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 $\alpha 1$ 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.