Cardon, L. R. , Goldstein, D. B. , Little, J. , Ioannidis, J. P. A. , & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics, 9*(5), 356-369.
McCrae, R.R., & Costa, Paul T. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience. *Personality and Individual Differences, 4*(3), 245-255.
McCrae, R.R., & Costa, Paul T. (2003). *Personality in adulthood: A five-factor theory perspective*. New York: Guilford Press.
McCrae, R.R., & Costa, Paul T. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. J. Boyle, G. Matthews & D. H. Saklofsk (Eds.), *Sage handbook of personality theory and assessment* (Vol. 1, pp. 273-294). London: Sage.
McCrae, R.R., & John, O. P. (1992). An introduction to the 5-factor model and its applications. *Journal of Personality, 60*(2), 175-215.
McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H., & Paurnonen, S.V. (1996). Evaluating replicebility of factors in the Revised NEO Personality inventory: confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology, 70*, 552-566.
McGue, Matt, Bouchard Jr, Thomas J, Iacono, William G, & Lykken, David T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture & psychology* (pp. 59-76). Washington, DC: American Psychological Association.
Mcrae, A.F., Matigian, N.A., Vadlamudi, L., Mulley, J.C., Mowry, B., Martin, N.G., . . . Visscher, P.M. (2007). Replicated effects of sex and genotype on gene expression in human lymphoblastoid cell lines. *Human molecular genetics, 16*(4), 364-373.
McRae, A.F., Wright, M.J., Hansell, N.K., Montgomery, G.W., & Martin, N.G. (2013). No association between general cognitive ability and rare copy number variation. *Behavior Genetics, 43*(3), 202-207.
Medland, S., & Neale, M.C. (2010). An integrated phenomic approach to multivariate allelic association. *European Journal of Human Genetics, 18*(2), 233-239.
Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*(2), 127-143.
Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29*(3), 223-236.
Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.
Metzker, M.L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics, 11*(1), 31-46.
Middeldorp, C.M., & Boomsma, D.I. (2009). Genetics and Psychopathology. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of Neuroscience for Behavioral Sciences*. New York: Wiley.
Middeldorp, C.M., Cath, D. C., Beem, A. L., Willemsen, G., & Boomsma, D.I. (2008). Life events, anxious depression and personality: a prospective and genetic study. *Psychological Medicine, 38*(11), 1557-1565.

Middeldorp, C.M., Cath, D. C., Van Dyck, R., & Boomsma, D.I. (2005). The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychological Medicine, 35*(5), 611-624.

Mineka, S., Watson, D., & Clark, L.A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology, 49*(1), 377-412.

Minică, C.C., Boomsma, D.I., Van Der Sluis, S., & Dolan, C.V. (2010). Genetic association in multivariate phenotypic data: power in five models. *Twin Research and Human Genetics, 13*(06), 525-543.

Minică, C.C., Dolan, C.V., Hottenga, J.-J., Willemsen, G., Vink, J.M., & Boomsma, D.I. (2013). The Use of Imputed Sibling Genotypes in Sibship-Based Association Analysis: On Modeling Alternatives, Power and Model Misspecification. *Behavior Genetics, 43*(3), 254-266.

Minică, C.C., Dolan, C.V., Kampert, M.M.D., Boomsma, D.I., & Vink, J.M. (2014). Sandwich corrected standard errors in family-based genomewide association studies. *European Journal of Human Genetics*.

Molenaar, D., Dolan, C.V., Wicherts, J.M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence, 38*(6), 611-624.

Molenaar, D., van der Sluis, S., Boomsma, D.I., & Dolan, C.V. (2012). Detecting specific genotype by environment interactions using marginal maximum likelihood estimation in the classical twin design. *Behavior Genetics, 42*(3), 483-499.

Mosing, M.A., Pedersen, N.L., Martin, N. G., & Wright, M.J. (2010). Sex differences in the genetic architecture of optimism and health and their interrelation: a study of Australian and Swedish twins. *Twin Res Hum Genet, 13*(4), 322-329.

Mroczek, D. K. (1992). *Personality and psychopathology in older men: The five factor model and the MMPI-2*. Unpublished doctoral dissertation. Boston University. Boston.

Mulaik, S. A. (1987). A brief-history of the philosophical foundations of exploratory factor-analysis. *Multivariate Behavioral Research, 22*(3), 267-305.

Muthén, B.O. , Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics, 9*(3), 313-324.

Muthén, L.K., & Muthén, B.O. . (1998-2007). *Mplus User's Guide* (5th ed.). Los Angeles: Muthén & Muthén.

Muthén, L.K., & Muthén, B.O. . (2007). MPlus (5.1) [Computer software]. Los Angeles: Muthén & Muthén.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., . . . Jamali, P. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature, 478*(7367), 57-63.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., . . . Ropers, H.H. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature, 478*(7367), 57-63. doi: 10.1038/nature10423

Najmabadi, H., Motazacker, M.M., Garshasbi, M., Kahrizi, K., Tzschach, A., Chen, W., . . . Abedini, S.S. (2007). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Human Genetics, 121*(1), 43-48.

Nance, W. E., Kramer, A. A., Corey, L. A., Winter, P. M., & Eaves, L.J. (1983). A causal-analysis of birth-weight in the offspring of monozygotic twins. *American Journal of Human Genetics, 35*(6), 1211-1223.

Narusyte, J., Neiderhiser, J. M., D'Onofrio, B. M., Reiss, D., Spotts, E. L., Ganiban, J., & Lichtenstein, P. (2008). Testing Different Types of Genotype-Environment Correlation: An Extended Children-of-Twins Model. *Developmental Psychology, 44*(6), 1591-1603.

Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., O.-M., M., . . . Daly, M.J. (2011). Testing for an unusual distribution of rare variants. *PLoS genetics, 7*(3), e1001322.

Neale, M.C. (2000). MxGui (1.7.03) [Computer software]. Richmond, VA: Virginia Commonwealth University. Retrieved from http://www.vcu.edu/mx/

Neale, M.C. (2007). Mx Examples. from http://www.vcu.edu/mx/examples.html

Neale, M.C., & Cardon, L. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers B.V.

Neale, M.C., Eaves, L.J., Kendler, K. S., & Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives - the effects of soft selection. *Behavior Genetics, 19*(2), 163-169.
Neale, M.C., Lubke, G., Aggen, S. H., & Dolan, C.V. (2005). Problems with using sum scores for estimating variance components: Contamination and measurement noninvariance. *Twin Research and Human Genetics, 8*(6), 553-568.
Neale, M.C., & McArdle, J. J. (2000). Structured latent growth curves for twin data. *Twin Res, 3*(3), 165-177.
Need, A.C., Attix, D.K., McEvoy, J.M., Cirulli, E.T., Linney, K.L., Hunt, P., . . . Shianna, K.V. (2009). A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Human molecular genetics, 18*(23), 4650-4661.
Neisser, U., Boodoo, G., Bouchard T.J. Jr, Boykin, A.W., Brody, N., Ceci, S.J., . . . Sternberg, R.J. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77.
Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., . . . Mefford, H.C. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics, 42*(9), 790-793.
Nigg, J.T., John, O.P., Blaskey, L.G., Huang-Pollock, C.L., Willicut, E.G., Hinshaw, S.P., & Pennington, B. (2002). Big Five dimensions and ADHD symptoms: Links between personality traits and clinical symptoms. *Journal of Personality and Social Psychology, 83*(2), 451-469.
Nivard, M.G., Mbarek, H., Hottenga, J.J., Smit, J.H., Jansen, R., Penninx, B.W., . . . Boomsma, D.I. (2013). Further confirmation of the association between anxiety and CTNND2: replication in humans. *Genes, Brain and Behavior*.
Norman, W.T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The journal of abnormal and social psychology, 66*(6), 574-583.
Norman, W.T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population*. University of Michigan, Department of Psychological Sciences. Ann Arbor.
Parker, J.D.A., Bagby, R.M., & Summerfeldt, L.J. (1993). Confirmatory factor analysis of the Revised NEO Personality Inventory. *Personality and Individual Differences, 15*(4), 463-466.
Peabody, D., & Goldberg, L.R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57*(3), 552-567.
Pearl, J. (2000). *Causality*. New York: Oxford University Press.
Pearson, H. (2006). What is a gene? *Nature, 441*(7092), 398-401.
Pedersen, N. L., & Reynolds, C. A. (1998). Stability and change in adult personality: Genetic and environmental components. *European Journal of Personality, 12*(5), 365-386.
Pervin, L.A. (1994). Further reflections on current trait theory. *Psychological Inquiry, 5*(2), 169-178.
Petrill, S. A., Lipton, P. A., Hewitt, J. K., & Plomin, R. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology, 40*(5), 805-812.
Petrill, S.A., Lipton, P.A., Hewitt, J.K., Plomin, R., Cherny, S.S., Corley, R., & DeFries, J.C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental psychology, 40*(5), 805.
Pietiläinen, K.H., Naukkarinen, J., Rissanen, A., Saharinen, J., Ellonen, P., Keränen, H., . . . Yki-Järvinen, H. (2008). Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS medicine, 5*(3), e51.
Pietiläinen, K.H., Róg, T., Seppänen-Laakso, T., Virtue, S., Gopalacharyulu, P., Tang, J., . . . Ruskeepää, A.-L. (2011). Association of lipidome remodeling in the adipocyte membrane with acquired obesity in humans. *PLoS biology, 9*(6), e1000623.
Pietiläinen, K.H., Sysi-Aho, M., Rissanen, A., Seppänen-Laakso, T., Yki-Järvinen, H., Kaprio, J., & Orešič, M. (2007). Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects–a monozygotic twin study. *PLoS ONE, 2*(2), e218.
Plomin, R. (1986). *Development, genetics, and psychology*. Hillsdale: Erlbaum.
Plomin, R., & Caspi, A. (1990). Behavioral genetics and personality. *Handbook of personality: Theory and research, 2*, 251-276.
Plomin, R., Defries, J. C., & Loehlin, John C. (1977). Genotype-environment interaction and correlation in analysis of human-behavior. *Psychological Bulletin, 84*(2), 309-322.

Plomin, R., Defries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral Genetics* (5th ed.). New York: W.H.Freeman & Co Ltd.

Plomin, R., Haworth, C.M.A., Meaburn, E.L., Price, T.S., & Davis, O.S.P. (2013). Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science, 24*(4), 562-568.

Plomin, R., & Spinath, F.M. (2004). Intelligence: genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*(1), 112-129.

Polderman, T.J.C., Gosso, M.F., Posthuma, D., van Beijsterveldt, T.C., Heutink, P., Verhulst, F.C., & Boomsma, D.I. (2006). A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *Acta neurol. belg, 106*, 191-207.

Posthuma, D., & Boomsma, D.I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics, 30*(2), 147-158.

Posthuma, D., & Boomsma, D.I. (2005). Mx scripts library: Structural equation modeling scripts for twin and family data. *Behavior Genetics, 35*(4), 499-505.

Prescott, C. A., Aggen, S. H., & Kendler, K. S. (1999). Sex differences in the sources of genetic liability to alcohol abuse and dependence in a population-based sample of US twins. *Alcoholism-Clinical and Experimental Research, 23*(7), 1136-1144.

Prescott, C. A., Aggen, S. H., & Kendler, K. S. (2000). Sex-specific genetic influences on the comorbidity of alcoholism and major depression in a population-based sample of US twins. *Archives of General Psychiatry, 57*(8), 803-811.

Price, A.L., Kryukov, G.V., de Bakker, P.I.W., Purcell, S.M., Staples, J., Wei, L.-J., & Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *The American Journal of Human Genetics, 86*(6), 832-838.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904-909.

Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? *The American Journal of Human Genetics, 69*(1), 124-137.

Purcell, S.M. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*(6), 554-571.

Purcell, S.M., Neale, Benjamin, Todd-Brown, Kathe, Thomas, Lori, Ferreira, Manuel AR, Bender, David, . . . Daly, Mark J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Purcell, S.M., Cherny, S.S., & Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics, 19*(1), 149-150.

Purcell, S.M., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., . . . Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Ramsden, S., Richardson, F.M., Josse, G., Thomas, M.S.C., Ellis, C., Shakeshaft, C., . . . Price, C.J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature, 479*(7371), 113-116.

Rao, D. C., Morton, N. E., & Yee, S. (1974). Analysis of family resemblance II. A linear-model for familial correlation. *American Journal of Human Genetics, 26*(3), 331-359.

Rapee, R.M., Barrett, P. M., Dadds, M. R., & Evans, L. (1994). Reliability of the DSM-III-R childhood anxiety disorders using structured interview - interrater and parent-child agreement *Journal of the American Academy of Child and Adolescent Psychiatry, 33*(7), 984-992.

Rapee, R.M., Schniering, C.A., & Hudson, J.L. (2009). Anxiety disorders during childhood and adolescence: origins and treatment. *Annual Review of Clinical Psychology, 5*, 311-341.

Rathouz, P. J., Van Hulle, C. A., Rodgers, J. L., Waldman, I. D., & Lahey, B. B. (2008). Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene-environment correlation. *Behavior Genetics, 38*(3), 301-315.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4, The Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

Raven, J.C. (1960). *Guide to the standard progressive matrices: sets A, B, C, D and E*. London: HK Lewis.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rebollo, I., de Moor, M.H.M., Dolan, C.V., & Boomsma, D.I. (2006). Phenotypic factor analysis of family data: Correction of the bias due to dependency. *Twin Research and Human Genetics, 9*(3), 367-376.
Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., . . . Chen, W. (2006). Global variation in copy number in the human genome. *Nature, 444*(7118), 444-454.
Rettew, D. C., Vink, J. M., Willemsen, G., Doyle, A., Hudziak, J. J., & Boomsma, D.I. (2006). The genetic architecture of neuroticism in 3301 Dutch adolescent twins as a function of age and sex: A study from the Dutch Twin Register. *Twin Research and Human Genetics, 9*(1), 24-29.
Reynolds, C. A., Finkel, D., Gatz, M., & Pedersen, N. L. (2002). Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models. *Experimental Aging Research, 28*(4), 407-433.
Reynolds, C. A., Finkel, D., McArdle, J. J., Gatz, M., Berg, S., & Pedersen, N. L. (2005). Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Developmental Psychology, 41*(1), 3-16.
Rice, F., Harold, G. T., & Thapar, A. (2002a). Assessing the effects of age, sex and shared environment on the genetic aetiology of depression in childhood and adolescence. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*(8), 1039-1051.
Rice, F., Harold, G. T., & Thapar, A. (2002b). The genetic aetiology of childhood depression: a review. *Journal of Child Psychology and Psychiatry, 43*(1), 65-79.
Rietveld, M.J.H., Dolan, C.V., Van Baal, G.C.M., & Boomsma, D.I. (2003). A twin study of differentiation of cognitive abilities in childhood. *Behavior Genetics, 33*(4), 367-381.
Rijsdijk, F. V., Vernon, P. A., & Boomsma, D.I. (2002). Application of hierarchical genetic models to Raven and WAIS subtests: A Dutch twin study. *Behavior Genetics, 32*(3), 199-210.
Rimmer, A., Phan, H., Mathieson, I., Lunter, G., & McVean, G. (2013). Platypus: A Haplotype-Based Variant Caller For Next Generation Sequence Data. from http://www.well.ox.ac.uk/platypus
Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L.J., Hoh, J., . . . Merikangas, K. R. (2009). Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression A Meta-analysis. *Jama-Journal of the American Medical Association, 301*(23), 2462-2471.
Ropers, H.H. (2008). Genetics of intellectual disability. *Current opinion in genetics & development, 18*(3), 241-250.
Ropers, H.H. (2010a). Genetics of early onset cognitive impairment. *Annual review of genomics and human genetics, 11*, 161-187.
Ropers, H.H. (2010b). Single gene disorders come into focus-again. *Dialogues in clinical neuroscience, 12*(1), 95.
Roysamb, E., Harris, J. R., Magnus, P., Vitterso, J., & Tambs, K. (2002). Subjective well-being. Sex-specific effects of genetic and environmental factors. *Personality and Individual Differences, 32*(2), 211-223.
Rummel, R. J. (1970). *Applied factor analysis*. Philadelphia: Inst Sci Inform Inc.
Rusk, N. , & Kiermer, V. (2008). Primer: Sequencing—the next generation. *Nature Methods, 5*(1), 15.
Sandhu, M.S., Weedon, M.N., Fawcett, K.A., Wasson, J., Debenham, S.L., Daly, A,, . . . Sherva, R. (2007). Common variants in WFS1 confer risk of type 2 diabetes. *Nature Genetics, 39*(8), 951-953.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*(12), 5463-5467.
Sarachana, T., Zhou, R., Chen, G., Manji, H.K., & Hu, V.W. (2010). Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. *Genome Med, 2*(4), 23.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514.
Sattler, J. M. (1992). *Assessment of children: WISC–III and WPPSI–R Supplemental*. San Diego, CA: Jerome M. Sattler.
Scheet, P., Ehli, E.A., Xiao, X., van Beijsterveldt, C.E.M., Abdellaoui, A., Althoff, R.R., . . . Huizenga, P.E. (2012). Twins, Tissue, and Time: An Assessment of SNPs and CNVs. *Twin Research and Human Genetics, 15*(6), 737-745.
Schmidt, F.L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162-173.

Schmit, M.J., & Ryan, A.M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*(6), 966-974.
Schousboe, K., Willemsen, G., Kyvik, K. O., Mortensen, J., Boomsma, D.I., Cornes, B. K., . . . Harris, J. R. (2003). Sex differences in heritability of BMI: A comparative study of results from twin studies in eight countries. *Twin Research, 6*(5), 409-421.
Shendure, J., & Ji, H.. (2008). Next-generation DNA sequencing. *Nature biotechnology, 26*(10), 1135-1145.
Silberg, J. L., Rutter, M., & Eaves, L.J. (2001). Genetic and environmental influences on the temporal association between earlier anxiety and later depression in girls. *Biological psychiatry, 49*(12), 1040-1049.
Silberg, J. L., Rutter, M., Neale, M.C., & Eaves, L.J. (2001). Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *British Journal of Psychiatry, 179*, 116-121.
Silventoinen, K., Posthuma, D., Van Beijsterveldt, T., Bartels, M., & Boomsma, D.I. (2006). Genetic contributions to the association between height and intelligence: Evidence from Dutch twin data from childhood to middle age. *Genes, Brain and Behavior, 5*(8), 585-595.
Slagboom, P. E., & Meulenbelt, I. (2002). Organisation of the human genome and our tools for identifying disease genes. *Biological Psychology, 61*(1-2), 11-31.
Snieder, H., van Doornen, L. J. P., & Boomsma, D.I. (1997). The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. *American Journal of Human Genetics, 60*(3), 638-650.
Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201-292.
Spearman, C. (1927). *The abilities of man*. London: Macmillan.
Stinissen, J., Willems, P.J., Coetsier, P., & Hulsman, W.L.L. (1970). *Manual for the Dutch translated and adapted version of the Wechsler Adult Intelligence Scale (WAIS)*. Lisse: Swets and Zeitlinger.
Stoel, R. D., De Geus, E. J. C., & Boomsma, D.I. (2006). Genetic analysis of sensation seeking with an extended twin design. *Behavior Genetics, 36*(2), 229-237.
Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin, 125*(4), 392-409.
Stout, W. E. . (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52*, 589-517.
Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426.
Sullivan, P.F., Neale, M.C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry, 157*(10), 1552-1562.
Sundet, J.M., Tambs, K., Magnus, P., & Berg, K. (1988). On the question of secular trends in the heritability of intelligence test scores: A study of Norwegian twins. *Intelligence, 12*(1), 47-59.
Teare, D.M., & Barrett, J.H. (2005). Genetic linkage studies. *The Lancet, 366*(9490), 1036-1044.
Thomsen, S. F., van der Sluis, S., Stensballe, L. G., Posthuma, D., Skytthe, A., Kyvik, K. O., . . . Bisgaard, H. (2009). Exploring the Association between Severe Respiratory Syncytial Virus Infection and Asthma. A Registry-based Twin Study. *American Journal of Respiratory and Critical Care Medicine, 179*(12), 1091-1097.
Topper, S., Ober, C., & Das, S. (2011). Exome sequencing and the genetics of intellectual disability. *Clinical genetics, 80*(2), 117-126.
Truett, K. R., Eaves, L.J., Walters, E. E., Heath, A. C., Hewitt, J. K., Meyer, J. M., . . . Kendler, K. S. (1994). A model system for analysis of family resemblance in extended kinships of twins. *Behavior Genetics, 24*(1), 35-49.
Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence, 42*, 83-88.
Trzaskowski, M., Shakeshaft, N.G., & Plomin, R. (2013). Intelligence indexes generalist genes for cognitive abilities. *Intelligence, 41*(5), 560-565.
Trzaskowski, M., Yang, J., Visscher, P.M., & Plomin, R. (2013). DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12. *Molecular psychiatry*.
Tucker-Drob, E. M. (2009). Differentiation of Cognitive Abilities Across the Life Span. *Developmental Psychology, 45*(4), 1097-1118.

Tupes, E.C., & Christal, R.E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*(2), 225-251.
Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., . . . Affourtit, J.P. (2009). A core gut microbiome in obese and lean twins. *Nature, 457*(7228), 480-484.
Vadlamudi, L., Dibbens, L.M., Lawrence, K.M., Iona, X., McMahon, J.M., Murrell, W., . . . Berkovic, S.F. (2010). Timing of de novo mutagenesis—a twin study of sodium-channel mutations. *New England Journal of Medicine, 363*(14), 1335-1340.
van Beijsterveldt, C.E.M., Groen-Blokhuis, M., Hottenga, J.-J., Franić, S., Hudziak, J.J., Lamb, D., . . . Schutte, N. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16*(1), 252-267.
van den Berg, S. M., Glas, C. A. W., & Boomsma, D.I. (2007). Variance decomposition using an IRT measurement model. *Behavior Genetics, 37*(4), 604-616.
van den Oord, E.J.C.G. (2000). Framework for identifying quantitative trait loci in association studies using structural equation modeling. *Genetic Epidemiology, 18*(4), 341-359.
van den Oord, E.J.C.G., Boomsma, D.I., & Verhulst, F. C. (1994). A study of problem behaviors in 10-year-old to 15-year-old biologically related and unrelated international adoptees. *Behavior Genetics, 24*(3), 193-205.
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*(3), 349-357.
van der Maas, H. L. J., Dolan, C.V., Grasman, Rppp, Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842-861.
van der Maas, H. L. J., Kan, Kees-Jan, & Borsboom, Denny. (2014). Intelligence Is What the Intelligence Test Measures. Seriously. *Journal of Intelligence, 2*(1), 12-15.
van der Sluis, S., Dolan, C.V., Neale, M.C., Boomsma, D.I., & Posthuma, D. (2006). Detecting Genotype–Environment Interaction in Monozygotic Twin Data: Comparing the Jinks and Fulker Test and a New Test Based on Marginal Maximum Likelihood Estimation. *Twin Research and Human Genetics, 9*(3), 377-392.
van der Sluis, S., Dolan, C.V., Neale, M.C., & Posthuma, D. (2008). Power calculations using exact data simulation: A useful tool for genetic study designs. *Behavior Genetics, 38*(2), 202-211.
van der Sluis, S., Posthuma, D., & Dolan, C.V. (2012). A note on false positives and power in G× E modelling of twin data. *Behavior Genetics, 42*(1), 170-186.
van der Sluis, S., Posthuma, D., & Dolan, C.V. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS genetics, 9*(1), e1003235.
van der Sluis, S., Verhage, M., Posthuma, D., & Dolan, C.V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS ONE, 5*(11), e13929.
Van der Valk, J.C., Stroet, T.M., & Boomsma, D.I. (1998). Quantitative genetic analysis of Internalizing and Externalizing Problems in a large sample of 3-year-old twins. *Twin Research, 1*, 25-33.
van Dongen, J.P., Draisma, H.H.M., Martin, N. G., & Boomsma, D.I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics, 13*(9), 640-653.
Van Dongen, S., & Gangestad, S.W. (2011). Human fluctuating asymmetry in relation to health and quality: a meta-analysis. *Evolution and Human behavior, 32*(6), 380-398.
Van Grootheest, D. S., Bartels, M., Van Beijsterveldt, C. E. M., Cath, D. C., Beekman, A. T., Hudziak, J. J., & Boomsma, D.I. (2008). Genetic and environmental contributions to self-report obsessive-compulsive symptoms in dutch adolescents at ages 12, 14, and 16. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(10), 1182-1188.
Van Haasen, P.P., De Bruyn, E.E.J., Pijl, Y., Poortinga, Y.H., Lutje-Spelberg, H.C., Vander Steene, G., . . . Stinissen, J. (1986). *Wechsler intelligence scale for children-revised, Dutch Version*. Lisse: Swets & Zetlinger BV.
Veltman, J.A., & Brunner, H.G. (2012). De novo mutations in human genetic disease. *Nature Reviews Genetics, 13*(8), 565-575.
Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18 (CBCL/4-18 Manual)*. Rotterdam: Academic Medical Centre Rotterdam / Erasmus University, Sophia Children's Hospital, Department of Child Psychiatry.

Vink, J. M., & Boomsma, D.I. (2002). Gene finding strategies. *Biological Psychology, 61*(1-2), 53-71.
Vinkhuyzen, A. A. E., van der Sluis, S., de Geus, E. J. C., Boomsma, D.I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes Brain and Behavior, 9*(3), 276-287.
Vinkhuyzen, A. A. E., van der Sluis, S., & Posthuma, D. (2010). Life events moderate variation in cognitive ability (g) in adults. *Molecular Psychiatry,* (15), 1197-1197
Visscher, P.M., Brown, M.A., McCarthy, M.I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics, 90*(1), 7-24.
Visscher, P.M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., . . . Boomsma, D.I. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. *The American Journal of Human Genetics, 81*(5), 1104-1110.
Waller, N. G., & Reise, S. P. (1992). Genetic and environmental-influences on item response pattern scalability. *Behavior Genetics, 22*(2), 135-152.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–Third Edition, Dutch version*. Lisse: Swets and Zeitlinger.
Wechsler, D., Kort, W., Compaan, E. L., Bleichrodt, N., Resing, W. C. M., & Schittkatte, M. (2002). *Wechsler Intelligence Scale for Children-Third Edition. Dutch version*. Lisse: Swets and Zeitlinger.
Weiss, L. A., Pan, L., Abney, M., & Ober, C. (2006). The sex-specific genetic architecture of quantitative traits in humans. *Nature Genetics, 38*(2), 218-222.
Widiger, R. A. (2005). A dimensional model of psychopathology. *Psychopathology, 38*(4), 211-214.
Willems, T.F., Gymrek, M., Highnam, G., Mittelman, D., & Erlich, Y. (2014). The Landscape of Human STR Variation. *bioRxiv*. doi: http://dx.doi.org/10.1101/004671
Willemsen, G., Vink, J.M., Abdellaoui, A., den Braber, A., van Beek, J.H.D.A., Draisma, H.H.M., . . . van Lien, R. (2013). The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data Collection. *16*(1), 271-281.
Winckler, W., Weedon, M.N., Graham, R.R., McCarroll, S.A., Purcell, S.M., Almgren, P., . . . Walker, M. (2007). Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes, 56*(3), 685-693.
Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58-79.
Wray, N. R., & Visscher, P. M. (2008). Populaton genetics and its relevance to gene mapping. In B. M. Neale, M. A. R. Ferreira, S. E. Medland & D. Posthuma (Eds.), *Statistical Genetics. Gene Mapping Through Linkage and Association* (pp. 87-110). New York: Taylor & Francis Group.
Yang, Jian, Lee, S Hong, Goddard, Michael E, & Visscher, Peter M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics, 88*(1), 76-82.
Yang, J., Weedon, M.N., Purcell, S.M., Lettre, G., Estrada, K., Willer, C.J., . . . Mangino, M. (2011). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics, 19*(7), 807-812.
Yeo, R.A., Gangestad, S.W., Liu, J., Calhoun, V.D., & Hutchison, K.E. (2011). Rare copy number deletions predict individual variation in intelligence. *PLoS ONE, 6*(1), e16339.
York, T.P., Miles, M.F., Kendler, K.S., Jackson-Cook, C., Bowman, M.L., & Eaves, L.J. (2005). Epistatic and environmental control of genome-wide gene expression. *Twin Research and Human Genetics, 8*(01), 5-15.
Yung, Y., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113-128.
Zuk, O., Hechter, E., Sunyaev, S.R., & Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences, 109*(4), 1193-1198.
Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*, 253-269.
Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 17*, 253-269.
Zwijnenburg, P.J., Meijers-Heijboer, H., & Boomsma, D.I. (2010). Identical but not the same: The value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet, 153B*, 1134–1149.

Nederlandse Samenvatting

Dit proefschrift, getiteld *Van Structural Equation Modeling naar Next-Generation Sequencing: Het Evoluerende Landschap van de Moderne Gedragsgenetica*, gaat over de ontologie van latente psychometrische variabelen en over de genetica van intelligentie. Hieronder geef ik een samenvatting van de voorgaande acht hoofdstukken.

In hoofdstuk 2 wordt de techniek van het *structural equation modeling* geïntroduceerd, zoals deze toegepast wordt in het klassieke tweelingonderzoek. Eerst bespreek ik de basismethode voor het analyseren van overeenkomsten tussen familieleden om conclusies te kunnen trekken over de effecten van niet-geobserveerde genetische factoren en omgevingsfactoren. Daarna komt de implementatie van verschillende modellen in *structural equation modeling* aan bod voor genetisch-informatieve studies, waaronder simplex en latente groeicurvemodellen. Vervolgens worden *structural equation*-modellen besproken die speciaal in de genetica zijn ontwikkeld, waaronder factoranalysemodellen met gemeenschappelijke en onafhankelijke factoren, modellen voor de interactie tussen genotype en omgeving, sekse-limitatiemodellen, en causale modellen waarbinnen wordt getracht de richting van causale effecten te bepalen. Het hoofdstuk eindigt met een discussie over het opnemen van gemeten genetische varianten in associatieanalyses gebaseerd op *structural equation modeling*.

Hoofdstuk 3 gaat over het gebruik van genetisch-informatieve analyses op het niveau van afzonderlijke items om meer te weten te komen over de ontologie van latente gedragsfenotypes (bijvoorbeeld depressie en algemene cognitieve vaardigheden). Hiervoor wordt gekeken naar de mediërende rol die ze spelen met betrekking tot genetische en omgevingsinvloeden. De beschikbaarheid van genetisch-informatieve data op itemniveau maakt het mogelijk om 1) te onderzoeken wat de empirische implicaties zijn van de realistische interpretatie van latente psychometrische eigenschappen, te weten hoe de geobserveerde covariatie gemedieerd wordt door genetische en omgevingsinvloeden, en 2) de dimensionaliteit van de latente covariantiestructuren (zowel genetisch als van de omgeving) te bestuderen die leidt tot de geobserveerde itemcovariatie. Ik merk op dat de typische problemen in het vaststellen van dimensionaliteit in psychometrisch onderzoek een gevolg kunnen zijn van de verschillen in de dimensionaliteit van genetische en omgevingscovariantiestructuren, en stel voor om genetisch-informatieve data op itemniveau te gebruiken als een hulpmiddel om meer inzicht te krijgen in de fenotypische dimensionaliteit.

In hoofdstuk 4 wordt de methodologie uit hoofdstuk 3 gebruikt om de ontologie en de genetische en omgevingsetiologie van de *internalizing syndrome dimensions* van de *Child Behavior Checklist* te onderzoeken (CBCL; Achenbach, 1991; Verhulst, Van der Ende, & Koot, 1996). De resultaten suggereren dat de verschillende syndromen van de CBCL beter begrepen kunnen worden als een samengesteld geheel van vrij te schatten genetische en omgevingsinvloeden dan als causale entiteiten die zouden leiden tot de waargenomen covariatie tussen symptomen. Bovendien geven de resultaten aan dat er een gemeenschappelijke genetische basis is voor angst, depressie en teruggetrokken gedrag, waarbij het onderscheid tussen deze syndromen voornamelijk ontstaat door individu-specifieke omgevingsfactoren. De bevindingen worden besproken in de context van de bekende moeilijkheid een onderscheid te maken tussen verschillende diagnostische categorieën zoals angst en depressie.

In hoofdstuk 5 wordt dezelfde methodologie gebruikt om 1) de houdbaarheid van de realistische interpretatie van de *Big Five* persoonlijkheidsdimensies (McCrae & Costa, 2008)

te onderzoeken, en 2) om de structuur van de genetische en omgevingscovariantiematrices te analyseren die ten grondslag liggen aan de geobserveerde covariatie van de persoonlijkheidsitems van de *NEO Five Factor Inventory* (NEO-FFI; Costa & McCrae, 1992). Het is opmerkelijk dat deze covariantiematrices vergelijkbare (vijf-factor) structuren vertonen, in tegenstelling tot de resultaten voor de CBCL. Echter, de latente persoonlijkheidsdimensies lijken de genetische en omgevingseffecten op de items niet volledig te verklaren, zoals te verwachten zou zijn onder de realistische interpretatie van de *Big Five*. Ik bespreek de implicaties hiervan voor de substantieve interpretatie van de *Big Five*.

Hoofdstuk 6 geeft een overzicht van de methodologie die gebaseerd is op het modelleren van genetische covariantiestructuren voor onderzoek naar angst en depressie bij kinderen, en het geeft een overzicht van de relevante bevindingen. De vragen die aan bod komen gaan verder dan alleen het inschatten van de bijdragen van genetische en omgevingsfactoren aan angst en depressie, wat relatief simpel is. Het blijkt dat er relatief consistente evidentie is voor a) kleine tot verwaarloosbare sekseverschillen in de genetische etiologie van deze aandoeningen, b) een belangrijke rol voor genetische factoren in het verklaren van stabiliteit door de tijd heen, c) een bijdrage van genetische factoren aan de comorbiditeit tussen angst en depressie, d) een mogelijke rol van genotype-omgevingsinteractie, e) een rol voor een correlatie tussen genotype en omgeving, en f) een ondergeschikte, misschien zelfs helemaal geen etiologische rol voor interactie tussen broertjes en zusjes.

De hoofdstukken 7-9 gaan over de genetica van intelligentie. In hoofdstuk 7 worden alle longitudinale data met betrekking tot verbale, non-verbale en algemene intelligentie samen geanalyseerd, zoals verzameld in het Jonge Nederlands Tweelingen Register (Bartels et al., 2007) in 2009. Ik heb een simplex model gebruikt om de genetische en niet-genetische oorzaken van de temporele stabiliteit van de data te analyseren. Gegeven deze uitkomsten bespreek ik vervolgens hoe de al bestaande longitudinale data het beste gebruikt kunnen worden in de context van *gene finding studies*. De hoge mate van stabiliteit van de gesommeerde genetische factoren geeft aan dat dezelfde set genen ten grondslag ligt aan de variatie in intelligentie tijdens de kinderjaren; dit rechtvaardigt het gebruik van een lineaire combinatie van scores van de verschillende leeftijdsgroepen voor genetisch associatieonderzoek.

De resultaten uit hoofdstuk 7 zijn gebruikt voor het modelleren van het fenotype in de associatiestudies van hoofdstuk 8 en 9. In hoofdstuk 8 wordt een onderzoek besproken waarin de associatie wordt getest tussen normale intelligentie en veelvoorkomende *single-nucleotide polymorphisms* (SNP's) in 43 genen die betrokken zijn bij intellectuele stoornissen. Het uitgangspunt was redelijk eenvoudig, namelijk dat de genetische variatie die van invloed is op continue, polygenetische eigenschappen (bijvoorbeeld normale intelligentie) op dezelfde plekken van het genoom kan liggen die ook verantwoordelijk zijn voor vergelijkbare monogenetische fenotypes (bijvoorbeeld intellectuele stoornissen). Hoewel er voor de afzonderlijke SNP's geen significanties werden gevonden, wezen de analyses met SNP's wel op een verrijking van de set van kandidaatgenen voor polymorfismes met betrekking tot intelligentie. Dit is het eerste onderzoek dat laat zien dat genen die betrokken zijn bij monogenetische aandoeningen met betrekking tot intellectueel functioneren ook normale intelligentie beïnvloeden.

In hoofdstuk 9 wordt het werk uit hoofdstuk 8 uitgebreid tot 168 genen die betrokken zijn bij intellectuele stoornissen, nu ook met gebruik van *next-generation exon sequencing*. Het gaat daarbij om de mogelijke effecten van zeldzame genetische variatie. In overeenstemming met de literatuur werd geen verrijking gevonden van de set van

kandidaatgenen voor mutaties die betrokken zijn bij normale intelligentie. Dit resultaat wordt besproken in de context van de literatuur.

Publication List

Journal articles and book chapters

Franić, S., Dolan, C.V., Borsboom, D., & Boomsma, D.I. (2012). Structural Equation Modeling in Genetics. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 617-635). New York: Guilford Press.

Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2013) Can Genetics Help Psychometrics? Improving Dimensionality Assessment Through Genetic Factor Modeling. *Psychological methods, 18(3)*, 406-433.

Franić, S., Dolan, C. V., Borsboom, D., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2014) Three-and-a-Half-Factor Model? The Genetic and Environmental Structure of the CBCL/6–18 Internalizing Grouping. *Behavior Genetics, 44*(3), 254-268.

Franić, S., Borsboom, D., Dolan, C. V., & Boomsma, D. I. (2013) The Big Five Personality Traits: Psychological Entities or Statistical Constructs? *Behavior Genetics*, advance online publication, doi: 10.1007/s10519-013-9625-7, 1-14.

Franić, S., Middeldorp, C. M., Dolan, C. V., Ligthart, L., & Boomsma, D. I. (2010) Childhood and Adolescent Anxiety and Depression: Beyond Heritability. *Journal of the American Academy of Child and Adolescent Psychiatry, 49(8)*, 820-829.

Franić, S., Dolan, C. V., van Beijsterveldt, C. E. M., Hulshoff Pol, H. E., Bartels, M., & Boomsma, D. I. (2014). Genetic and Environmental Stability of Intelligence in Childhood and Adolescence. *Twin Research and Human Genetics, 17*(03), 151-163.

Franić, S., Groen-Blokhuis, M.M., Dolan, C.V., Kattenberg, M.V., Xiao, X., Scheet, P.A., Ehli, E.A., Davies, G.E., van der Sluis, S., Abdellaoui, A., Hansell, N.K., Martin, N.G., Hudziak, J.J., van Beijsterveldt, C.E.M., Swagerman, S., Hulshoff Pol, H.E., de Geus, E.J.C., Bartels, M., Ropers, H.H., Hottenga, J.J., & Boomsma, D.I. (2014) IQ: Shared Genetic Basis between Mendelian Disorders and a Polygenic Trait. *Under review.*

Franić, S., Dolan, C.V., Broxholme, J., Hu, H., Zemojtel, T., Davies, G.E., Nelson, K., Ehli, E.A., the Childhood Intelligence Consortium, Ropers, H.-H., & Boomsma, D.I. (2014) Mendelian and polygenic inheritance of intelligence: a common set of causal genes? Using Next-Generation Sequencing to examine the effects of 168 cognitive disability genes on normal-range intelligence. *Under review.*

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand Scholten, A., & **Franić, S**. (2009) The end of construct validity. In R. V. Lissitz (Ed.), *The Concept of Validity* (pp. 135-166). Charlottle, NC: Information Age Publishing.

Groen-Blokhuis, M. M., **Franić, S.**, Van Beijsterveldt, C. E. M., de Geus, E., Davies, G.E., Ehli, E.A., Xiao, X., Scheet, P.A., Althoff, R., Hudziak, J.J., Middeldorp, C. M., Boomsma, D.I. (2013) A prospective study of the effects of breastfeeding and FADS2 polymorphisms on cognition and hyperactivity:attention problems. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 457-465.

Benyamin, B., St Pourcain, B., Davis, O.S.P., Davies, G., Hansell, N.K., Brion, M.J.A, Kirkpatrick, R.M., Cents, R.A.M., **Franić, S.**, Miller, M.B., Haworth, C.M.A., Meaburn, E., Price, T.S., Evans, D.E., Timpson, N., Kemp, J., Ring, S., McArdle, W., Medland, S.E., Yang, J., Harris, S.E., Liewald, D.C., Scheet, P., Xiao, X., Hudziak, J.J., de Geus, E.J.C, Wellcome Trust Case Control Consortium 2 (WTCCC2), Jaddoe, V.W.V., Starr, J.M., Verhulst, F.C., Pennell, C., Tiemeier, H., Iacono, W.G., Palmer, L.J., Montgomery, G.W., Martin, N.J., Boomsma, D.I., Posthuma, D., McGue, M.,

Wright, M.J., Smith, G.D., Deary, I.J., Plomin, R. & Visscher, P.M. (2013) Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry, 16(10),* 996-1005.

Van Beijsterveldt, C.E.M., Groen-Blokhuis, M.M., Hottenga, J.J., **Franić, S.**, Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J.C., Bartels, M., & Boomsma, D.I. (2013) The Young Netherlands Twin Registry (YNTR)- Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16(1),* 252-267.

Hansell, N.K., **Franić, S.**, Boomsma, D.I., Martin, N.G., et al. (2013). Complex Relational Processing: Genetic Relationship with Reasoning, Working Memory, and Full-scale IQ and the Search for Loci Contributing to Covariation. *Submitted for publication*

Nivard, M., Van der Sluis, S., **Franić, S.**, Middeldorp, C., Visser, I., Boomsma, D.I., Cramer, A.O.J., & Borsboom, D. (2013). Phenotypic complexity as genetic dark matter: A network explanation of missing heritability. *Manuscript in preparation*

Published abstracts

Franić S., Borsboom, D., Dolan, C. V., & Boomsma, D. I. (2013) Big Five Personality Traits: Psychological Entities or Statistical Constructs? In: Hewitt, J. (ed) Behavior Genetics Association 43rd Annual Meeting Abstracts, Marseille, France, pp. 517

Franić, S., Dolan, C.V., Groen-Blokhuis, M., Hottenga, J.J., van Beijsterveldt,C.E.M., Sullivan, P., Smit, D., Scheet, P., Hudziak, J., de Geus, E., Boomsma, D.I. (2012) Shared genetic bases between Mendelian disorders and a polygenic trait? Testing for an association between genes underlying autosomal recessive cognitive disorders and full-scale IQ. In: Hewitt, J. (ed) Behavior Genetics Association 42nd Annual Meeting Abstracts, Edinburgh, Scotland, pp. 933

Nivard, M., van der Sluis, S., **Franić, S.**, Cramer, A.O.J., Visser, I., Middeldorp, C., Lubke, G., Boomsma, D.I., Borsboom, D. (2012) Phenotypic complexity as genetic dark matter: A network explanation of missing heritability. In: Hewitt, J. (ed) Behavior Genetics Association 42nd Annual Meeting Abstracts, Edinburgh, Scotland, pp. 933

Franić, S., Dolan, C. V., Borsboom, D., van Beijsterveldt, C. E., & Boomsma, D. I. (2010). Multivariate genetic analysis of longitudinally measured cognitive abilities. In: Hewitt, J. (ed) Behavior Genetics Association 40th Annual Meeting Abstracts, Seoul, South Korea, pp. 792

Franić, S., Dolan, C.V., Borsboom, D., Van Beijsterveldt, C.E.M., Hudizak, J., & Boomsma, D.I. (2009). How genetics can help psychometrics: Determining the dimensionality of the Internalizing grouping of the Dutch Version of the Child Behavioral Checklist. In: Hewitt, J. (ed) Behavior Genetics Association 39th Annual Meeting Abstracts, Minneapolis, USA, pp. 652

Acknowledgements

This work was supported by Nuffic (HSP Huygens Scholarships HSP-HP.08/ 298; HSP-HP.10/ 151), the Ministry of Science, Education and Sports of the Republic of Croatia, the Dutch Organization for Scientific Research (NWO 668.772; NWO 433-09-220; NWO 051.02.060, NWO-MagW 480-04-004; NWO/SPI 56-464-14192), the European Research Council (ERC-230374), the National Institute of Mental Health (Grand Opportunity grants 1RC2MH089951-01 and 1RC2 MH089995-01), the National Institutes of Health (NIH, R01D0042157-01A), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), and the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). I would like to thank all twins and family members participating in NTR research.