Genetic and Environmental Influences on Personality: A Study of Twins Reared Together Using the Self- and Peer Report NEO-FFI Scales

Rainer Riemann  Alois Angleitner
University of Bielefeld, Germany

Jan Strelau
University of Warsaw, Poland

ABSTRACT  Previous behavior-genetic research on personality has been almost exclusively based on self-report questionnaire measures. The purpose of this research was to measure personality constructs via self- and peer reports on the items of the NEO Five-Factor Inventory (Costa & McCrae, 1989). The sample included 660 monozygotic and 200 same sex and 104 opposite sex dizygotic twin pairs. We collected self- and two independent peer reports for each of the twins. Our analysis of self-report data replicates earlier findings of a substantial genetic influence on the Big Five ($h^2 = .42$ to $.56$). We also found this influence for peer reports. Our results validate findings based solely on self-reports. However, estimates of genetic contributions to phenotypic variance were substantially higher when based on peer reports ($h^2 = .57$ to $.81$) or self- and peer reports ($h^2 = .66$ to $.79$) because these data allowed...
us to separate error variance from variance due to nonshared environmental influences. Correlations between self- and peer reports reflected the same genetic influences to a much higher extent than identical environmental effects.

Human behavior-genetic research has provided important insights into the etiology of individual differences in personality traits (Loehlin, 1992; Plomin & McClearn, 1993; Rose, 1995; Rowe, 1994). This research has focused most strongly on Extraversion and Neuroticism as the two major dimensions of personality. In recent years, however, the five-factor model has gained a prominent position within personality research (for reviews see Digman, 1990; John, 1990; John, Angleitner, & Ostendorf, 1988; McCrae, 1989). Numerous studies have shown that five broad factors—Extraversion or Surgency, Agreeableness, Conscientiousness, Emotional Stability or Neuroticism, and Culture/Intellect or Openness to Experience—account for much of the common variance in personality rating and questionnaire data. They are meaningful dimensions for the description of individual differences in personality and robustly found in various sources of data. The NEO Personality Inventory-Revised (NEO-PI-R; Costa & McCrae, 1992) and the short version of the NEO-PI, the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1989) are now standard questionnaire measures of the five-factor model.

Four major research designs that are used in human behavior-genetic studies can be distinguished: (a) studies of twins reared together in the same family, (b) studies of twins who have been reared in different families, (c) studies of adoptive families (ideally including measures of the adoptees' biological parents), and (d) family studies analyzing the data of participants who are genetically related to varying degrees. All of these designs have their particular strengths and weaknesses, and firm conclusions regarding the etiology of individual differences are based on an integrative analysis of data based on all four designs (e.g., Bouchard & McGue, 1981; Loehlin, 1992). The most popular design, the comparison of monozygotic (MZ) and dizygotic (DZ) twins reared together, requires several assumptions to estimate genetic and environmental effects on traits: (a) There are only two types of twins (MZ and DZ) and DZ twins share on average 50% of their genes (i.e., there are no effects of assortative mating), (b) twins

1. There is some evidence of assortative mating for Openness and Conscientiousness (see McCrae, 1996). Assortative mating increases the correlation of additive and
are a representative sample of the population to which one wishes to generalize, (c) the effects of environmental influences shared by twins is no greater for MZ than for DZ twins ("equal environments" assumption), and (d) gene-environment correlation and interaction have a negligible effect on the trait under study. Finally, studies of twins reared together do not allow simultaneous estimation of environmental effects shared by the twins and nonadditive genetic influences (e.g., due to genetic dominance). The validity of these assumptions is thoroughly discussed in the behavior-genetic literature (see, e.g., Eaves, Eysenck, & Martin, 1989; Loehlin, 1992). From this discussion it can be concluded that studies of twins reared together provide a valuable tool for behavior-genetic research. The strength of this design is that large samples are available representing the whole range of genotypes and environments.

Combining data from twin, adoption, and family studies into an extremely large data set, Loehlin (1992) thoroughly reviewed the then available behavior-genetic research on the five-factor model of personality. All of the studies summarized by Loehlin used self-report questionnaires that, with one exception (Bergeman et al., 1993), had not been constructed with the goal of measuring the Big Five. Instead, Extraversion and Neuroticism were frequently measured by the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1964), the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975), the EPQ-Revised (Eysenck & Eysenck, 1991), or short versions of one of these questionnaires. Measures related to Agreeableness comprised scales developed to measure masculinity versus femininity, aggression, altruism, empathy, and related constructs. Conscientiousness was indexed by measures of, for example, socialization, task orientation, and conformity. Finally, Culture/Openness was assessed, among others, by scales for flexibility, artistic interests, science interests, and achievement via independence.

Two behavior-genetic models could reasonably be fitted to this extensive data collection. One of these models involved effects of additive genes (A), genetic epistasis (I), shared twin or sibling environments (C), nonadditive genetic effects for DZ twins. Thus, the failure to take assortative mating into account results in an overestimation of effects due to environmental influences shared by siblings and an underestimation of nonadditive genetic effects. Because we neither have appropriate estimates of assortative mating for our sample nor a thorough understanding of its etiology (see Neale & Cardon, 1992), we did not correct the correlations of genetic effects for DZ twins in our models.
and a combination of specific (not shared by siblings) environmental, genotype-environment interaction, and measurement error (E). The other model included effects of A, C, and E, and of a special MZ twin environment (contributing to the similarity of MZ twins reared together). Given the evidence on the “equal environments” assumption, we focus on the data for the first model here. Several conclusions can be drawn from Loehlin’s (1992) meta-analysis. For all five factors, additive genetic effects were the strongest source of the phenotypic variance, explaining 43% of the variance in Culture/Openness, 32% of Extraversion, 27% of Neuroticism, 24% of Agreeableness, and 22% of Conscientiousness. Conversely, shared sibling environment effects contributed little to phenotypic variance. They were negligible for Extraversion (2%) and small for Culture/Openness (6%), Conscientiousness (7%), Neuroticism (7%), and Agreeableness (11%). Nonadditive genetic effects had a noticeable influence on all dimensions (11% to 17%) except Culture/Openness (2%).

Highly similar results are reported by Bouchard (1994) for the Minnesota study of twins reared apart and twins reared together using scales or factors of the Multidimensional Personality Questionnaire (MPQ; Tellegen et al., 1988) to measure the five factors, although for these data a model involving A and E effects provide the most appropriate description of the data. Earlier studies of twins reared together, however, yielded heritability estimates that were somewhat higher and indicated less influence of nonadditive genetic factors.

Two studies of twins reared together (Beer & Rose, 1995; Jang, Livesley, & Vernon, 1996) that used direct measures of the five-factor model, the NEO-PI and the NEO-FFI (Costa & McCrae, 1985) yielded results similar to those reported by Loehlin (1992). The estimates of genetic effects on the phenotypic variance of the NEO-PI scales reported by Jang et al. (1996) were 61% for Openness, 53% for Extraversion, 44% for Conscientiousness, 41% for Neuroticism, and 41% for Agreeableness. However, the results of Bergeman et al.’s (1993) study of twins reared together and twins reared apart were markedly different. Bergeman et al. used a Swedish translation of short preliminary NEO-PI scales for Openness (25 items), Conscientiousness (10 items), and Agreeableness (10 items). They found small genetic effects for Conscientiousness (29%) and Agreeableness (12%). Shared environmental influences were substantial for Agreeableness (21%) and noticeable for Conscientiousness (11%). The results for Openness were in line with previous research. For Agreeableness and
Conscientiousness, twins reared apart were substantially less similar than twins reared together. This result, however, was not replicated in Bouchard’s (1994) analysis and may, as the authors concede, be attributed to psychometric limitations of the measures or to peculiarities of the Swedish sample.

In general, the data summarized here are in good agreement with many behavior-genetic studies of personality traits. They indicate substantial heritability of all dimensions of the five-factor model. Genetic factors explain about 40% to 50% of the variance in self-reported personality traits; environmental influences shared by siblings or twins are absent or marginal (Bouchard, 1994). Nonadditive genetic effects or, alternatively, a special MZ environment resemblance are important for Neuroticism, Extraversion, and Conscientiousness.

These conclusions for adult personality measures rely almost exclusively on self-report measures, which are used because they allow researchers to collect reliable and valid personality measures for several hundred or even several thousand participants. Since its beginning, however, the systematic study of personality has used multiple methods (e.g., Allport, 1937; Cattell, 1950; Fiske, 1971; Heymans & Wiersma, 1906), primarily because they allow examination of method-specific effects. This is important for personality constructs like traits, which summarize divergent modes of responding. In addition, by including multiple sources of data in behavior-genetic research, the influence of environmental effects that are not shared by twins can be separated from effects due to measurement error.

Although it may be argued that people generally know their own behaviors, thoughts, and feelings better than observers, there are reasons to believe that self-reports of twins reared together are a problematic source of data. Specifically, several authors (e.g., Borkenau, 1993; Carey, 1986; Heath, Neale, Kessler, Eaves, & Kendler, 1992; Loehlin, 1986; Rose, 1995; Saudino & Eaton, 1991) have suggested that self-reports may be distorted by “contrast effects,” which result in an underestimation of the similarity between DZ twins and may obscure real shared environmental influences or inflate the estimates of nonadditive genetic influences.

We can conceive of several processes that would result in a distortion of self-reports provided by twins reared together. For example, real behavioral differences may be emphasized by DZ twins because they compare their own behavior to that of their co-twin instead of the population average. Alternatively, MZ twins may discount behavioral differences because of their generally high similarity.
A second shortcoming of self-report data in behavior-genetic research is that no personality judgment directly reflects behavioral differences (except the answering of questionnaire items). To varying extents responses to questionnaire items or rating scales require inferences and interpretations. Thus, it remains unclear to what extent heritability coefficients and estimates of environmental influences reflect content or stylistic (response-set) variance. This argument is not limited to studies of twins reared together.

In the present study of twins reared together, we use peer reports in addition to self-reports on the items of the NEO-FFI to measure the Big Five. Peer reports are an established source of data in personality research (Funder, 1987; Kenny, 1994). Depending on the number of peers in the study and additional moderator variables (Chaplin, 1991; Paunonen, 1989) they correlate consistently but far from perfectly among each other and with self-report data. Several factors have been identified that mediate consensus among peers (Kenny, 1991). Assimilation effects in peer report data can be reduced to the degree that peers know one of the twins very well but have little or no contact with the other. More important, peers judging only one of a pair of DZ twins should be less susceptible to contrast effects.

Our study has five goals. First, we will estimate environmental and genetic influences using an established measure of the five-factor model. Estimates will be compared between the methods. If the estimates for self-report data deviate consistently from those for peer report data, this would be an indicator of contrast or assimilation effects. Second, peer report data provided by more than one peer allow the separation of effects due to environmental influences on personality traits that are not shared by twins from measurement error. Third, we will provide heritability estimates for personality constructs that are measured by more than one method. Fourth, and closely related to the third point, we will estimate genetic and environmental correlations between self- and peer report measures, which are important if we are to understand the genetic and environmental basis of the correlation between self- and peer reports. For example, even if we find high heritability estimates for both sources of data, they may reflect different genetic effects. Thus, a joined analysis of self- and peer reports may obscure the underlying processes. If this is the case, more detailed analyses should be carried out for both measures separately. On the other hand, the identification of the components that make up the correlation between self- and peer reports provides information on the basis of this agreement.
Finally, we will test for age differences in estimates of genetic and environmental influences on personality measures (differential heritability). Since our study has a cross-sectional design, age effects are confounded with cohort effects. Recent meta-analyses of differential heritability (McCartney, Harris, & Bernieri, 1990; Pedersen, 1993) found that heritability is somewhat lower in older age groups than in younger groups, but due to methodological problems and inconsistent findings, the size of this effect is still undetermined. Bergeman et al. (1993) found no significant age effects for Openness, Agreeableness, and Conscientiousness.

**METHOD**

**Participants**

Participants were 509 female and 151 male MZ twin pairs and 149 female, 51 male, and 104 opposite sex twin pairs diagnosed as DZ. All twin pairs volunteered for this study. Because there is no population-based twin register in Germany, all participants were approached through announcements in newspapers, magazines, radio and TV stations, or through twin clubs and twin meetings. Participants were between 14 and 80 years old ($M = 32.98$, $SD = 13.40$ years). There were no marked differences between MZ and DZ twins with respect to age, education, or occupational status.

**Zygosity Determination**

Zygosity of twins was determined using physical similarity criteria. Participants completed a zygosity questionnaire (Oniszczenko, Angleitner, Strelau, & Angert, 1993) in which they were asked to describe and compare themselves with their co-twin on a number of physical characteristics (e.g., height, hair color, texture of hair, eye color) and to judge the extent of twin confusion by parents, relatives, peers, and strangers. From a subsample of 110 (84 MZ, 38 DZ) same sex twin pairs, blood samples were obtained during a separate medical examination. For these participants, zygosity was determined by the analysis of single-gene markers with a probability of less than .0001 to falsely classify a DZ pair as MZ. Discriminant function scores and classification functions were calculated in this subsample to predict the zygosity of the remaining 750 same sex twin pairs. Among the 15 variables entered in the analysis, extent of confusion by strangers, differences in eye color, height, hair color, and suffering from illness yielded the highest standardized discriminant function coefficients. The predicted zygosity was correct 93.64% of the time in the subsample, a result which is in the range of hit rates typically achieved.
by zygosity questionnaires (Lykken, 1978; Nichols & Bilbro, 1966; Plomin, DeFries, & McClearn, 1990). Misclassifications occurred more frequently for DZ twins than for MZ twins: Six DZ twins were classified as MZ, while only one MZ twin was classified as DZ.

**Materials**

A number of temperament and personality questionnaires and the inventory for zygosity diagnosis were mailed to participants. Among these measures was the German version of the NEO-FFI (Costa & McCrae, 1989; Borkenau & Ostendorf, 1993), an established measure of the five-factor model of personality. The peer report version of the NEO-FFI was identical to the self-report version, except that the first-person form had been changed to the third-person form.

**Procedure**

Participants were instructed to complete the questionnaires independently. In addition, each twin gave the peer report versions of the questionnaires to two peers who knew one twin but (preferably) not the co-twin very well. Participants returned self- and peer reports by mail; the latter were sealed by the peers in separate envelopes.

Peers were mostly friends (62%), relatives (16%), spouses (10%), and colleagues (9%) who had known the participants for 11.06 (SD = 10.46) years on average and who, on the whole, judged their acquaintance with the target person as “very good” or “good” (82%). Very few (1%) indicated that they had little or very little knowledge about the target. The majority of peers were female (62%).

**Analyses**

Personality scores on the NEO-FFI change significantly with age and are correlated with the respondents’ sex (Borkenau & Ostendorf, 1993). Because the ages of twin pairs were obviously perfectly correlated, age was expected to contribute to the correlation between twins. Covariance between twins due to maturation may thus be misrepresented as caused by environmental influences shared by twins (Eaves et al., 1989; McGue & Bouchard, 1984). In the same manner sex differences resulted in spuriously higher correlations because most of our participants were same sex twins.

To control for artificial sex and linear age effects on the correlations among our MZ and DZ twin samples, age- and sex-corrected scores were computed. We regressed the twins’ NEO-FFI scores on their age and sex and computed standardized residuals (i.e., the standardized difference between
We used maximum-likelihood model fitting to estimate the influence of genetic and environmental sources of variance on phenotypic individual differences. Goodness-of-fit indexes and parameter estimates were obtained using LISREL 8.03 (Jöreskog & Sörbom, 1993). These analyses were limited to age- and sex-corrected scores of same sex twin pairs using variance-covariance matrices. Due to the small number of male DZ twins, more detailed analyses of genetic and environmental effects on male and female phenotypes would add little, if any, additional information. We considered four sources of variance in our analyses: (a) additive genetic effects at multiple loci (A), (b) genetic dominance effects at multiple loci (D), (c) environmental effects shared by twins (C), and (d) nonshared environmental influences (E). When the analyses were based exclusively on self-report data (Model 1), E confounded variance due to nonshared environmental effects and error variance. The availability of two peer reports per twin (Model 2) and of self- and peer reports (Model 3) allowed us to separate error variance from nonshared environmental variance. The models fitted to our data are depicted in Figure 1 (a–c).

Model 1 is a standard model for the analysis of twin data. The twins’ phenotypic personality traits were measured by a single indicator (i.e., NEO-FFI scale). We fixed the paths from phenotypic trait expressions (P1 and P2) to their respective indicators to unity. In Model 2 there were two indicators of each phenotypic trait, and the paths from traits to their measures and the measures’ error variances were estimated. Model 3 was the rater bias model described by Neale and Cardon (1992). It differed from Model 2 insofar as two method factors were introduced. Paths from the method factors to the measures were estimated for each method (self- vs. peer report) separately. They were constrained to equality within methods and between MZ and DZ twin groups. Although self- and peer reports were provided by different judges for each twin, we decided to include method factors in the model because the similarity among self- and among peer ratings, even if they stem from different persons, may be greater than the similarity between self- and peer ratings. For example, self-raters who were reared together and shared genetic influences may have employed more similar meaning systems than peer raters, who showed no family resemblance. In turn, peer ratings may have reflected shared stereotypes on the relation between physical and psychological characteristics or rare but striking behaviors of twins to a higher degree than self-report measures.

In a study of twins reared together, D and C effects cannot be estimated simultaneously. Following the procedure described in detail by Neale and Cardon (1992) we tested the following hypotheses: (a) The data do not indicate any family resemblance (E model), (b) family resemblance is caused by additive genetic effects (AE model), (c) family resemblance is caused by
Figure 1
Structural Models for Self-Report, Peer Report, and Self- and Peer Report Data

Note. A = additive genetic effects; D = genetic dominance effects; C = shared environmental effects; E = specific environmental effects; P₁ and P₂ = phenotype of Twin 1 or 2; R₁₁ to R₂₄ = peer report on Twin 1 or 2 by Rater 1 to 4; SR₁ and SR₂ = self-report of Twin 1 or 2; PR₁ and PR₂ = peer report (averaged across two raters) on Twin 1 or 2; Mₛ = method effects for self-reports; Mₚ = method effects for peer reports. Parameters alpha, beta, and gamma indicate the degree of shared environment, additive genetic, and genetic dominance relationship between members of a pair: α = 1.0 for all twins; β = 1.0 for MZ twins; β = .5 for DZ twins; γ = 1.0 for MZ twins; γ = .25 for DZ twins.
environmental influences shared by twins (CE model), (d) family resemblance is due to additive genetic and dominance genetic effects (ADE model), and (f) family resemblance is caused by additive genetic plus shared environmental influences (ACE model). Thus, we identified the best-fitting model by starting from the model with the fewest parameters (usually the AE model) that was not rejected significantly. We then tested whether models with more parameters (ADE, ACE) fitted the data significantly better than the reduced model using $\chi^2$ difference tests. However, we did not use the DE model because it is highly unlikely that there are genetic dominance effects without additive genetic effects (Roberts, 1967). If either of the two environmental models could not be rejected, this result is also reported.

**RESULTS**

**Psychometric Quality of the Data**

The internal consistency of the self- and peer report scales of the NEO-FFI for Neuroticism, Extraversion, Agreeableness, and Conscientiousness was acceptable. The internal consistency for Openness to Experience can be judged as satisfactory. On average, the internal consistency was marginally higher for the peer reports than for self-reports (see Table 1), and it was lower for Openness than for the other traits.

Altogether, the agreement between the two raters was good. On average, Spearman-Brown corrected intraclass correlations between the peer ratings were higher than .60 (see Table 1). Rater agreement showed no substantial differences between MZ and DZ twins but moderate differences between male and female targets. Averaged across scales, rater agreement was .66 for male and .59 for female MZ twins and .60 for both female and male DZ twins.

Correlations between self-reports and averaged peer reports were in the range of .46 to .60, indicating high validity for the self- and peer report measures. They were only slightly inflated by age and sex differences. Table 1 lists correlations of age- and sex-corrected scores.

**Self-Report Data**

We calculated twin similarities (intraclass correlations) for raw scores and for age- and sex-corrected scores. Both analyses yielded nearly identical results. Therefore we limit our presentation of results to the corrected data.

Intraclass correlations for the self-report data (Table 2) corresponded closely to the data summarized by Loehlin (1992). The greatest deviation
**Table 1**
NEO Five-Factor Inventory: Psychometric Characteristics for Self- and Peer Report Measures

<table>
<thead>
<tr>
<th>Scales</th>
<th>Scale reliability (Cronbach’s alpha)</th>
<th>Agreement among Peers (ICC 2,2)</th>
<th>Correlations between self- and peer reports</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self ((N = 1,928))</td>
<td>Peer (^{b}) ((N = 3,856))</td>
<td>((N = 1,928))</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.85</td>
<td>.85</td>
<td>.63</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.80</td>
<td>.79</td>
<td>.65</td>
</tr>
<tr>
<td>Openness</td>
<td>.63</td>
<td>.62</td>
<td>.59</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.69</td>
<td>.78</td>
<td>.59</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.82</td>
<td>.86</td>
<td>.61</td>
</tr>
<tr>
<td>Mean</td>
<td>.77</td>
<td>.79</td>
<td>.61</td>
</tr>
</tbody>
</table>

Note. Cronbach’s alpha was calculated on raw scores. The remaining statistics are based on scale values corrected for age and sex effects.

a. Values are Spearman-Brown corrected intraclass correlations (ICC 2,2; Shrout & Fleiss, 1979).

b. Statistics were calculated for each twin separately and then averaged across twins.

from Loehlin’s results was the correlation between DZ twins for Extraversion (.20 in Loehlin’s meta-analysis vs. .28 for our sample). This correlation can be traced back to the similarity between female DZ twins (.41). Although correlations for female DZ twin pairs were higher and those for female MZ twins were lower than the correlations for the respective male twins, only the difference between male and female DZ twins on the Extraversion scale was significant \(T = 1.51, df = 71, 158, p < .05; Alsawalmeh & Feldt, 1992\).

Maximum-likelihood model fitting using same sex MZ and DZ twins (Model 1) showed that a reduced model encompassing additive genetic effects (A) and specific environmental and error influences (E) was most appropriate to represent the self-report data for every NEO-FFI scale. Goodness-of-fit statistics and maximum-likelihood parameter estimates for additive genetic and specific environmental influences (including measurement error) are shown in Table 3. Genetic and environmental contributions to the phenotypic variance of the Neuroticism, Extraversion, Openness, and Conscientiousness scales are almost identical. Additive genetic effects explained between 52% and 56% of the variability on these scales. Additive genetic influences on the Agreeableness scores were substantially smaller (42%).
### Table 2

Intraclass Correlations (ICC 1,1) of the NEO Five-Factor Inventory Self-Report Scales for Monozygotic (MZ) and Dizygotic (DZ) Twins

<table>
<thead>
<tr>
<th>Scales</th>
<th>MZ</th>
<th>DZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N = 660)</td>
<td>Male (N = 151)</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.53</td>
<td>.60</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.56</td>
<td>.59</td>
</tr>
<tr>
<td>Openness</td>
<td>.54</td>
<td>.55</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.42</td>
<td>.49</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.54</td>
<td>.52</td>
</tr>
<tr>
<td>Mean</td>
<td>.52</td>
<td>.55</td>
</tr>
</tbody>
</table>

Note. Statistics are based on scores corrected for age and sex effects.
Table 3
NEO Five-Factor Inventory Self-Report: Percentage of Phenotypic Variance Explained by Additive Genetic ($a^2$) and Specific Environmental Effects ($e^2$)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$</th>
<th>$a^2$</th>
<th>$e^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>AE</td>
<td>2.03</td>
<td>4</td>
<td>.73</td>
<td>.52</td>
<td>.48</td>
</tr>
<tr>
<td>Extraversion</td>
<td>AE</td>
<td>3.50</td>
<td>4</td>
<td>.48</td>
<td>.56</td>
<td>.44</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>AE</td>
<td>3.38</td>
<td>4</td>
<td>.50</td>
<td>.53</td>
<td>.47</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>AE</td>
<td>3.17</td>
<td>4</td>
<td>.53</td>
<td>.42</td>
<td>.58</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>AE</td>
<td>9.93</td>
<td>4</td>
<td>.03</td>
<td>.53</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. See text for description of the models. $A =$ additive genetic effects; $E =$ specific environmental and error influences.

Peer Report Data

Intraclass correlations based on age- and sex-corrected peer report data averaged across the two peers are summarized in Table 4. For both groups of twins, correlations of averaged peer reports were generally smaller than those for self-reports if no correction for attenuation was applied.

As in the self-report data, maximum-likelihood model fitting suggested that a model allowing for additive genetic and specific environmental effects described the covariance among peer reports of MZ and DZ twin pairs most appropriately. However, for the Agreeableness scale we could not dismiss an environmental model (CE) unequivocally. Fitting an ADE model to the peer reports for Neuroticism resulted in parameter estimates for the additive genetic component that were close to zero. Thus, for this scale a model encompassing genetic dominance ($D$) and specific environmental effects ($E$) yielded the same fit ($\chi^2 = 8.29$, $df = 10$, $p = .60$) as the ADE model. This reflects the twin similarities, which were much higher for MZ twin pairs (.40) than for same sex DZ twin pairs (.05). An AE model was less appropriate for Neuroticism but could not be rejected as an adequate description of the data.

The ACE model yielded the best fit for peer reports on the Extraversion, Openness to Experience, and Agreeableness scales. However, for none of these scales did the inclusion of shared environment influence ($C$) improve the fit significantly. The CE model could not be rejected for Agreeableness. The structure of peer reports on the Conscientiousness scale was best described by an ADE model. Inclusion of the $D$
<table>
<thead>
<tr>
<th>Scales</th>
<th>MZ</th>
<th>DZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>(N = 660)</td>
<td>(N = 151)</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.40</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>(.63)</td>
<td>(.66)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.38</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>(.58)</td>
<td>(.64)</td>
</tr>
<tr>
<td>Openness</td>
<td>.49</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>(.82)</td>
<td>(&gt;1.0)</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.32</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>(.55)</td>
<td>(.66)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.41</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>(.70)</td>
<td>(.67)</td>
</tr>
<tr>
<td>Mean</td>
<td>.40</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>(.67)</td>
<td>(.66)</td>
</tr>
</tbody>
</table>

Note. Statistics are based on scores corrected for age and sex effects. Values in parentheses are intraclass correlations corrected for attenuation. 

a. Corrected intraclass correlation for Openness omitted.
parameter, however, did not improve the fit significantly. The model that allowed solely for specific environmental influences was rejected for each scale at the .001 significance level.

Maximum-likelihood parameter estimates for all models that could not be rejected are summarized in Table 5. Because our model for peer report data allows for a separation of variance due to specific environmental influences and measurement error, we yielded substantially higher estimates for genetic influences on personality variables compared to the self-report data. Genetic influences were highest for Openness (81%) and Conscientiousness (71%). About 60% of the variance on the Neuroticism, Extraversion, and Agreeableness scales was explained by genetic influences.

**Self- and Peer Report Data**

To obtain the most reliable parameter estimates we analyzed self- and peer report data in a joint model. First, we tested whether a model without method factors (Model 2) fits the variance-covariance matrices for self- and peer reports. For all scales except Agreeableness, this model had to be rejected. However, the model including method factors (Model 3) yielded a substantially better fit also for Agreeableness. Thus, we restrict our presentation to the results for Model 3. In some of our analyses the loadings on one of the method factors approached zero and could not be identified. If this was the case, we fixed the loadings on the method factor for this and all nested analyses to zero.

The reduced AE model was most appropriate for Openness to Experience and Agreeableness. For Extraversion the ACE model yielded the best fit, but this model was not significantly superior to the AE model ($\chi^2 = 1.11, df = 1, ns$). For Openness to Experience, both the ACE and the AE models were rejected at the .05 level. Inclusion of shared environmental influences for this scale did not improve the fit significantly ($\chi^2 = 2.05, df = 1, ns$). Allowing for genetic dominance variance improved the fit for Agreeableness only marginally ($\chi^2 = 0.05, df = 1, ns$).

For Neuroticism and Conscientiousness, genetic dominance was an important source of variance. Inclusion of D improved the fit for Neuroticism significantly ($\chi^2 = 4.60, df = 1, p < .05$) over the AE model. Fitting an ADE model for the Neuroticism data resulted in a parameter estimate for additive genetic variance close to zero. Thus a reduced DE model yielded the same fit. For Conscientiousness the comparison
Table 5
NEO Five-Factor Inventory Peer Report:
Percentage of Phenotypic Variance Explained by Additive Genetic (a^2), Genetic Dominance (d^2), Shared Environmental (c^2),
and Specific Environmental Effects (e^2)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>p</th>
<th>a^2</th>
<th>d^2</th>
<th>c^2</th>
<th>e^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>AE</td>
<td>11.89</td>
<td>10</td>
<td>.29</td>
<td>.61</td>
<td></td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DE</td>
<td>8.29</td>
<td>10</td>
<td>.60</td>
<td></td>
<td>.63</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>AE</td>
<td>16.09</td>
<td>10</td>
<td>.10</td>
<td>.60</td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>AE</td>
<td>12.94</td>
<td>10</td>
<td>.23</td>
<td>.81</td>
<td></td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>AE</td>
<td>8.41</td>
<td>10</td>
<td>.59</td>
<td>.57</td>
<td></td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>8.64</td>
<td>10</td>
<td>.57</td>
<td></td>
<td>.52</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>AE</td>
<td>16.01</td>
<td>10</td>
<td>.10</td>
<td>.71</td>
<td></td>
<td></td>
<td>.29</td>
</tr>
</tbody>
</table>

Note. See text for description of the models. A = additive genetic effects; E = specific environmental and error influences; D = genetic dominance effects; C = shared environment influence.

of an ADE model to an AE model only marginally failed the level of significance (\( \chi^2 = 3.47, df = 1, p < .07 \)). For all scales, solely environmental models (CE and E) had to be rejected.

Measuring personality traits by a combination of self- and peer reports resulted in higher estimates of genetic variance components for all but the Openness to Experience scale. About 70% of the phenotypic variance on each scale was due to additive genetic or genetic dominance effects (see Table 6). Genetic effects were strongest for Openness to Experience (79%) and Conscientiousness (about 75%). Additive genetic effects explained 68% of the variation in Extraversion and 66% of the variance in Agreeableness. Genetic dominance accounted for 68% of the variance in Neuroticism.

Genetic and Environmental Correlations between Self- and Peer Reports

Usually genetic/environmental correlations are estimated to explain the covariation between different traits, not methods. Because phenotypic correlations of self- and peer reports of personality traits are not perfect (even when corrected for attenuation) it is informative to estimate the extent to which genetic or environmental factors shared by both measures contribute to the observed correlation between self- and peer reports (see Table 1). Note that the high genetic (or environmental)
Table 6

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( p )</th>
<th>( a^2 )</th>
<th>( d^2 )</th>
<th>( c^2 )</th>
<th>( e^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>DE</td>
<td>9.17</td>
<td>12</td>
<td>.61</td>
<td>.68</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>AE</td>
<td>10.12</td>
<td>12</td>
<td>.61</td>
<td>.68</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>AE</td>
<td>23.06</td>
<td>12</td>
<td>.03</td>
<td>.79</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>AE</td>
<td>9.64</td>
<td>12</td>
<td>.65</td>
<td>.66</td>
<td>.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>AE</td>
<td>19.97</td>
<td>12</td>
<td>.07</td>
<td>.75</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADE</td>
<td>16.50</td>
<td>11</td>
<td>.12</td>
<td>.10</td>
<td>.66</td>
<td>.24</td>
<td></td>
</tr>
</tbody>
</table>

Note. See text for description of the models. D = genetic dominance effects; E = specific environmental and error influences; A = additive genetic effects.

correlations are not a necessary consequence of high genetic (or environmental) influences on the measures in question. If, for example, we found perfect heritability for self- and peer report measures of a trait and a nonperfect correlation between these measures, then both measures capture partially nonshared genetic effects. If the genetic correlations between self- and peer reports were low, this would imply that they are based on different genetic influences (mediated by physiological processes), a conclusion that is at odds with an understanding of these measures as indicators of a single disposition.

We used maximum-likelihood parameter estimates (LISREL 8) to calculate genetic and environmental correlations between self- and averaged peer reports. Environmental and genetic covariance matrices were derived from a Cholesky factorization; these matrices were then transformed to correlations (see Neale & Cardon, 1992, p. 252, for a detailed description of the model and calculation of correlations). Note that these correlations are independent of the ordering of the latent variables (Loehlin, 1996). For all NEO-FFI scales an AE model was compatible with the data and CE and E models had to be rejected significantly. Only for Neuroticism did a DE model yield fit values that were marginally better than those for the AE model \( (\chi^2 = 8.02, df = 14, \text{ vs. } \chi^2 = 9.94, df = 14) \).

Genetic correlations ranged from .77 for Neuroticism to .87 for Extraversion (Table 7). Environmental correlations were substantially smaller. They were smallest for Openness to Experience (.25), Agreeableness...
Table 7
Genetic and Environmental Correlations between Self- and Averaged Peer Reports for the NEO Five-Factory Inventory Scales: Correlations of Additive Genetic (A) or Genetic Dominance (D) and Specific Environmental Effects (E) Based on Maximum-Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$</th>
<th>A</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>AE</td>
<td>9.93</td>
<td>14</td>
<td>.77</td>
<td>.77</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DE</td>
<td>8.02</td>
<td>14</td>
<td>.88</td>
<td>.77</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>AE</td>
<td>10.13</td>
<td>14</td>
<td>.75</td>
<td>.87</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>AE</td>
<td>22.39</td>
<td>14</td>
<td>.07</td>
<td>.86</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>AE</td>
<td>13.72</td>
<td>14</td>
<td>.47</td>
<td>.84</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>AE</td>
<td>20.16</td>
<td>14</td>
<td>.13</td>
<td>.81</td>
<td>.26</td>
<td></td>
</tr>
</tbody>
</table>

Note. See text for descriptions of the models.

(.25), and Conscientiousness (.26) and somewhat higher for Neuroticism (.33, .32) and Extraversion (.38).

Age/Cohort Effects
Since the age range of the participants in our study covers nearly 7 decades of development, we did additional analyses to study whether genetic and environmental effects change with age. The correction for sex and age effects controls for spurious correlations between twins; it may, however, still be the case that twins grow apart during their life span development or that environmental influences are weaker in older adults. Correlations with age (−.14 for Neuroticism, −.09 for Extraversion, −.10 for Openness, .07 for Agreeableness, .19 for Conscientiousness) indicated significant but small mean level changes of NEO-FFI self-report scores with age/cohort. In the peer report data only the correlations for Agreeableness (.06) and Conscientiousness (.22) were significant.

We first correlated absolute intrapair differences on the NEO-FFI scales with age separately for self- and peer reports. These correlations were computed for raw scores and age- and sex-corrected scores. Both analyses yielded nearly identical results and we restrict our presentation to intrapair differences based on raw scores (Table 8). None of these correlations can be regarded as substantial, but most were negative, which indicates that the older twins tended to be more similar. Only three of these correlations were significant, those for MZ twins'
Table 8
Correlations between Age and Absolute Intrapair Differences on NEO Five-Factor Inventory Scales Based on Raw Scores for Self- and Peer Reports

<table>
<thead>
<tr>
<th>Scale</th>
<th>MZ Self-report</th>
<th>Peer report</th>
<th>DZ same sex Self-report</th>
<th>Peer report</th>
<th>DZ opposite sex Self-report</th>
<th>Peer report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>-.09*</td>
<td>.03</td>
<td>-.12</td>
<td>.01</td>
<td>-.05</td>
<td>.20*</td>
</tr>
<tr>
<td>Extraversion</td>
<td>-.04</td>
<td>-.07</td>
<td>.01</td>
<td>.00</td>
<td>-.02</td>
<td>.06</td>
</tr>
<tr>
<td>Openness</td>
<td>-.04</td>
<td>-.05</td>
<td>-.07</td>
<td>-.01</td>
<td>-.08</td>
<td>-.15</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-.04</td>
<td>-.03</td>
<td>-.02</td>
<td>.02</td>
<td>.12</td>
<td>-.14</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.10**</td>
<td>-.05</td>
<td>-.13</td>
<td>-.13</td>
<td>-.03</td>
<td>-.05</td>
</tr>
</tbody>
</table>

Note. MZ = monozygotic; DZ = dizygotic.
*p < .05
**p < .01.

self-reports on Neuroticism and Conscientiousness, and the positive correlation for opposite sex twins’ peer reports on Neuroticism. The significant correlations for MZ twins’ self-reports were matched by higher (negative) correlations for DZ twins on the same scales.

Intrapair differences are a simple but psychometrically problematic measure. Thus we employed a second strategy to reveal age/ages-group effects. We divided our sample into four groups: older MZ twins (age 29 to 80 years, \(N = 322\)), younger MZ twins (age 14 to 28 years, \(N = 338\)), and corresponding DZ twin groups (older, \(N = 101\); younger, \(N = 99\)). Maximum-likelihood parameter estimates were obtained for self- and peer reports (age- and sex- corrected scores) using the best-fitting models described above. The goodness-of-fit was compared between a reduced model with equal genetic and environmental effects for the age groups, and a model allowing for separate estimates of these parameters for older and younger twins. In most of the comparisons the model allowing for age effects was not significantly superior to the reduced model. In self-reports significant age effects were found for Agreeableness (\(\chi^2 = 6.57, df = 2, p < .05\)) and Conscientiousness (\(\chi^2 = 12.13, df = 2, p < .01\)). For Agreeableness, parameter estimates were .48 for additive genetic effects (.52 for nonshared environment) in the younger group and .35 (.65) for the older group. For Conscientiousness, parameter estimates differed only marginally. Additive genetic effects were .52 for the younger and .53 for the older group. No significant age effects
were found in peer reports, although the effect for Conscientiousness was nearly replicated ($\chi^2 = 5.40, df = 2, p < .07$; additive genetic effect for younger group, .69, vs. .78 for older group).

To summarize, the results of these two analyses do not support general claims about the association between age and genetic or environmental effects on personality. We found stronger age effects in self-reports than in peer reports. These were most consistent for Conscientiousness but, contrary to expectations, on this scale older twins were more similar than younger twins.

**DISCUSSION**

The results of our study support six major conclusions. First, our analyses of self-report data replicate earlier findings of a substantial genetic influence on all broad personality dimensions measured by the NEO-FFI. Second, this influence is not only demonstrated for self-report data but also for peer reports. Third, the pattern of genetic and environmental influences is similar for self- and peer report data; thus our analyses validate findings solely based on self-reports and invalidate the hypothesis that self-report data of DZ twins are substantially distorted by contrast effects. Fourth, self-report studies of single personality measures tend to underestimate genetic influences because they do not allow the separation of error variance from variance due to specific environmental influences. Fifth, correlations between self- and peer reports reflect the same genetic influences to a much higher extent than identical environmental influence. Finally, there is little evidence of differential heritability across the age cohorts. We shall briefly discuss each of these conclusions in turn.

**Self-Reports**

If we compare the correlations of self-reports between MZ and DZ twins that were obtained in our study to those of the twin studies summarized by Loehlin (1992), Bergeman et al. (1993), and Jang et al. (1996), striking similarities are obvious. Consistent differences between the five scales are hardly noticeable across the studies, although MZ twin correlations tend to be highest for Extraversion and lowest for Agreeableness. DZ twins tend to correlate highest on Openness. Our heritability estimates are highly similar to those reported by Jang et al. (1996) for the NEO-PI-R and Bouchard’s (1994) results for twins
riemann et al.

reared together. They are higher than those reported by Loehlin (1992) and Bergeman et al. (1993), and genetic dominance effects (or the influence of more similar MZ twins' environments) are absent in our data. This difference may be caused by the differences in the questionnaire measures, divergent procedures regarding the recruitment of twins, and the absence of twins reared apart in our study.

**Peer Reports**

The analysis of peer report data clearly demonstrates family resemblance for the five dimensions under study. Comparing the results for self-report and peer report data across the NEO-FFI scales, we found a substantial similarity. A model that explains family resemblance via additive genetic and specific environmental effects is most appropriate for each self-report scale and cannot be rejected for any peer report measure. However, tendencies that are observed for the self-reports seem to be stronger in the peer report data. For self-reports we found the highest difference between MZ twin and DZ twin correlations for Neuroticism and the smallest difference for Agreeableness. For these scales, models allowing for genetic dominance and shared environmental effects, respectively, could not be rejected in the domain of peer report data.

It is important to note that the differences for these two scales are in opposite directions. For Neuroticism the difference between MZ and DZ twins is more pronounced (due to low DZ twin correlations) in peer reports, whereas for Agreeableness this difference becomes smaller (due to low MZ twin correlations). This pattern of results contradicts suggestions that self-reports of MZ and DZ twins are generally distorted by assimilation or contrast effects and validates self-report studies.

**Self- and Peer Reports**

We regard the joint analysis of self- and peer reports as providing the most reliable and valid parameter estimates in our study because it utilizes all available measures and derives trait scores from independent sources. It emphasizes the importance of additive genetic and nonshared environmental influences on Extraversion, Openness to Experience, and Agreeableness, and of genetic dominance effects for Neuroticism and Conscientiousness. Despite the similarities between self- and peer report data, both sources of data result in substantially different esti-
mates of environmental and genetic influences on the five personality variables under study. The increase in the averaged estimates of genetic influences from .51 for self-reports to .66 for peer report measures is a consequence of separating error variance from nonshared environmental influences in the latter analyses. This is comparable to a correction for attenuation in classical test theory. Subtracting error variance from nonshared environmental effects (taking the average internal consistency of the NEO-FFI scales as a crude estimate of true score variance) simulates this effect for the self-report data and yields heritability estimates of about .70.

The finding that heritability estimates for the joint analysis of self- and peer reports are somewhat higher (.71 on average) than those for peer reports is not merely a trivial result of aggregating self- and peer reports into a more reliable measure because error variance is controlled for in both analyses. The joint analysis demonstrates that the broad personality traits (constructs) measured by independent judgments are strongly determined by genetic influences. This result suggests that the covariance between self- and peer reports is largely due to shared genetic effects, which is confirmed by our analysis of genetic and environmental correlations. The correlations of additive genetic effects were on average higher than .80. Nonshared environmental effects are reflected commonly by self- and peer reports to a substantially smaller extent (.32). This result implies that behavior-genetic research can profit from the joint analyses of self- and peer report data.

None of our analyses shows marked differences between the dimensions of the five-factor model regarding genetic versus environmental influences. Although heritability estimates were lowest for Agreeableness in all of our analyses, the difference between this scale and Neuroticism and Extraversion is only marginal in the analyses involving peer reports. The latter data result in estimates of genetic influences on Openness to Experience and Conscientiousness somewhat higher than those for the remaining scales. Although previous studies have shown the importance of nonadditive genetic effects for Neuroticism, Conscientiousness, and Extraversion, these effects are particularly high in our data for Neuroticism and Conscientiousness. The robustness of these findings has to be tested in independent samples.

The present analyses are based on only one instrument to measure the five-factor model, the NEO-FFI. Thus it remains to be seen how the inclusion of multiple measures of the five-factor model, in combination with multiple perspectives (self- vs. peer), will affect our findings. An
extension of the multitrait multi-observer approach presented in this study may help to clarify the genetic and environmental structure of personality traits. A thorough understanding of this structure is also necessary for the interpretation of molecular-genetic studies linking personality traits to quantitative gene loci (Cloninger, Adolfsson, & Svrakic, 1996; Ebstein et al., 1996; Plomin, 1995).

Age Effects

The absence of substantial differential heritability of the five factors across age/cohort groups replicates earlier findings of Bergeman et al. (1993). It is, however, at odds with recent meta-analyses. This discrepancy is difficult to explain. It would be informative to examine whether differential heritability is systematically associated with specific facets of the five-factor model. Likewise, differential heritability may be stronger in comparisons including children and adolescents, who are underrepresented in Bergeman et al.'s and the present study.

CONCLUSION

In conclusion, the present study has shown that the inclusion of peer reports into behavior-genetic studies of personality variables provides a valuable tool. Peer report data provide the opportunity to validate the results of self-report studies and, more important, they shift the focus from personality measures to personality constructs. Further research should include additional sources of data (e.g., objective personality measures or physiological indicators) or multiple questionnaires to conduct systematic multitrait multimethod behavior-genetic studies of personality.

REFERENCES

Twin Study of NEO-FFI Scales


This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.