Psychological Bulletin
2015, Vol. 141, No. 4, 769–785


0033-2909/15/$12.00 http://dx.doi.org/10.1037/bul0000017

# Heritability of Personality: A Meta-Analysis of Behavior Genetic Studies

Tena Vukasović and Denis Bratko
University of Zagreb

The aim of this meta-analysis was to systematize available findings in the field of personality heritability and test for possible moderator effects of study design, type of personality model, and gender on heritability estimates. Study eligibility criteria were: personality model, behavior genetic study design, self-reported data, essential statistical indicators, and independent samples. A total of 134 primary studies with 190 potentially independent effect sizes were identified. After exclusion of studies that did not meet inclusion criteria and/or met 1 of the exclusion criteria, the final sample included 62 independent effect sizes, representing more than 100,000 participants of both genders and all ages. Data analyses were performed using the random-effects model, software program R package metafor. The average effect size was .40, indicating that 40% of individual differences in personality were due to genetic, while 60% are due to environmental influences. After correction for possible publication bias the conclusion was unaltered. Additional analyses showed that personality model and gender were not significant moderators of personality heritability estimate, while study design was a significant moderator with twin studies showing higher estimates, .47, compared to family and adoption studies, .22. Personality model also was not a significant moderator of heritability estimates for neuroticism or extraversion, 2 personality traits contained in most personality trait theories and/or models. This study is the first to empirically test and confirm moderator effect of study design on heritability estimates in the field of personality. Limitations of the study, as well as suggestion for future studies, are discussed.

*Keywords:* personality, behavior genetics, heritability, meta-analysis

*Supplemental materials:* http://dx.doi.org/10.1037/bul0000017.supp



Most people have their own implicit personality theory. We could even go as far as to say that there are as many implicit personality theories as there are people in the world. Even experts in the field of personality psychology are not unanimous when it comes to defining personality. Larsen and Buss (2005) provided a definition of *personality* stating: "Personality is the set of psychological traits and mechanisms within the individual that are organized and relatively enduring and that influence his or her interactions with, and adaptations to, the intrapsychic, physical, and social environments" (p. 4). The premise of a smaller number of relatively stable personality dimensions is typical of the trait perspective, the dominant theoretical perspective in the field of personality psychology in the past decades. Even though authors have started from the same premise, they developed a wide array of personality models differing in both number and conceptualization of basic personality traits. However, there are two traits that have in some form been a part of almost every personality model or theory—neuroticism and extraversion. Due to the limited space, we give a very brief overview of four personality trait taxonomies relevant to this study and a brief introduction to the field of behavior genetics.

This article was published Online First May 11, 2015.

Tena Vukasović and Denis Bratko, Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb.

Correspondence concerning this article should be addressed to Tena Vukasović, Department of Psychology, Faculty of Humanities and Social Sciences, Ivana Lučića 3, 10000 Zagreb, Croatia. E-mail: tvukasov@ffzg.hr

## Theoretical Frameworks

Cattell (1943) developed a trait taxonomy consisting of 16 intercorrelated primary personality traits at the first level, with the possibility of calculating second-order factors. Eysenck (1970) also developed a hierarchical personality model, but with three independent traits at the top of hierarchy: neuroticism, extraversion, and psychoticism. Tellegen (1982), motivated to provide a better understanding of the lower levels of personality, developed a questionnaire measuring 11 primary scales and three higher order factors: negative emotionality, positive emotionality, and constraint. In the 1990s there was at least an initial consensus on a general taxonomy of personality traits known as the Big Five model (Goldberg, 1990) measuring emotional stability, extraversion, agreeableness, conscientiousness, and intellect or the five-factor model (Costa & McCrae, 1992) measuring neuroticism, extraversion, agreeableness, conscientiousness, and openness. John, Naumann, and Soto (2008) demonstrated that the number of publications related to either the Big Five personality traits or to the influential models developed earlier by Cattell and Eysenck, identified by the keyword searches of the PsycINFO database in the last 30 years, showed a paradigm shift to the integrative Big Five trait taxonomy.

As Bouchard and Loehlin (2001) showed by comparing six different trait taxonomies, various models could be seen as "dividing up much the same pie in slightly different ways, with some unevenness in the density of coverage at the lower levels" (p. 247). For example, the impulsiveness facet is included under psychoticism in Eysenck's (1970) model and under neuroticism in the

five-factor model (Costa & McCrae, 1992). Its opposite, control, is a part of the higher order trait constraint in Tellegen's (1982) taxonomy, and there is no equivalent facet in Cattell's (1943) taxonomy. A logical question that arises is can we assume that constructs bearing the same labels in different personality models are truly commensurable? There are number of studies that have examined whether different personality measures actually measure the same constructs. But it is only recently that Pace and Brannick (2010) conducted a meta-analysis to give a quantitative review of a multitude of personality scales that examines their commensurability. Their conclusion was that different measures of personality only taxonomically measure the same construct, with scales of the "same" construct correlating only modestly in many cases. Estimated mean convergent validities among all measures were generally below .50, with convergent validity highest among extraversion scales. When the analysis was limited only to measures based on the five-factor model, some improvement occurred with convergent validities in the low to mid .60s (with the exception of openness to experience, which was around .50), still lower then would be expected. These results indicate that the assumption of commensurability is questionable. What could we expect if these taxonomically same, but psychometrically somewhat different, traits were used in behavior genetic studies? Would personality heritability estimates vary depending on a model or would they be the same? There is no straightforward answer to this question. Even if heritability estimates were the same that would not necessarily mean they were based on the same construct. It would only suggest the conclusion of nondifferential personality heritability.

*Behavior genetics* is a scientific field focusing on identifying the genetic and environmental sources of individual differences in a specific characteristic or trait, known as phenotype. If we decompose the total variance of individual differences in a specific phenotype to its components, we can distinguish between additive genetic and nonadditive genetic influences on one hand, and shared environmental and nonshared environmental influences on the other. Heritability, a statistical parameter used in behavior genetics, is the proportion of phenotypic differences among individuals that can be attributed to genetic differences in a particular population. Broad-sense heritability involves all additive and nonadditive sources of genetic variation, whereas narrow-sense heritability is limited to additive genetic variance (Plomin, DeFries, McClearn, & McGuffin, 2008).

Human behavior genetic research focuses on studying naturally occurring genetic and environmental variation and utilizes data from different types of genetically related individuals. There are three broad types of research designs in the field of classical behavior genetics: twin studies, adoption studies, and family studies. Adoption and twin studies are like experiments that can be used to assess the relative contributions of nature and nurture to familial resemblance (Plomin et al., 2008). Results of twin, adoption, and family studies and of combination of these designs converge on the conclusion that personality is substantially heritable, with estimates of genetic contributions to individual differences ranging between 30 and 50% (Plomin et al., 2008). However, it also is known that heritability estimates may vary depending on some methodological variables, such as study design. There has been a strong trend of family and adoption studies that have shown systematically lower estimates of genetic contributions, closer to 30%, compared to twin studies, closer to 50% (Plomin et al., 2008). Family study design has a limitation of not being able to separate additive genetic and shared environmental influences and thus cannot provide a heritability estimate, but can show evidence of familial aggregation. This statistical parameter, computed by doubling the parent–offspring correlation, includes both additive genetic and shared environmental influences that contribute to members of the same family being more similar. In theory, this statistical parameter is the upper limit of the narrow-sense heritability (Plomin, DeFries, & McClearn, 1990). On the other hand, as stated earlier, empirical findings tend to show that this is not the case—twin studies systematically show higher estimates compared to family studies. Previous reviews of personality heritability (e.g., Bouchard, 1993, 1997; Eaves, Eysenck, & Martin, 1989; Johnson, Vernon, & Feiler, 2008; Loehlin, 1992) have shown strong evidence of significant genetic effects to individual differences in personality.

There also have been studies examining potential gender differences in personality heritability (e.g., Eaves, Heath, Neale, Hewitt, & Martin, 1998; Fanous, Gardner, Prescott, Cancro, & Kendler, 2002; Finkel & McGue, 1997; Rettew et al., 2006), but the results have not been entirely consistent. However, a number of studies have shown that there may be relevant gender differences affecting personality heritability estimates. Results based on three large samples from the United States, Australia, and Finland, using a form of (Eysenck & Eysenck, 1975; Eysenck, Eysenck, & Barret, 1985) questionnaire, showed evidence of sex differences in the genetic architecture of neuroticism favoring a greater relative contribution of nonadditive genetic effects in males (Eaves et al., 1998). Finkel and McGue (1997) examined evidence for sex limitation of heritability in the Multidimensional Personality Questionnaire (MPQ) scales in a sample from the Minnesota Twin Family Registry, and concluded that there were significant sex differences in heritability for three of the scales (alienation, control, and absorption). Rettew et al. (2006) examined the magnitude of genetic and environmental influences to variation in adolescent neuroticism assessed by Amsterdamse Biografische Vragenlijst, a self-report personality instrument similar in content to the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975), as a function of age and sex. They concluded that different genes are expressed in Dutch boys and girls in adolescence.

As far as we are aware, there has not been an attempt to systematize available findings using standard meta-analytic procedures and correction techniques (e.g., publication bias, sensitivity analyses) in the field of personality heritability nor has there been an attempt to empirically test for possible moderator effects of study design, type of personality model, or gender on personality heritability estimates.

## The Current Study

The aim of this study was to quantitatively synthesize the existing body of research in the field of personality heritability and to form new knowledge by testing for potential moderating effects of personality model, study design, and gender on heritability estimates. To achieve this aim we formulated five problems and hypotheses. Our first problem was to estimate the average effect of genetic contributions to individual differences in personality. Based on the existing body of research, we hypothesize that there will be a statistically significant genetic effect on individual dif-

ferences in personality. Our second problem was to test for potential moderator effect of study design on heritability estimate. Because primary research using family study and adoption study designs consistently result in lower estimates of genetic contributions to individual differences in personality compared to twin study design, we hypothesize that study design will be a statistically significant moderator of heritability estimates in the stated direction. Our third problem was to test for potential moderator effect of the type of personality model on heritability estimate. Personality is a latent construct conceptualized in somewhat different ways depending on the theory or model, but because there is no firm theoretical framework from which to extrapolate hypotheses of differential heritability, we will use null hypothesis. Our forth problem was to test for potential moderator effect of gender on average heritability estimate. Because primary research and available literature showed no consistent evidence of gender based differential heritability, we will use null hypothesis. The final problem was to test for potential moderator effect of the type of personality model on heritability estimate of two personality traits included in some form in different personality models—neuroticism and extraversion. Neuroticism and extraversion are two personality traits included in almost every personality trait model or theory, but depending on the model they capture somewhat different aspects of the latent constructs. Because we do not have a firm theoretical framework from which to extrapolate hypotheses of differential heritability, we use null hypothesis both for moderating effect of personality model on neuroticism and on extraversion.

## Method

### Sample of Studies

We used multiple methods to obtain relevant research for inclusion in the current study. First, computerized database searches of PsycINFO, ESCOhost, and ERIC were used to generate a pool of potential articles. The following search terms were used: *personality, heritability, familial aggregation, twin study, family study, adoption study, five-factor model, FFM, Big Five, B5, neuroticism, emotional stability, extraversion, agreeableness, consciousness, openness to experience, psychoticism, positive affect, negative affect, constraint, Cattell, Eysenck,* and *Tellegen*. Second, we reviewed all available hard copy issues of scientific journals covering the field of behavior genetics: *Behavior Genetics, Twin Research,* and *Twin Research and Human Genetics*. Third, we reviewed all online copies of the journal *Twin Research and Human Genetics* via electronic repository of Cambridge University Press, as well as online copies of the journal *Behavior Genetics*. Fourth, we reviewed references from every located primary study. Fifth, we reviewed official web pages (if available) of all first authors of the primary studies considered for inclusion in the meta-analysis. Sixth, via e-mail we contacted first authors of primary studies, who did not provide sufficient information for coding and were based on a sample not used in any study included in the meta-analysis. Seventh, we reviewed all the studies included in the Johnson et al. (2008) review.

Finally, we used three strategies to overcome file-drawer effects. Both PsycINFO and ERIC index dissertations and other unpublished work, which thus were captured in the literature search. Also, we reviewed all available abstracts from conferences organized by the Behavior Genetics Association, European Association of Personality Psychology, and the International Society for the Study of Individual Differences. After our reference list of all studies considered for the meta-analysis was formed, it was placed on a protected web page on July 13, 2013, and all members of the Behavior Genetics Association were contacted via e-mail with a request to check the reference list at the provided link and to (i) suggest additional studies that may have been overlooked in the literature search, and (ii) to supply their unpublished data that met the inclusion criteria provided in the e-mail. The web page with the reference list was available until September 1, 2013. Fourteen researchers[1] contacted Tena Vukasović, 10 recommended published articles and book chapters, and four offered their unpublished data. All recommendations were considered for inclusion, but all met one of the exclusion criteria (i.e., published after December 31, 2010, used a sample already included in the meta-analysis). Authors who offered their unpublished data were contacted for additional information. One data set did not meet the inclusion criteria of personality model used (Finland–Framingham personality measurement, Type A personality), while three unpublished data sets were included in the meta-analysis (Germany–the Jena Twin Study of Social Attitudes [JeTSSA]; The Netherlands–Twin Registry; United States–the Vietnam Era Twin Study of Aging [VETSA]).

### Selection Criteria

Inclusion and exclusion criteria were defined to specify the targeted population of primary studies relevant for this meta-analysis. To be included studies had to meet all seven inclusion criteria, and were not allowed to meet any of the eight exclusion criteria. Inclusion criteria were the following:

1. Primary study used one of the following personality models: Cattell, Eysenck, Tellegen, or five-factor model.
2. Primary study reported self-reported data for broad personality domains.
3. Primary study reported statistical indicators essential for effect size calculations: intraclass correlations (twin study), parent–offspring correlations or regression of offspring on midparent score (family study), biological parent–offspring correlation (adoption study).
4. Primary study reported standard error/standard deviation (or variance) of the effect size.
5. Primary study used one of the behavior genetic study designs: twin study, family study, adoption study.

[1] We would like to take this opportunity to thank all the colleagues who helped us by providing their insight, especially those allowing us to include their unpublished data: P. Merjonen (personal communication, July 22, 2013): Framingham personality measurement; D. Boomsma and S. Franjić (personal communication, July 24, 2013): The Netherlands Twin Registry; C. E. Franz (personal communication, July 24, 2013): The Vietnam Era Twin Study of Aging; C. Kandler (personal communication, August 6, 2013): domain level intraclass correlations for JeTSSA sample not reported in Kandler, Riemann, Spinath, and Angleitner (2010).

6. Primary study reported basic sample information (number of participants, gender structure).

7. For published primary studies: published prior to January 1, 2011.

First seven exclusion criteria were opposite of the stated inclusion criteria. The final exclusion criterion was,

8. Samples included in the meta-analysis had to be independent.

   8.1. In case there were multiple publications based on a single sample or longitudinal data in the same publication, the same sample was not included twice. The sample that was genetically most informative, either (i) used rare study design or rare biological relatives or (ii) had a larger sample and thus a more precise estimate, was included.

   8.2. In case there was more than one effect size in the same study (e.g., intraclass correlations from two different personality questionnaires), the effect size based on the psychometrically more reliable measure was included.

In all, 134 published primary studies met the inclusion criteria based on the content of their abstracts.[2] A pdf or a hard copy of each of these publications was obtained for coding. These publications were then examined by Tena Vukasović to determine whether they presented sufficient statistics for an effect size calculation. This procedure, together with six unpublished studies, led to a sample of 190 potentially independent effect sizes (illustration of the selection process is provided in Figure 1). If primary studies were deemed eligible, but did not provide sufficient information for coding and were based on a sample not used in any study included in the meta-analysis, we contacted the authors for additional information via e-mail. We contacted the first author of three articles and of these one responded stating that he no longer had access to the data (Oniszczenko & Jakubowska, 2005). Consequently all three primary studies were not included in the meta-analysis (Distel et al., 2009; Jang et al., 2006; Oniszczenko & Jakubowska, 2005).

Literature search and review procedures led to a final sample of 62 independent effect sizes from 39 published primary studies (13 contributing multiple independent effect sizes) and six unpublished studies. The structure of 128 effect sizes excluded from the meta-analysis was as follows: 16 effect sizes were based on a personality model not specified in the inclusion criteria, 46 were based on a sample already included in the meta-analysis, 60 did not provide sufficient information for calculating an effect size (57 of which also were based on a sample already included in the meta-analysis), two used genetically noninformative sample, one reported data only for lower level of personality, and three were not primary studies, but review papers (for characteristics of each potentially independent effect size considered for inclusion in the meta-analysis see online supplemental material, Appendix B). The problem of dependent samples used in multiple publications is due to many longitudinal studies (e.g., twin registers, national volunteer samples) that reported personality data collected on (mostly) same samples at different time points. Effect sizes included in this meta-analysis were based on self-reported personality estimates of 113,452 participants. See Table 1 for a summary of all studies included in the meta-analysis.

## A Description of Personality Heritability Studies Included in This Meta-Analysis

As shown in Table 1, the final sample of primary studies included in this meta-analysis consists of different study designs and different personality models, but none of the primary studies using Cattell (1943) personality model satisfied inclusion criteria without satisfying one of the exclusion criteria. From this point forward, we analyze and discuss three personality models included in the meta-analysis: Eysenck's (1970), Tellegen's (1982), and the five-factor model (FFM) of personality.

A more detailed structure of effect sizes included in the sample is presented in Table 2. As can be seen, different study designs are not equally represented, with twin studies overrepresented compared to family and adoption studies.

This is not surprising if we take into account that there is a number of twin registers around the world. On the other hand, adoption studies are an extremely valuable type of behavior genetic design, but are very rare. The effect sizes of different personality models also are unequally represented, with Eysenck's (1970) model being overrepresented compared to Tellegen's (1982) and FFM. Again, this finding reflects the use of personality measures in (most) twin registers around the world, where personality questionnaires developed from Eysenck's personality framework were used in most of the early twin studies (e.g., United Kingdom, Australia, The Netherlands), while Tellegen's model has almost exclusively been used in Minnesota. The FFM has become the dominant personality framework in the past two decades and as such has been included in most of the registers, projects, and individual studies.

Basic structure of primary studies included in this meta-analysis, based on gender, age, country of origin, and year of publication are presented. Males and females are equally represented with 35 effect sizes based on male and female data combined, 14 on male samples and 13 on female samples. Average age of participants is 31, with the age range from 9 to 92 years. Participants included in the final sample come from four continents (Asia, Australia, Europe, North America), and 12 countries (United States, $k = 18$; Australia, $k = 10$; United Kingdom, $k = 7$; The Netherlands, $k = 6$; Croatia, $k = 6$; Sweden, $k = 3$; Finland, $k = 2$; Norway, $k = 2$; Germany, $k = 2$; Canada, $k = 1$; South Korea, $k = 1$; Russia, $k = 1$; and 3 international online studies). Even though this sample represents all the available data, obviously there is a large gap in our knowledge of personality heritability in Africa and South America. There are new projects and studies being developed in Asia, but at the time most were not focused on personality as a phenotype. The oldest study included in the sample was published in 1951, 26% of the sample was published between 1951

[2] See online supplemental material for the reference list of all published studies considered for the meta-analysis (Appendix A).

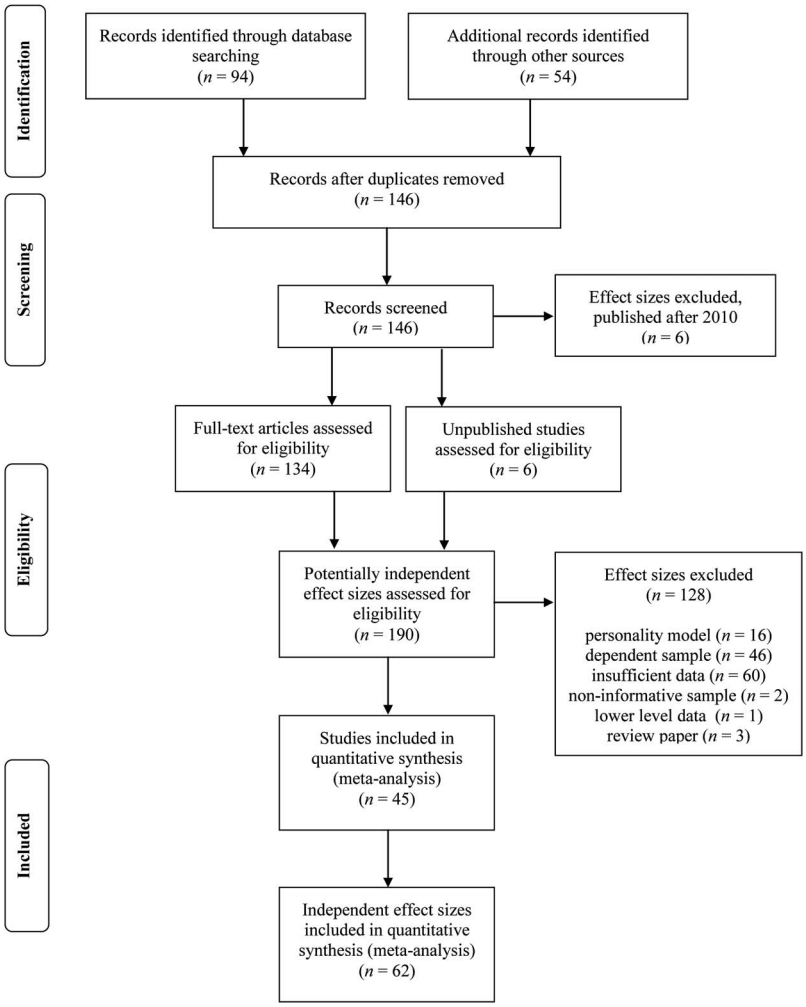


*Figure 1.* Flowchart illustrating the selection process.

and 1990, 24% between 1991 and 2000, 40% between 2001 and 2009, and 10% unpublished.

## Coding the Studies

During the coding of primary studies (from October 2011 to August, 2013), we recorded information about a range of characteristics of the publication, sample, study design, methodology, effect size, and coding process. These characteristics included the following: (a) authors, title, type of publication, name of publication, language, country, project; (b) sampling procedure, overall number of participants, number of subgroups of participants (monozygotic [MZ] twin pairs, dizygotic [DZ] twin pairs, parent–offspring pairs, mother–offspring pairs, father–offspring pairs, biological parent–adopted offspring pairs), average age, age range; (c) behavior genetic study design, personality model, personality questionnaire, study quality; (d) average effect size calculation, average effect size standard error of estimate, average effect size sample size, and average effect size reliability coefficient; effect size calculation for each personality trait, effect size standard error of estimate for each personality trait, effect size sample size for each personality trait, and effect size reliability coefficient for each personality trait; and (e) coder, date of coding, duration of coding, comments. Our intention in coding this information was threefold. The first was to collect descriptive information about the type of samples, study designs, and personality models being used in behavior genetic studies of personality heritability. The second was to collect information about the samples and data collection procedures used, as well as information about the authors and research projects responsible for the study, to identify independent samples. The third was to examine the impact of some of these variables through moderator analysis.

We took several steps to ensure that coding was reliable. After developing coding sheets and coding manual, and establishing the coding procedures all coders took part in a coding training to make sure the coding manual was used consistently. As part of the training all five coders had to familiarize themselves with coding sheets and manual, and independently code five studies that differed in study design and personality model.

Table 1
*Study Design, Personality Model, Gender, Age, and Country of Origin of Participants, Average Personality Heritability Effect Size Estimate, Standard Error of Estimate, and Units of Analysis (n)*

| No. | Authors | Design | Model | Gender | Age | Country | *ES* | *SE* | *n* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Tellegen et al. (1988) | TRT | Tellegen | M/F | 22.2 | United States (MN) | .59 | .044 | 217 |
| 2 | Floderus-Myrhed, Pedersen, & Rasmuson (1980) | TRT | Eysenck | M | 32.0 | Sweden | .47 | .016 | 2,277 |
| 3 | Floderus-Myrhed et al. (1980) | TRT | Eysenck | F | 32.0 | Sweden | .54 | .014 | 2,717 |
| 4 | Hur (2007) | TRT | Eysenck | M/F | 17.5 | South Korea | .42 | .036 | 531 |
| 5 | Saudino, Pedersen, Lichtenstein, McClearn, & Plomin (1997) | TRT | Eysenck | M/F | 42.2 | Russia | .55 | .081 | 79 |
| 6 | Rose, Koskenvuo, Kaprio, & Sarna, & Langinvainio (1988) | TRT | Eysenck | M | 37.0 | Finland | .40 | .026 | 1,027 |
| 7 | Rose et al. (1988) | TRT | Eysenck | F | 37.0 | Finland | .46 | .022 | 1,293 |
| 8 | Baker & Daniels (1990) | TRT | Eysenck | M/F | 35.2 | United States (CA) | .44 | .378 | 103 |
| 9 | Jang, Livesley, & Vernon (1996) | TRT | FFM | M/F | 31.2 | Canada | .47 | .070 | 123 |
| 10 | Riemann, Angleitner, & Strelau (1997) | TRT | FFM | M/F | 33.0 | Germany (BiLSAT) | .52 | .028 | 660 |
| 11 | Loehlin, McCrae, Costa, & John (1998) | TRT | FFM | M/F | — | United States (NMTS) | .44 | .122 | 807 |
| 12 | Johnson, McGue, & Krueger (2005) | TRT | Tellegen | M/F | 59.4 | United States (MN) | .46 | .163 | 467 |
| 13 | Losoya, Callor, Rowe, & Goldsmith (1997) | TRT | FFM | M/F | 34.5 | United States (OR) | .48 | .115 | 45 |
| 14 | Lensvelt-Mulders & Hettema (2001) | TRT | FFM | M/F | 31.5 | Netherlands | .44 | .355 | 100 |
| 15 | Loehlin & Martin (2001) | TRT | Eysenck | M | 23.2 | Australia | .37 | .054 | 255 |
| 16 | Loehlin & Martin (2001) | TRT | Eysenck | M | 37.2 | Australia | .42 | .046 | 327 |
| 17 | Loehlin & Martin (2001) | TRT | Eysenck | M | 61.4 | Australia | .36 | .273 | 208 |
| 18 | Loehlin & Martin (2001) | TRT | Eysenck | F | 23.2 | Australia | .40 | .039 | 453 |
| 19 | Loehlin & Martin (2001) | TRT | Eysenck | F | 37.8 | Australia | .40 | .112 | 1,105 |
| 20 | Loehlin & Martin (2001) | TRT | Eysenck | F | 61.2 | Australia | .41 | .038 | 470 |
| 21 | Gillespie, Evans, Wright, & Martin (2004) | TRT | Eysenck | M | 12.0 | Australia | .48 | .221 | 240 |
| 22 | Gillespie et al. (2004) | TRT | Eysenck | F | 12.0 | Australia | .41 | .073 | 129 |
| 23 | South, Krueger, Johnson, & Iacono (2008) | TRT | Tellegen | M/F | 17.8 | United States (MN) | .52 | .026 | 798 |
| 24 | Weiss, Bates, & Luciano (2008) | TRT | FFM | M/F | 44.9 | United States (MIDUS) | .45 | .042 | 365 |
| 25 | Vernon, Martin, Schermer, & Mackie (2008) | TRT | FFM | M/F | 41.2 | United States/Canada | .54 | .041 | 300 |
| 26 | Kendler, Myers, Potter, & Opalesky (2009) | TRT | FFM | M/F | — | International (web) | .45 | .038 | 441 |
| 27 | Bratko & Butković (2007) | TRT | Eysenck | M/F | 17.3 | Croatia | .47 | .090 | 75 |
| 28 | Rettew et al. (2006)[a] | TRT | Eysenck | M | 15.5 | Netherlands | — | — | 517 |
| 29 | Rettew et al. (2006)[a] | TRT | Eysenck | F | 15.5 | Netherlands | — | — | 639 |
| 30 | Eysenck & Prell (1951)[a] | TRT | Eysenck | M/F | — | United Kingdom | — | — | 25 |
| 31 | Eysenck (1956)[a] | TRT | Eysenck | M/F | 13.8 | United Kingdom | — | — | 26 |
| 32 | Valera & Berenbaum (2001) | TRT | Eysenck | M/F | 23.6 | United States | .60 | .095 | 45 |
| 33 | Franz (unpublished) | TRT | Tellegen | M/F | 55.4 | United States (VETSA) | .42 | .145 | 603 |
| 34 | Boomsma & Franić (unpublished) | TRT | FFM | M/F | 32.3 | Netherlands | .46 | .020 | 1,528 |
| 35 | Kandler (unpublished) | TRT | FFM | M/F | 31.5 | Germany (JeTSSA) | .54 | .188 | 316 |
| 36 | Eaves, Heath, Neale, Hewitt, & Martin (1998) | TFS | Eysenck | M | 53.0 | United States (VA) | .40 | .033 | 646 |
| 37 | Eaves et al. (1998) | TFS | Eysenck | F | 53.0 | United States (VA) | .45 | .021 | 1,418 |
| 38 | Finkel & McGue (1997) | TFS | Tellegen | M | 37.8 | United States (MN) | .52 | .049 | 220 |
| 39 | Finkel & McGue (1997) | TFS | Tellegen | F | 37.8 | United States (MN) | .52 | .036 | 406 |
| 40 | Rettew, Rebollo-Mesa, Hudziak, Willemsen, & Boomsma (2008)[a] | TFS | Eysenck | M | 15.5 | Netherlands | — | — | 291 |
| 41 | Rettew et al. (2008)[a] | TFS | Eysenck | F | 15.5 | Netherlands | — | — | 403 |
| 42 | Pincombe, Luciano, Martin, & Wright (2007)[a] | TFS | FFM | M/F | 20.2 | Australia | — | — | 103 |
| 43 | Tellegen et al. (1988) | TRA | Tellegen | M/F | 40.7 | United States (MN) | .52 | .110 | 44 |
| 44 | Shields (1958) | TRA | Eysenck | M/F | — | United Kingdom | .56 | .111 | 38 |
| 45 | Pedersen, Plomin, McClearn, & Friberg (1988) | TRA | Eysenck | M/F | 58.6 | Sweden | .28 | .095 | 95 |
| 46 | Vukasović, Bratko, & Butković (2009) | FS | FFM | M/F | 16.8 | Croatia | .28 | .090 | 115 |
| 47 | Bratko et al. (unpublished) | FS | FFM | M/F | 12.3 | Croatia | .09 | .043 | 539 |
| 48 | Bratko et al. (unpublished) | FS | FFM | M/F | 16.4 | Croatia | .15 | .082 | 148 |
| 49 | Bratko et al. (unpublished) | FS | FFM | M/F | 21.1 | Croatia | .25 | .074 | 171 |
| 50 | Tambs, Sundet, Eaves, Solaas, & Berg (1991) | FS | Eysenck | M | 21.9 | Norway | .22 | .093 | 118 |
| 51 | Tambs et al. (1991) | FS | Eysenck | F | 21.9 | Norway | .34 | .078 | 147 |
| 52 | Ahern, Johnson, Wilson, McClearn, & Vandenberg (1982) | FS | Eysenck | M/F | 16.9 | United States (HI) | .05 | .049 | 415 |
| 53 | Insel (1974) | FS | Eysenck | M | 17.0 | United Kingdom | .04 | .104 | 94 |
| 54 | Wray, Birley, Sullivan, Visscher, & Martin (2007)[a] | FS | Eysenck | M/F | 35.1 | Australia | — | — | 4,253 |
| 55 | Kendler et al. (2009) | FS | FFM | M/F | — | International (web) | .28 | .068 | 201 |
| 56 | Coppen, Cowie, & Slater (1965) | FS | Eysenck | M | 37.8 | United Kingdom | .56 | .116 | 53 |
| 57 | Coppen et al. (1965) | FS | Eysenck | F | 35.9 | United Kingdom | .24 | .111 | 79 |
| 58 | Eaves et al. (1999) | FS | Eysenck | M | — | United States (VA) | .22 | .021 | 2,218 |
| 59 | Eaves et al. (1999) | FS | Eysenck | F | — | United States (VA) | .28 | .014 | 4,542 |

Table 1 (*continued*)

| No. | Authors | Design | Model | Gender | Age | Country | *ES* | *SE* | *n* |
|---|---|---|---|---|---|---|---|---|---|
| 60 | Martin et al. (2000) | FS | Eysenck | M/F | 40.3 | United Kingdom | .30 | .007 | 20,427 |
| 61 | Bratko & Marušić (1997) | FS | FFM | M/F | 18.0 | Croatia | .14 | .092 | 118 |
| 62 | Scarr, Webber, Weinberg, & Wittig (1981) | AS | Eysenck | M/F | 18.5 | United States | .23 | .092 | 113 |

*Note*. Design = type of behavior genetic study design; model = type of personality model; *ES* = average effect size defined as personality heritability estimate; *SE* = standard error of the effect size; TRT = twins reared together; Eysenck = Eysenck's personality model; M = male; F = female; TFS = twins reared together and their family members; TRA = twins reared apart; FFM = five-factor model of personality; Tellegen = Tellegen's personality model; FS = family study; AS = adoption study; BiLSAT = Bielefeld Longitudinal Study of Adult Twins; NMTS = National Merit Twin Study; MIDUS = National Survey of Midlife Development in the United States; VETSA = Vietnam Era Twin Study of Aging; JeTSSA = Jena Twin Study of Social Attitudes.

[a] Primary studies reporting data for only one personality trait so an average *ES* was not calculated.



Afterward coders discussed and compared coding of each study. Meetings of the coding team were held monthly to calculate interrater reliability coefficients and discuss coding process for each study.

All published and three unpublished primary studies were coded by Tena Vukasović and by two additional independent coders,[3] which resulted in 95% of the sample being triple coded by independent coders. Interrater agreement was calculated to ensure the reliability of the coding process. Krippendorff's alpha (Hayes & Krippendorff, 2007), an interrater reliability coefficient, calculated for coding of effect sizes and standard errors indicated a very high level of coding reliability (for effect sizes = .97; for standard errors = .90).

## Effect Size Calculation

We developed coding procedures with appropriate formulas for twin studies, family studies, and adoption studies to parsimoniously reflect heritability estimates from different study designs. In studies of twins reared together, the procedure included several steps and was as follows:

1. Intraclass correlations for each personality trait were transformed into Fisher's $r_z$.
2. Average[4] $r_z$ for MZ and DZ twins was calculated.
3. Average $r_z$ for MZ and DZ twins was transformed into intraclass correlations.

   3.1. If the average MZ and DZ intraclass correlations did not suggest nonadditive genetic contributions ($r_{MZ} < 2r_{DZ}$), Falconer's formula $h^2 = 2(r_{MZ} - r_{DZ})$ was used to calculate the heritability estimate. Standard error for Falconer's heritability estimate was calculated as

$$SE(h^2) = 2\sqrt{\frac{(1 - r_{MZ}^2)^2}{n_{MZ}} + \frac{(1 - r_{DZ}^2)^2}{n_{DZ}}}$$

   (Floderus-Myrhed, Pedersen, & Rasmuson, 1980)

   3.2. If the average MZ and DZ intraclass correlations did suggest nonadditive genetic contributions ($r_{MZ} > 2r_{DZ}$), the MZ intraclass correlation was noted as the heritability estimate. Standard error for intraclass correlations was calculated as

Table 2
*Number of Independent Effect Sizes Included in the Meta-Analysis by Five Study Designs and Four Personality Models*

| Study design | Personality model | | | | |
|---|---|---|---|---|---|
| | Cattell | Eysenck | Tellegen | FFM | Total |
| TRT | 0 | 21 | 4 | 10 | 35 |
| TFS | 0 | 4 | 2 | 1 | 7 |
| TRA | 0 | 2 | 1 | 0 | 3 |
| FS | 0 | 10 | 0 | 6 | 16 |
| AS | 0 | 1 | 0 | 0 | 1 |
| Total | 0 | 38 | 7 | 17 | 62 |

*Note*. There were eight primary studies (4 TRT, 3 TFS, 1 FS) reporting data for only one personality trait so an average effect size was not calculated (reporting data only for neuroticism: Eysenck & Prell (1951); Rettew et al. (2006, separately for males and for females); Wray, Birley, Sullivan, Visscher, & Martin (2007); reporting data only for extraversion: Eysenck (1956); Pincombe, Luciano, Martin, & Wright (2007); Rettew et al. (2008, separately for males and for females). Cattell = Cattell's personality model; Tellegen = Tellegen's personality model; Eysenck = Eysenck's personality model; FFM = five-factor model of personality; TRT = twins reared together; TFS = twins reared together and their family members; TRA = twins reared apart; FS = family study; AS = adoption study.

[3] Four coders were selected as the best psychology graduate students in their class. They underwent an introduction course in meta-analysis organized by Tena Vukasović, and passed the coding training before being included in the coding process.

[4] We want to emphasize that the provided formulas for computing standard errors are appropriate for single traits. However, the computation of standard errors for the heritability estimates when averaging multiple intraclass correlations (or parent–offspring correlations) may result in overestimated variance of average intraclass correlation. Because multiple intraclass correlations for different traits based on the same group of twins are not independent, the correct computation of the variance is difficult and requires computing the covariance/correlation between the estimated intraclass correlations. We are not aware of an equation to obtain such an estimate, but to minimize potential bias in our results we used bootstrapping (R 3.1.2. package boot) to obtain *p* values and confidence intervals that should be at least asymptotically correct. Bootstrap confidence interval calculations ($CI_{normal}$ [.32, .47], $CI_{basic}$ [.32, .47], $CI_{studentized}$ [.35, .43], $CI_{percentile}$ [.32, .47], $CI_{knha}$ [.32, .47]) based on 10,000 bootstrap samples indicate that our original results [.35, .43] did not result in too wide confidence intervals.

$$SE = \sqrt{\frac{(1 - r^2)^2}{n}}$$ (Rose, Koskenvuo, Kaprio, Sarna, & Langinvainio, 1988)

In studies of twins reared apart, the procedure included several steps and was as follows:

1. MZ intraclass correlations for each personality trait were transformed into Fisher's $r_z$.
2. Average $r_z$ for MZ twins was calculated.
3. Average $r_z$ for MZ twins was transformed into intraclass correlation and noted as the heritability estimate. Standard error for intraclass correlations was calculated.

In family and adoption studies, the procedure included several steps and was as follows:

1. Regression of offspring on midparent score coefficients (or biological parent–offspring correlation coefficients) for each personality trait were transformed into Fisher's $r_z$.
2. Average $r_z$ was calculated.
   2.1. Average $r_z$ was transformed into regression coefficient and noted as the heritability estimate. Standard error for Pearson's $r$ was calculated as:

$$SE = \frac{\sqrt{1 - r^2}}{\sqrt{n - 2}}.$$

   2.2. Average $r_z$ was transformed into correlation coefficient, multiplied by two for familial aggregation coefficient and noted as the estimate of the upper level of heritability. Standard error for Pearson's $r$ was calculated.

## Statistical Analyses

We assumed that our effect sizes were sampled from a universe of possible sample sizes based on the postulate of quantitative genetic theory stating that heritability is a population parameter. Therefore, we conducted a random-effects model, as it assumes that sample effect sizes differ from the population effect size by sampling error plus random variability among studies. We conducted all analyses in R 3.1.2. package metafor (Viechtbauer, 2010).

## Main Analyses

We tested the first hypothesis that there will be a statistically significant genetic effect on individual differences in personality. To do this a meta-analysis was conducted for average personality heritability estimates and additional 11 analyses were conducted for each personality trait in three personality models included in this study (Eysenck's, Tellegen's, FFM) separately. In each analysis, only independent effect sizes were included. For example, in a primary study using the FFM an average personality heritability estimate was calculated following the steps described earlier. This estimate was used in the first meta-analysis. Heritability estimates for each of the five personality traits included in the FFM were used in five separate meta-analyses (e.g., heritability estimate for agreeableness was used in a meta-analysis of agreeableness, while heritability estimate for conscientiousness was used in a meta-analysis of conscientiousness). Following these analyses, we were able to estimate the average personality heritability, as well as heritability estimates for 11 different personality traits.

## Moderator Analyses

We tested the second hypothesis that study design would be a statistically significant moderator of heritability estimates with family and adoption study designs having lower estimates compared to twin study design, and the third hypothesis that there would be no evidence of differential heritability depending on the personality model used. We also tested the fourth hypothesis that there would be no evidence of differential heritability based on gender. These analyses were performed on average personality heritability estimates. To better understand which moderator had the strongest effect when controlling for the effects of the other moderators, we first tested each of them separately and then simultaneously. Finally, we tested the fifth hypothesis that there would be no evidence of differential heritability of neuroticism and of extraversion depending on the personality model used. To do this two additional meta-analyses were conducted, one for neuroticism and another for extraversion heritability estimates.

# Results

The analyses on the heritability of personality involved 62 independent effect sizes from 45 primary studies. Eight primary studies reported data for only one personality trait (neuroticism or extraversion), so in this case an average effect size was not calculated, but the effect sizes for individual traits were included in the analyses for heritability of that trait.

## Mean Effect Size

**Hypothesis 1: Personality heritability.** As shown in Table 3, there is a statistically significant genetic effect on individual differences in personality, with an average heritability estimate of 0.39, $SE = 0.020$, 95% CI [.35, .43]. In other words, 39% of individual differences in personality are due to genetic and 61% to environmental effects. A graphical display of the result is shown as a forest plot in Figure 2. The forest plot indicates both the average heritability estimates (squares varying in proportion to the weight the individual study accorded in the analysis) with confidence intervals, as well as an estimate of the overall summary effect size at the bottom of the figure (a diamond at which the center represents the point estimate of summary effect size, and its width represents the limits of its 95% confidence interval).

The $Q$ statistic was statistically significant, $k = 54$, $Q(53) = 719.75$, $p < .001$, which suggests heterogeneity. We provide an additional indicator of the amount of heterogeneity $I^2$, which is the proportion of total variance that is due to heterogeneity rather than

Table 3
*Average Effect Size Estimate (ES), 95% Confidence Interval (CI), and Number of Effect Sizes (k) by Study Design and Personality Model*

| | Personality model | | | |
|---|---|---|---|---|
| Study design | Eysenck *ES*, 95% CI, *k* | Tellegen *ES*, 95% CI, *k* | FFM *ES*, 95% CI, *k* | Total *ES*, 95% CI, *k* |
| Twins | .44, [.42, .47], 21 | .53, [.49, .56], 7 | .48, [.45, .51], 10 | .47, [.45, .49], 38 |
| FS and AS | .24, [.17, .32], 10 | — | .19, [.11, .27], 6 | .22, [.17, .28], 16 |
| Total | .38, [.33, .42], 31 | .53, [.49, .56], 7 | .37, [.29, .45], 16 | .39, [.35, .43], 54 |

*Note*. Tellegen = Tellegen's personality model; Eysenck = Eysenck's personality model; FFM = five-factor model of personality; twins = twins reared together, twins reared together and their family members, and twins reared apart; FS = family study; AS = adoption study.



chance. Approximate guidelines for interpreting this statistic suggest that $I^2 = 75\%$ is considered a large heterogeneity (Higgins & Thompson, 2002). In this analysis $I^2 = 93.53\%$, suggesting that almost all of the variation in effect sizes is due to heterogeneity. Because both $Q$ and $I^2$ suggest heterogeneity, we continued with the planned moderator analyses.

To test the second part of our first hypothesis, additional 11 analyses were conducted for each personality trait in three personality models included in this study (Eysenck's, Tellegen's, FFM) separately. As shown in Table 4 there is a statistically significant genetic effect on individual differences in each of the 11 personality traits included in this study, with trait heritability estimates ranging from .30, 95% CI [.21, .38] for psychoticism in Eysenck's model, to .51, 95% CI [.44, .58] for positive emotionality and constraint in Tellegen's model.

**Publication bias.** Meta-analyses are susceptible to the so-called file-drawer problem (Rosenthal, 1979), by which published studies are more likely to be those that have found significant effects than those that have not. In a meta-analysis on the efficacy of psychological, educational, and behavioral treatments Lipsey and Wilson (1993) confirmed that oversampling of published studies in a meta-analysis does indeed upwardly bias treatment effect estimates. They showed that the mean effect size estimates for both published and unpublished

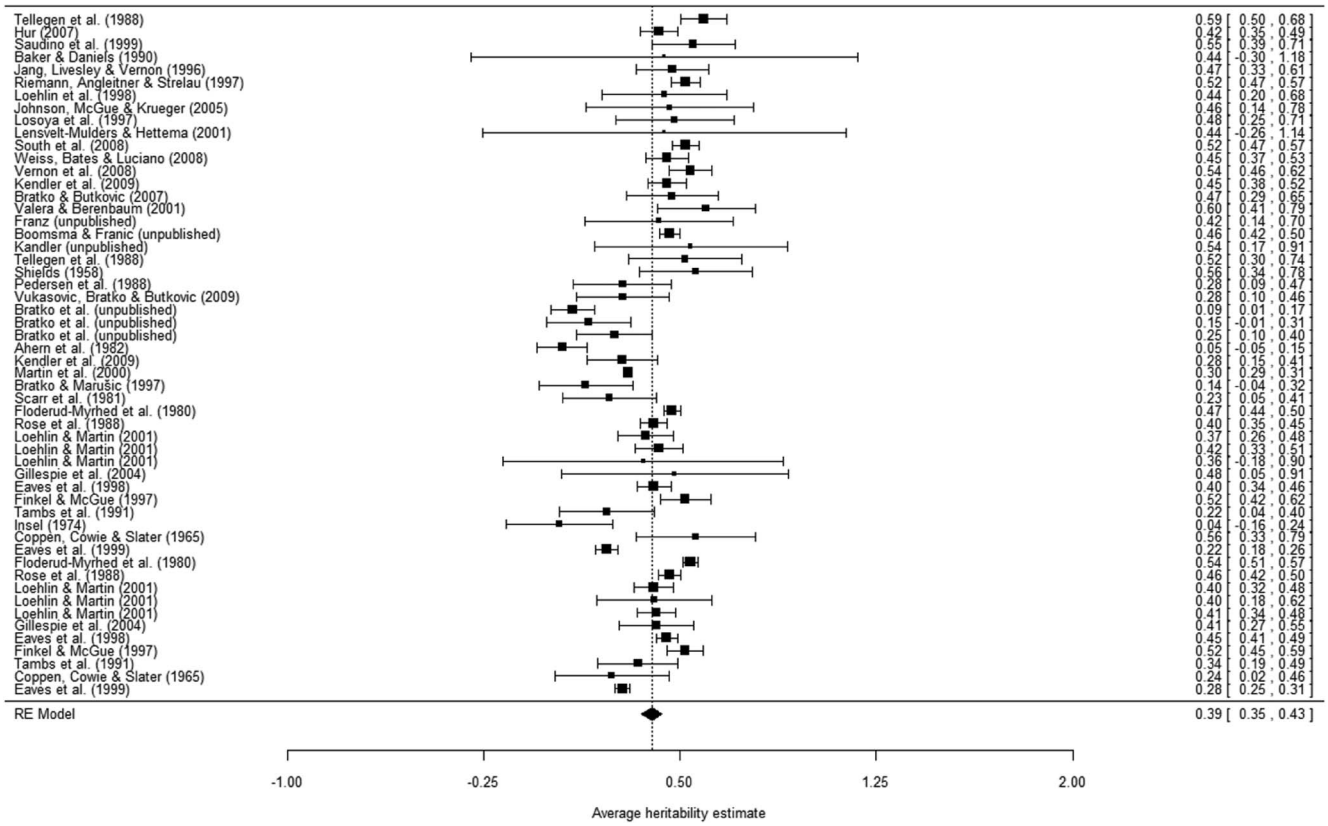


*Figure 2.* Forest plot for average personality heritability estimates ($k = 54$). RE = random-effects model.

Table 4
*Average Effect Size Estimate (ES), 95% Confidence Interval (CI), and Number of Effect Sizes (k) by Three Personality Models*

| Personality model | *ES* | *SE* | 95% CI | *k* |
|---|---|---|---|---|
| Eysenck | | | | |
| Neuroticism | .39 | .025 | [.34, .43] | 35 |
| Extraversion | .42 | .030 | [.37, .48] | 34 |
| Psychoticism | .30 | .044 | [.21, .38] | 17 |
| Tellegen | | | | |
| Negative emotionality | .47 | .021 | [.43, .51] | 7 |
| Positive emotionality | .51 | .036 | [.44, .58] | 7 |
| Constraint | .51 | .035 | [.44, .58] | 7 |
| Five-factor model | | | | |
| Neuroticism | .37 | .047 | [.28, .47] | 16 |
| Extraversion | .36 | .043 | [.28, .45] | 17 |
| Openness | .41 | .051 | [.31, .51] | 16 |
| Agreeableness | .35 | .035 | [.28, .42] | 16 |
| Conscientiousness | .31 | .045 | [.22, .40] | 16 |

*Note.* *SE* = standard error of the effect size.

studies of treatment efficacy fall in the positive range, but published studies have a higher effect size compared to unpublished studies (0.53 vs. 0.39).

We sought to address this issue by including unpublished studies of personality heritability. However, this did not preclude the possibility that substantial number of studies with null findings were left excluded. We therefore proceeded to empirically examine the presence of publication bias in four ways. First, we calculated fail-safe *N* for average personality heritability estimate based on two procedures: Rosenthal's and Orwin's. Based on Rosenthal's (1979) fail-safe *N* procedure, we would need additional 91,784 studies averaging null results that must be in the file drawers before the overall probability of a Type I error was brought to 5% level of significance, or in other words—for the result of our meta-analysis to become statistically insignificant. Based on Orwin's (1983) fail-safe *N* procedure, we would need additional 2,076 effect sizes with a null effect size for the result of our meta-analysis to become .01, or we would need additional 159 effect sizes with a null effect size for the result of our meta-analysis to become .10. This means that we would need to include more than twice as many effect sizes as in the present meta-analysis with an estimate of personality heritability of zero only to reduce our estimate from .39 to .10. Because such a large number of unpublished studies with null results can be considered unlikely, we concluded that the results of this meta-analysis are relatively robust to publication bias.

Second, as shown in Figure 3 we created a funnel plot for average personality heritability estimates, in which we plotted a measure of precision (standard error) by the effect size (average heritability estimate). In this type of graph, larger studies (i.e., more precise estimates) appear at the top and smaller studies appear at the bottom. When a plot has a funnel shape, as the sample size increases the studies converge more closely around the true mean (Cooper, Hedges, & Valentine, 2009). If the funnel plot is symmetric, with the left side of the graph being a mirror image of the right side of the graph, it is interpreted as suggesting that publication bias is unlikely to have extorted the results. Our funnel plot is generally symmetrical indicating that publication bias was unlikely to have influenced our findings. An atypical overrepresentation of larger studies including cohorts and national samples resulted in a nonfunnel shape but when study design was included as a moderator the studies converged more closely around the true mean.

Third, we conducted a trim and fill analysis, which is used to statistically evaluate the mirror symmetry of the funnel plot. This analysis includes four steps. First, effect sizes on the left and on the right side of the graph are compared and the ones not having a mirror image on the opposite side are temporarily removed, leaving the graph symmetrical. Second, removed effect sizes are returned to the analysis together with their simulated mirror images on the opposite side. Third, the corrected meta-analysis is performed on the sample including simulated effect sizes. Finally, the corrected results based on a simulated sample are compared to the results of the original meta-analysis. A large discrepancy is interpreted as suggesting that

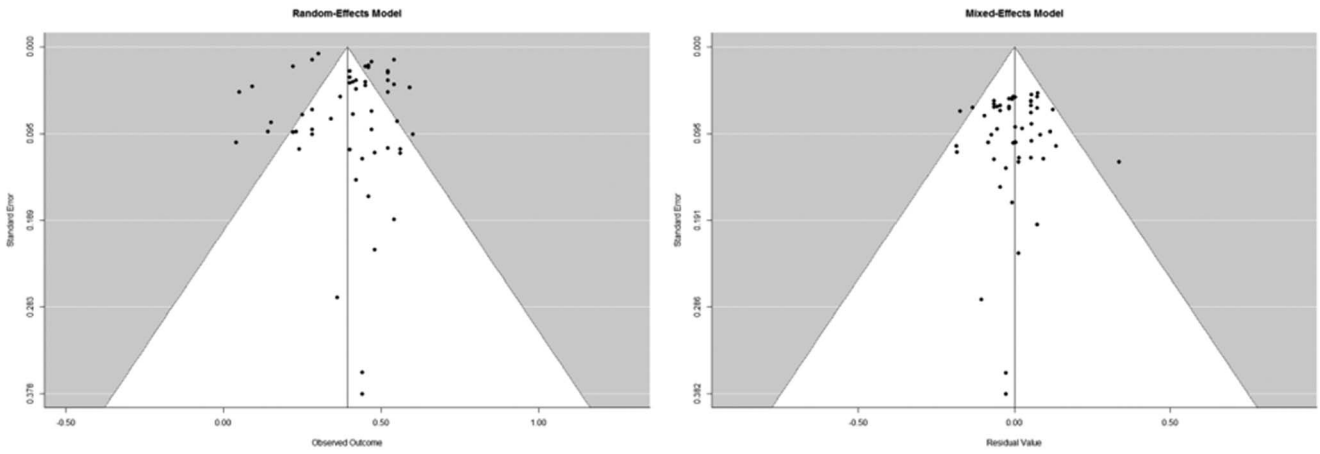


*Figure 3.* Funnel plot for a model without moderators (random-effects model) and a model with study design as moderator (mixed-effects model) for average personality heritability estimates ($k = 54$).

publication bias has likely extorted the results. In our sample trim and fill analysis indicated that there are seven "missing" effect sizes on the left side of the funnel plot and after including these simulated effect sizes in the sample the analysis indicated an average heritability estimate of .37, 95% CI [.33, .41]. In other words, the results of trim and fill analysis show that 37% of individual differences in personality are due to genetic, and 63% to environmental effects. This estimate is very similar to the original estimate of 39% with almost completely overlapping confidence intervals, suggesting that it is not likely that publication bias extorted our results.

**Sensitivity analysis.** We performed a sensitivity analysis on a subsample of studies to test how robust and stable our results are in terms of sample structure. Discrepancy between the sensitivity analysis results and the original results from the full sample is interpreted as an indicator of unstable results. We identified one potential outlier based on eight different influential case diagnostics[5] using the influence function of R 3.1.2 package metafor (Viechtbauer, 2010). The potential outlier was excluded from the sample and results of a repeated meta-analysis indicated an average personality heritability estimate of .40, 95% CI [.36, .44]. In other words, the sensitivity analysis suggests that 40% of individual differences in personality are due to genetic effects. This estimate is very similar to the original estimate of 39% with almost completely overlapping confidence intervals, indicating that our results are quite robust and stable, not affected by the change in the sample structure. After a closer examination of the potential outlier (Ahern, Johnson, Wilson, McClearn, & Vandenberg, 1982), we confirmed that this study was methodologically sound. The most probable reason for its identification as a potential outlier was its very small effect size. Finally, we concluded that there was no valid methodological or theoretical reason to exclude this study from the meta-analysis thus confirming the average personality heritability estimate of .39, which is based on a larger sample of methodologically sound primary studies compared to the sensitivity analysis.

**Multilevel model.** Data included in this meta-analysis have a two-level hierarchical structure with some studies providing more than one estimate of heritability. Even though we ensured no overlap in the data used to obtain the estimates, this did not exclude the possibility that multiple estimates from the same study used to compute one average heritability estimate per study, may be more similar to each other than estimates derived from different studies. To account for this we used a meta-analytic multilevel model with two random effects, at the study level and at the level of the estimates (Konstantopoulos, 2011; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). This analysis showed that the estimated variance at the study level (.0154) is larger than the estimated variance at the estimate level (.0011). Based on the estimated variance components we can compute an intraclass correlation for true effects coming from the same study (study variance/(study variance + estimate variance)), which turns out to be .93. It is obvious that the underlying true heritability values coming from the same study correlate very strongly with each other. We then tested whether the addition of the study level random effect significantly improves the model fit. The likelihood ratio test (LRT) was highly significant (LRT = 14.42; $df = 1$; $p < .001$) suggesting that such an approach should be used. If we take a closer look at the results from the original analysis, $h^2 = .3928$, 95% CI [.35, .43], $SE = .020$, $Q(53) = 719.75$, $p < .001$; and from the multilevel model analysis, $h^2 = .3997$, [.35, .45], $SE = .024$, $Q(53) = 840.56$, $p < .001$; we see that the heritability estimate changed by only .0069. In sum, even though the hierarchical multilevel model statistically fits the data better our initial results and conclusions are consistent.

## Moderator Analyses

**Hypothesis 2 and 3: Study design and personality models.** To test our second and third hypothesis we conducted moderator analyses, testing for potential moderator effect of study design and the type of personality model on average heritability estimates. We first tested study design and type of personality model separately and then simultaneously. Results showed that study design is a statistically significant moderator of heritability estimates in the expected direction, $Q\text{M}(1) = 98.31$, $p < .001$, $I^2 = 70.57\%$, $R^2 = 80.42\%$,[6] with family and adoption studies resulting in lower estimates of genetic contributions to individual differences in personality compared to twin studies (.22 vs. .47). The second moderator analysis showed that type of personality model was not a significant moderator of heritability estimates, $Q\text{M}(2) = 3.66$, $p = .06$, $I^2 = 93.06\%$, $R^2 = 6.37\%$, confirming our null-hypothesis and suggesting no evidence of differential heritability of personality. We then tested both moderators simultaneously to better understand which moderator had the strongest effect when controlling for the effects of the other. The results again showed that study design was a statistically significant moderator of heritability estimates in the expected direction, while the type of personality model was not, $Q\text{M}(3) = 102.49$, $p < .001$, $I^2 = 69.43\%$, $R^2 = 81.25\%$, $p_{\text{studydesign}} < .001$, $p_{\text{personalitymodel}} = .20$.

**Hypothesis 4: Gender.** We conducted a third moderator analysis to test our fourth hypothesis of potential moderator effect of gender on average personality heritability estimates. The structure of our final sample allowed us to test male ($k = 12$) versus female ($k = 11$) average heritability estimates. The moderator analysis showed that gender was not a significant moderator of heritability estimates, $Q\text{M}(1) = .80$, $p = .37$, $I^2 = 90.16\%$, $R^2 = 0\%$, confirming our null-hypothesis and suggesting no evidence of gender differential heritability of personality.

We finally tested all three moderators (study design, personality model, and gender) simultaneously to better understand which moderator had the strongest effect when controlling for the effects of the others. The results again showed that study design was the only statistically significant moderator of heritability estimates in the expected direction, $Q\text{M}(4) = 48.60$, $p < .001$, $I^2 = 57.73\%$, $R^2 = 82.89\%$, $p_{\text{studydesign}} < .001$, $p_{\text{personalitymodel}} = .07$, $p_{\text{gender}} = .13$.

[5] The influence function calculates the following leave-one-out diagnostics for each study: externally standardized residual, DFFITS value (indicating how many standard deviations the predicted [average] effect for the ith study changes after excluding the ith study from the model fitting), Cook's distance, covariance ratio, the leave-one-out amount of (residual) heterogeneity, the leave-one-out test statistic for the test of (residual) heterogeneity, DFBETAS value(s) (indicating how many standard deviations the estimated coefficient[s] change[s] after excluding the ith study from the model fitting).

[6] The amount of heterogeneity accounted for by the mixed-effects model.

**Hypothesis 5: Extraversion and neuroticism.** Two additional meta-analyses were conducted to test the fourth hypothesis, that there would be no evidence of differential heritability of neuroticism nor of extraversion depending on the personality model used. Average heritability estimate of neuroticism was .39 ($k = 58$; $SE = 0.021$, 95% CI [.35, .43]), and average heritability estimate of extraversion was .42 ($k = 58$; $SE = 0.023$, [.37, .46]). Moderator analyses showed that type of personality model was not a statistically significant moderator of heritability estimates of neuroticism, $QM(2) = 0.86$, $p = .36$, $I^2 = 94.63\%$, $R^2 = 0\%$; nor of extraversion, $QM(2) = 3.09$, $p = .08$, $I^2 = 95.21\%$, $R^2 = 6.49\%$; again confirming the null-hypothesis and suggesting no evidence of differential heritability.

## Discussion

In the present meta-analysis, we synthesized research on personality heritability from different behavior genetic study designs and different personality models. Analyses were based on the results of 45 primary studies with 62 independent effect sizes, six of which were unpublished. Because the sample in this meta-analysis included self-reports from more than 100,000 participants, from four continents and 12 countries, covering an age range from 9 to 92 years, these results have larger statistical power and can have wider generalization than any single primary study. Also, as far as we know this is the first time potential moderator effects of study design, type of personality model, and gender on personality heritability estimate were statistically tested.

### Average Personality Heritability

Overall, what is the average effect of genetic contributions to individual differences in personality? According to this meta-analysis, the answer is 40%. When one looks at several different publication bias indicators and the sensitivity analysis it seems that this estimate is largely unaffected by the file-drawer effect and is quite robust. This result is mostly in line with recent findings in the literature. Plomin et al. (2008) stated that results of behavior genetic studies converge on the conclusion that personality is substantially heritable, with estimates of genetic contributions to individual differences ranging between 30 and 50%. Johnson et al. (2008) conducted a review of more than 50 years of research in behavioral genetic studies of personality and the average heritability coefficient, calculated by the average MZT and DZT correlations and Falconer's formula, was 48%, while the average heritability estimate based on the average correlation of MZ twins reared apart was 31%. In the most recent effort to synthesize existing data on heritability of neuroticism and extraversion van den Berg et al. (2014) used a novel approach of item-response theory (IRT) for harmonization of neuroticism and extraversion phenotypes across nine different inventories and 23 cohorts in the Genetics of Personality Consortium. The IRT-based scores for neuroticism and extraversion based on six cohorts and more than 29,000 twin pairs were 48 and 49% heritable, respectively. Turkheimer, Pettersson, and Horn (2014) concluded in their review paper that personality is heritable and that wide traits (i.e., extraversion and neuroticism) are approximately 40% heritable. This conclusion is supported by our results of 40% average personality heritability estimate, and 39% and 42% heritability estimates of neuroticism and extraversion, respectively.

Our result of average personality heritability of 40% is somewhat lower than the estimates of 48% (Johnson et al., 2008), and 48 and 49% (van den Berg et al., 2014). It is interesting to notice that both higher estimates of 48 and 49% were based only on twin data, which we know result in higher estimates of genetic contributions to individual differences in personality compared to family and adoption studies, and are in line with our estimate of average personality heritability based on twin data (47%).

### Study Design as a Moderator of Personality Heritability

Moderator analyses showed that study design was a significant moderator of average personality heritability estimate with twin studies showing higher estimates compared to family and adoption studies (.47 vs. .22). Analyses also showed that personality model was not a significant moderator of average personality heritability estimate, nor was it a significant moderator of neuroticism and extraversion heritability estimates, two personality traits contained in some form in almost every personality theory and model. These conclusions are based on the empirical data available at this point in time. Some possible explanations of these findings are discussed.

For years behavior genetic studies have indicated the trend of lower heritability estimates from family and adoption studies compared to twin studies (e.g., Plomin, DeFries, Knopik, & Neiderhiser, 2013). This finding was surprising because familial aggregation estimate, a statistical indicator calculated by doubling the biological parent–offspring correlations from family and adoption studies should be the upper limit of heritability estimate. We know that family members living together share both additive genetic and shared environmental influences, which should make them more similar. If we double their correlation, we are doubling both the additive genetic and the shared environmental influences thus resulting in an upwardly bias estimate. How is it then possible that heritability estimates from twin studies are systematically larger? What effects could explain these findings? This question has been the focus of a scientific debate in which Plomin, Corley, Caspi, Fulker, and DeFries (1998) offered three possible substantive explanations: maturation effects, cohort effects, and nonadditive genetic effects. However, maturation effects (i.e., age differences between parents and offspring) and cohort effects seem to be less probable explanations, which leaves the third possibility—nonadditive genetic effects. Heritability estimates from twin studies include all additive as well as nonadditive genetic effects, whereas family and adoption studies include only additive genetic effects. A number of primary studies have confirmed the nonadditive genetic effect on individual differences in personality (e.g., Eaves et al., 1998, 1999; Finkel & McGue, 1997; Hur, 2007; Keller, Coventry, Heath, & Martin, 2005; Plomin et al., 1998). In summary, personality traits are heritable and nonadditive genetic effects probably have a role in explaining individual differences in personality.

### Differential Heritability of Personality

The question of differential personality heritability has been a matter of interest for more than three decades. After analyzing personality data of 850 twin pairs who completed the California

Psychological Inventory (CPI; Gough, 1956), and were included in the 1962 National Merit Scholarship Qualifying Test at the age of 16, Loehlin (1978) concluded there was no firm evidence of differential heritability of personality traits. Loehlin (1982) tackled this problem again using Swedish versions of Eysenck's neuroticism and extraversion scales in a sample of more than 13,000 twin pairs. Again, he concluded that there was no evidence of differential heritability between extraversion and neuroticism. However, in the same study he reanalyzed the data of the National Merit twin sample and derived seven orthogonal factor scales from the CPI item pool. Results indicated evidence of differential heritability of personality scales and Loehlin (1982) concluded that differential heritability of personality scales may be found if one employs sufficiently large samples and extends the search to personality dimensions independent of extraversion and neuroticism. Three decades after his first study on differential heritability of personality, Loehlin (2012) revisited this question using a similar approach but a considerably larger sample of 1,771 adult Australian twin pairs. Results indicated there was evidence of differential heritability for 11 clusters derived from Cloninger's and Eysenck's personality questionnaires items, but not for broad extraversion and neuroticism composites based on them. Again, he concluded that there is small variability in heritability estimates of wide personality dimensions, but the final conclusion of differential genetic and environmental structure of personality still has to be formulated. Turkheimer et al. (2014) also concluded in their review that personality is not differentially heritable.

Based on our results we must conclude that there is no evidence of differential heritability of personality from three personality models, as well as no evidence for differential heritability of neuroticism or extraversion between these models. We find that our results confirm Loehlin's (1978, 1982, 2012) findings of no differential heritability between wide personality dimensions, but still leave the question of differential heritability of personality traits on different levels of personality hierarchy open. Further research is needed before we can conclude if there is evidence of differential personality heritability at facet level, the hierarchically lower level of personality, or at the higher order factors of the Big Five labeled alpha and beta (Digman, 1997) or stability and plasticity (DeYoung, Peterson, & Higgins, 2002).

Results of our final moderator analysis indicate that there is also no evidence of differential personality heritability based on gender. This finding does not eliminate the possibility of difference in genetic architecture or gene expression in males and females. It only shows that the proportion of phenotypic differences among individuals, which can be attributed to genetic differences, is not statistically different for males and females.

After conducting 11 separate analyses for each personality trait in three personality models included in this study (Eysenck's, Tellegen's, FFM), we can see that trait heritability estimates range from .30 [.21, .38] for psychoticism in Eysenck's model, to .51 [.44, .58] for positive emotionality and constraint in Tellegen's model. We address possible reasons for these traits showing the highest and lowest heritability estimates. If we look at Table 3, we can see that all primary studies of Tellegen's model included in this meta-analysis are based on twin study design. There are publications reporting MPQ family data, but they only report kinship correlations for 11 primary scales and not for three higher order factors (e.g., Carey, 2002). As we demonstrated, twin studies result in statistically higher personality heritability estimates compared to family and adoption studies. We expect that in some future meta-analysis of personality heritability, after sufficient number of primary studies based on family and adoption study design using Tellegen's model of personality are conducted the estimates of positive emotionality, negative emotionality, and constraint would be somewhat lower.

To help us better understand lower heritability estimate of psychoticism, we turn to Heath and Martin's (1990) study of 2,903 adult same-sex Australian twin pairs. In this study the authors applied multivariate genetic analysis, a generalization of factor analysis and behavior genetic analysis, to responses to items of the psychoticism scale of the Eysenck Personality Questionnaire. Results indicated that the structure of genetic influences on the items of the psychoticism scale differed from the structure of environmental influences, thus confirming the etiologic heterogeneity of the psychoticism scale. The genetic correlation between suspiciousness items and items reflecting unconventional or tough-minded attitudes or hostility to others was negative, but the environmental correlation was positive. The authors concluded that very different genetic and environmental structures would be expected to produce low heritability estimates in some analyses. This may explain why the psychoticism scale has the lowest heritability estimate in our study.

Johnson et al. (2008) also estimated weighted mean broad sense heritability coefficients for core neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness based on data from different kinships. These estimates were 43, .54, .48, .49, and .47, respectively, which is somewhat higher compared to our heritability estimates of single traits. However, confidence intervals of heritability estimates from Johnson et al. (2008) and 11 single trait estimates from our study are overlapping.

To conclude, this meta-analysis confirmed Loehlin's (1978, 1982, 2012) findings of no evidence of differential heritability of personality at the personality hierarchy level of wide personality dimensions or traits.

## Methodological Implications, Limitations, and Future Directions

The present study provided an important empirical review of personality heritability in different types of study designs and in different types of personality models. The main aim of this meta-analysis was to systematize available findings in the field of personality heritability and test for possible moderator effects of study design, type of personality model, and gender on heritability estimates, but as in any systematic review we also identified certain blind spots in the field and are able to give guidance for future research directions.

There are at least four important and somewhat surprising insights that this systematic review has provided. First, in the final sample there are no effect sizes based on Cattell's personality model. This is in part due to our inclusion criteria, which did not include: personality scales of ego strength, super ego strength, and self-sentiment strength (Cattell, Rao, & Schuerger, 1985), spouse pairs (Nagoshi, Johnson, & Honbo, 1992), or parents' ratings of children (Loehlin, Horn, & Willerman, 1981). But it also reflects the situation in the literature in which there is an underrepresentation of Cattell's personality model in behavior genetic studies.

This may be considered a surprising finding for two reasons. First, Cattell's personality questionnaires were widely used in the past and included in some large projects at the time (e.g., Hawaii Family Study of Cognition, HFSC; Ahern et al., 1982). Second, Cattell was one of the pioneers of statistical and psychometric analyses in the field of behavior genetics with several publications exploring the advantages of the multiple abstract variance analysis (MAVA) over the twin method (e.g., Cattell, 1953, 1960). However, Cattell obviously was a pioneer, but it seems his personality model and questionnaires had been replaced with new models in most behavior genetic studies and projects.

A second surprising finding is the number of dependent and overlapping samples in different publications of twin registry samples over the years. Because our exclusion criteria stated that samples included in the meta-analysis had to be independent, we made an effort to identify all independent samples, subsamples, and effect sizes. This procedure identified 46 effect sizes that were dependent, and 57 additional effect sizes that did not state statistical indicators necessary for effect size calculations and were also at least partly dependent to samples already included in the meta-analysis. In summary, 103 out of 190 potentially independent effect sizes considered for this meta-analysis (54%) had to be excluded due to overlapping samples. This shows us the current state in behavior genetic personality research using specific personality models and samples, but it also shows there are valuable replications within each twin registry sample.

The third somewhat surprising finding is the uneven representation of behavior genetic study designs in the literature. The fact that there are many twin registers around the world explains the overrepresentation of studies based on twins reared together in our sample (56%). The organizational and ethical complexity of adoption studies also explains the underrepresentation of this design (1.6%) in our sample, as well as studies based on twins reared apart (5%). The surprising finding is the relative underrepresentation of family studies (26%), because this is one of the basic behavior genetic designs and is based on biological relatives that are easily attainable in any population (i.e., parents and their offspring). One probable reason for this underrepresentation is the development of new designs within the twin study design, which include twins reared together as well as their parents, siblings, children, and spouses, allowing for simultaneous use of different data sources. This type of study design is a combination of twin and family design and makes 11% of our sample. Due to genetic dependency between family members, we were limited to using data from only one type of relatives from this type of study design. Personality data from MZ and DZ twin pairs were chosen because it was based on larger sample size compared to other possible types of genetically related pairs. If we wanted to include parent–offspring correlations from the same study from which twin pair correlations were already included, we would have dependent effect sizes in our meta-analysis because "offspring" were actually the twins.

The fourth surprising finding is the uneven representation of populations and samples around the world. In the present meta-analysis most of the effect sizes are based on samples from United States ($k = 18$), Australia ($k = 10$), and western European countries such as United Kingdom ($k = 7$) and The Netherlands ($k = 6$). There is only one effect size from South Korea, and no effect sizes from any country from South America or Africa. This makes any future research of personality heritability from less studied populations extremely valuable.

However, it is also important to note that our review was limited in several ways. First, it was limited to studies employing behavior genetic design and only three personality models. This lead to small cell sizes in some of our analyses (e.g., zero primary studies using Tellegen's model of personality and family of adoption design). Second, our sample size was drastically downsized due to dependent and overlapping samples in different publications of twin registry samples over the years. This shows us the current state in behavior genetic personality research using specific personality models, but limits our power to generalize the results of this meta-analysis to all personality theories or models. Third, this study is limited to self-reported personality data and should be interpreted as such. More primary research is needed before a meta-analysis on different methods of personality assessment and personality heritability is undertaken. However, we believe that first such effort should be based on peer reports in not so distant future. For example, Riemann, Angleitner, and Strelau (1997) conducted a German twin study on a sample of 660 MZ and 304 DZ twin pairs comparing heritability estimates based on self- and peer report data. Heritability estimates based on self-report data were in line with our results for twin studies ranging from .42 to .56. Heritability estimates based on peer report data also suggest substantial heritability and were somewhat higher ranging from .57 to .81. However, peer report estimates were based on two raters and may not be directly comparable to self-reports. Fourth, almost all of the primary studies in the present meta-analysis consisted of participants from North America, Europe, and Australia with extreme under representation of Asian population, and no representation of South American and African populations. Future efforts should be made to acquire data from less studied parts of the world, and the easiest first step could be to include personality questionnaires in existing twin registers in Asia.

In summary, a valuable primary study of personality heritability in the future might try to use family or adoption study design, a new independent sample, a sample from Asia, Africa, or South America, a psychometrically validated personality questionnaire, and a peer report method of personality assessment.

## References

References marked with an asterisk indicate studies included in the meta-analysis.

*Ahern, F. M., Johnson, R. C., Wilson, J. R., McClearn, G. E., & Vandenberg, S. G. (1982). Family resemblances in personality. *Behavior Genetics, 12,* 261–280. http://dx.doi.org/10.1007/BF01067847

*Baker, L. A., & Daniels, D. (1990). Nonshared environmental influences and personality differences in adult twins. *Journal of Personality and Social Psychology, 58,* 103–110. http://dx.doi.org/10.1037/0022-3514.58.1.103

Bouchard, T. J., Jr. (1993). Genetic and environmental influences on adult personality: Evaluating the evidence. In I. Deary & J. Hettema (Eds.), *Basic issues in personality* (pp. 15–44). Dordrecht, The Netherlands: Kluwer Academic. http://dx.doi.org/10.1007/978-94-011-1660-2_2

Bouchard, T. J., Jr. (1997). The genetics of personality. In K. Blum & E. P. Noble (Eds.), *Handbook of psychiatric genetics* (pp. 273–296). Boca Raton, FL: CRC Press.

Bouchard, T. J., Jr., & Loehlin, J. C. (2001). Genes, evolution, and personality. *Behavior Genetics, 31,* 243–273. http://dx.doi.org/10.1023/A:1012294324713

*Bratko, D., & Butković, A. (2007). Stability of genetic and environmental effects from adolescence to young adulthood: Results of Croatian longitudinal twin study of personality. *Twin Research and Human Genetics, 10,* 151–157. http://dx.doi.org/10.1375/twin.10.1.151

*Bratko, D., & Marušić, I. (1997). Family study of the Big Five personality dimensions. *Personality and Individual Differences, 23,* 365–369. http://dx.doi.org/10.1016/S0191-8869(97)00081-0

Carey, G. (2002). *Human genetics for the social sciences*. Thousand Oaks, CA: Sage.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38,* 476–506. http://dx.doi.org/10.1037/h0054116

Cattell, R. B. (1953). Research designs in psychological genetics with special reference to the multiple variance method. *American Journal of Human Genetics, 5,* 76–93.

Cattell, R. B. (1960). The multiple abstract variance analysis equations and solutions: For nature-nurture research on continuous variables. *Psychological Review, 67,* 353–372. http://dx.doi.org/10.1037/h0043487

Cattell, R. B., Rao, D. C., & Schuerger, J. M. (1985). Heritability in the personality control system: Ego strength (C), super ego strength (G) and the self sentiment (Q3), by the MAVA model, Q-data, and maximum likelihood analyses. *Social Behavior and Personality, 13,* 33–41. http://dx.doi.org/10.2224/sbp.1985.13.1.33

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

*Coppen, A., Cowie, V., & Slater, E. (1965). Familial aspects of "neuroticism" and "extraversion." *The British Journal of Psychiatry, 111,* 70–83. http://dx.doi.org/10.1192/bjp.111.470.70

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33,* 533–552. http://dx.doi.org/10.1016/S0191-8869(01)00171-4

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256. http://dx.doi.org/10.1037/0022-3514.73.6.1246

Distel, M. A., Trull, T. J., Willemsen, G., Vink, J. M., Derom, C. A., Lynskey, M., . . . Boomsma, D. I. (2009). The five-factor model of personality and borderline personality disorder: A genetic analysis of comorbidity. *Biological Psychiatry, 66,* 1131–1138. http://dx.doi.org/10.1016/j.biopsych.2009.07.017

Eaves, L., Eysenck, H., & Martin, N. (1989). *Genes, culture and personality: An empirical approach*. London, England: Academic.

*Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K., . . . Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research, 2,* 62–80. http://dx.doi.org/10.1375/136905299320565933

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

*Eysenck, H. J. (1956). The inheritance of extraversion-introversion. *Acta Psychologica, 12,* 95–110. http://dx.doi.org/10.1016/0001-6918(56)90010-5

Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London, England: Methuen.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult versions)*. Kent, England: Hodder & Stoughton.

*Eysenck, H. J., & Prell, D. B. (1951). The inheritance of neuroticism: An experimental study. *Journal of Mental Science, 97,* 441–465. http://dx.doi.org/10.1192/bjp.97.408.441

Eysenck, S. B. G, Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6,* 21–29. http://dx.doi.org/10.1016/0191-8869(85)90026-1

Fanous, A., Gardner, C. O., Prescott, C. A., Cancro, R., & Kendler, K. S. (2002). Neuroticism, major depression and gender: A population-based twin study. *Psychological Medicine, 32,* 719–728. http://dx.doi.org/10.1017/S003329170200541X

*Finkel, D., & McGue, M. (1997). Sex differences and nonadditivity in heritability of the Multidimensional Personality Questionnaire scales. *Journal of Personality and Social Psychology, 72,* 929–938. http://dx.doi.org/10.1037/0022-3514.72.4.929

*Floderus-Myrhed, B., Pedersen, N., & Rasmuson, I. (1980). Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs. *Behavior Genetics, 10,* 153–162. http://dx.doi.org/10.1007/BF01066265

*Gillespie, N. A., Evans, D. E., Wright, M. M., & Martin, N. G. (2004). Genetic simplex modeling of Eysenck's dimensions of personality in a sample of young Australian twins. *Twin Research, 7,* 637–648. http://dx.doi.org/10.1375/1369052042663814

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229. http://dx.doi.org/10.1037/0022-3514.59.6.1216

Gough, H. G. (1956). *California Psychological Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1,* 77–89. http://dx.doi.org/10.1080/19312450709336664

Heath, A. C., & Martin, N. G. (1990). Psychoticism as a dimension of personality: A multivariate genetic test of Eysenck and Eysenck's psychoticism construct. *Journal of Personality and Social Psychology, 58,* 111–121. http://dx.doi.org/10.1037/0022-3514.58.1.111

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558. http://dx.doi.org/10.1002/sim.1186

*Hur, Y.-M. (2007). Evidence for nonadditive genetic effects on Eysenck personality scales in South Korean twins. *Twin Research and Human Genetics, 10,* 373–378. http://dx.doi.org/10.1375/twin.10.2.373

*Insel, P. M. (1974). Maternal effects in personality. *Behavior Genetics, 4,* 133–143.

Jang, K. L., Livesley, W. J., Ando, J., Yamagata, S., Suzuki, A., Angleitner, A., . . . Spinath, F. (2006). Behavioral genetics of the higher-order factors of the Big Five. *Personality and Individual Differences, 41,* 261–272. http://dx.doi.org/10.1016/j.paid.2005.11.033

*Jang, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the Big Five personality dimensions and their facets: A twin study. *Journal of Personality, 64,* 577–592. http://dx.doi.org/10.1111/j.1467-6494.1996.tb00522.x

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York: NY: Guilford Press.

Johnson, A. M., Vernon, P. A., & Feiler, A. R. (2008). Behavioral genetic studies of personality: An introduction and review of the results of 50+ years of research. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment. Vol. 1*: *Personality theories and models* (pp. 145–173). London, England: Sage.

*Johnson, W., McGue, M., & Krueger, R. F. (2005). Personality stability in late adulthood: A behavioral genetic analysis. *Journal of Personality, 73,* 523–552. http://dx.doi.org/10.1111/j.1467-6494.2005.00319.x

Kandler, C., Riemann, R., Spinath, F. M., & Angleitner, A. (2010). Sources of variance in personality facets: A multiple-rater twin study of self-peer, peer-peer, and self-self (dis)agreement. *Journal of Personality, 78,* 1565–1594. http://dx.doi.org/10.1111/j.1467-6494.2010.00661.x

Keller, M. C., Coventry, W. L., Heath, A. C., & Martin, N. G. (2005). Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behavior Genetics, 35,* 707–721. http://dx.doi.org/10.1007/s10519-005-6041-7

*Kendler, K. S., Myers, J., Potter, J., & Opalesky, J. (2009). A web-based study of personality, psychopathology and substance use in twin, other relative and relationship pairs. *Twin Research and Human Genetics, 12,* 137–141. http://dx.doi.org/10.1375/twin.12.2.137

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2,* 61–76. http://dx.doi.org/10.1002/jrsm.35

Larsen, R. J., & Buss, D. M. (2005). *Personality psychology: Domains of knowledge about human nature* (2nd ed.). New York, NY: McGraw-Hill.

*Lensvelt-Mulders, G., & Hettema, J. (2001). Analysis of genetic influences on the consistency and variability of the Big Five across different stressful situations. *European Journal of Personality, 15,* 355–371. http://dx.doi.org/10.1002/per.414

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209. http://dx.doi.org/10.1037/0003-066X.48.12.1181

Loehlin, J. C. (1978). Are CPI scales differently heritable: How good is the evidence? *Behavior Genetics, 8,* 381–382. http://dx.doi.org/10.1007/BF01067402

Loehlin, J. C. (1982). Are personality traits differentially heritable? *Behavior Genetics, 12,* 417–428. http://dx.doi.org/10.1007/BF01065633

Loehlin, J. C. (1992). *Genes and environment in personality development*. Newbury Park, CA: Sage.

Loehlin, J. C. (2012). The differential heritability of personality item clusters. *Behavior Genetics, 42,* 500–507. http://dx.doi.org/10.1007/s10519-011-9515-9

Loehlin, J. C., Horn, J. M., & Willerman, L. (1981). Personality resemblance in adoptive families. *Behavior Genetics, 11,* 309–330. http://dx.doi.org/10.1007/BF01070814

*Loehlin, J. C., & Martin, N. (2001). Age changes in personality traits and their heritabilities during the adult years: Evidence from Australian twin registry samples. *Personality and Individual Differences, 30,* 1147–1160. http://dx.doi.org/10.1016/S0191-8869(00)00099-4

*Loehlin, J. C., McCrae, R. R., Costa, P. T., Jr., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality, 32,* 431–453. http://dx.doi.org/10.1006/jrpe.1998.2225

*Losoya, S. H., Callor, S., Rowe, D. C., & Goldsmith, H. H. (1997). Origins of familial similarity in parenting: A study of twins and adoptive siblings. *Developmental Psychology, 33,* 1012–1023. http://dx.doi.org/10.1037/0012-1649.33.6.1012

*Martin, N., Goodwin, G., Fairburn, C., Wilson, R., Allison, D., Cardon, L. R., & Flint, J. (2000). A population-based study of personality in 34,000 sib-pairs. *Twin Research, 3,* 310–315.

Nagoshi, C. T., Johnson, R. C., & Honbo, K. A. M. (1992). Assortative mating for cognitive abilities, personality, and attitudes: Offspring from the Hawaii family study of cognition. *Personality and Individual Differences, 13,* 883–891. http://dx.doi.org/10.1016/0191-8869(92)90005-A

Oniszczenko, W., & Jakubowska, U. (2005). Genetic determinants and personality correlates of sociopolitical attitudes in a Polish sample. *Twin Research and Human Genetics, 8,* 47–52. http://dx.doi.org/10.1375/twin.8.1.47

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics, 8,* 157–159. http://dx.doi.org/10.3102/10769986008002157

Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the "same" construct? A meta-analytic investigation. *Personality and Individual Differences, 49,* 669–676. http://dx.doi.org/10.1016/j.paid.2010.06.014

*Pedersen, N. L., Plomin, R., McClearn, G. E., & Friberg, L. (1988). Neuroticism, extraversion, and related traits in adult twins reared apart and reared together. *Journal of Personality and Social Psychology, 55,* 950–957. http://dx.doi.org/10.1037/0022-3514.55.6.950

*Pincombe, J. L., Luciano, M., Martin, N. G., & Wright, M. J. (2007). Heritability of NEO PI-R extraversion facets and their relationship with IQ. *Twin Research and Human Genetics, 10,* 462–469. http://dx.doi.org/10.1375/twin.10.3.462

Plomin, R., Corley, R., Caspi, A., Fulker, D. W., & DeFries, J. (1998). Adoption results for self-reported personality: Evidence for nonadditive genetic effects? *Journal of Personality and Social Psychology, 75,* 211–218. http://dx.doi.org/10.1037/0022-3514.75.1.211

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics* (6th ed.). New York, NY: Worth.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer* (2nd ed.). San Francisco, CA: Freeman.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics* (5th ed.). New York, NY: Worth.

*Rettew, D. C., Rebollo-Mesa, I., Hudziak, J. J., Willemsen, G., & Boomsma, D. I. (2008). Non-additive and additive genetic effects on extraversion in 3314 Dutch adolescent twins and their parents. *Behavior Genetics, 38,* 223–233. http://dx.doi.org/10.1007/s10519-008-9192-5

*Rettew, D. C., Vink, J. M., Willemsen, G., Doyle, A., Hudziak, J. J., & Boomsma, D. I. (2006). The genetic architecture of neuroticism in 3301 Dutch adolescent twins as a function of age and sex: A study from the Dutch twin register. *Twin Research and Human Genetics, 9,* 24–29. http://dx.doi.org/10.1375/twin.9.1.24

*Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. *Journal of Personality, 65,* 449–475. http://dx.doi.org/10.1111/j.1467-6494.1997.tb00324.x

*Rose, R. J., Koskenvuo, M., Kaprio, J., Sarna, S., & Langinvainio, H. (1988). Shared genes, shared experiences, and similarity of personality: Data from 14,288 adult Finnish co-twins. *Journal of Personality and Social Psychology, 54,* 161–171. http://dx.doi.org/10.1037/0022-3514.54.1.161

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

*Saudino, K. J., Pedersen, N. L., Lichtenstein, P., McClearn, G. E., & Plomin, R. (1997). Can personality explain genetic influences on life events? *Journal of Personality and Social Psychology, 72,* 196–206. http://dx.doi.org/10.1037/0022-3514.72.1.196

*Scarr, S., Webber, P. L., Weinberg, R. A., & Wittig, M. A. (1981). Personality resemblance among adolescents and their parents in biologically related and adoptive families. *Journal of Personality and Social Psychology, 40,* 885–898. http://dx.doi.org/10.1037/0022-3514.40.5.885

*Shields, J. (1958). Twins brought up apart. *The Eugenics Review, 50,* 115–123.

*South, S. C., Krueger, R. F., Johnson, W., & Iacono, W. G. (2008). Adolescent personality moderates genetic and environmental influences

on relationships with parents. *Journal of Personality and Social Psychology, 94,* 899–912. http://dx.doi.org/10.1037/0022-3514.94.5.899

*Tambs, K., Sundet, J. M., Eaves, L., Solaas, M. H., & Berg, K. (1991). Pedigree analysis of Eysenck Personality Questionnaire (EPQ) scores in monozygotic (MZ) twin families. *Behavior Genetics, 21,* 369–382. http://dx.doi.org/10.1007/BF01065973

Tellegen, A. (1982). *Brief manual for the Differential Personality Questionnaire*. Unpublished manuscript, University of Minnesota, Minneapolis.

*Tellegen, A., Lykken, D. T., Bouchard, T. J., Jr., Wilcox, K. J., Segal, N. L., & Rich, S. (1988). Personality similarity in twins reared apart and together. *Journal of Personality and Social Psychology, 54,* 1031–1039. http://dx.doi.org/10.1037/0022-3514.54.6.1031

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology, 65,* 515–540. http://dx.doi.org/10.1146/annurev-psych-113011-143752

*Valera, E. M., & Berenbaum, H. (2001). A twin study of alexithymia. *Psychotherapy and Psychosomatics, 70,* 239–246. http://dx.doi.org/10.1159/000056261

van den Berg, S. M., de Moor, M. H., McGue, M., Pettersson, E., Terracciano, A., Verweij, K. J., . . . Boomsma, D. I. (2014). Harmonization of neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: An application of the item response theory. *Behavior Genetics, 44,* 295–313. http://dx.doi.org/10.1007/s10519-014-9654-x

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45,* 576–594. http://dx.doi.org/10.3758/s13428-012-0261-6

*Vernon, P. A., Martin, R. A., Schermer, J. A., & Mackie, A. (2008). A behavioral genetic investigation of humor styles and their correlations with the Big-5 personality dimensions. *Personality and Individual Differences, 44,* 1116–1125. http://dx.doi.org/10.1016/j.paid.2007.11.003

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://www.jstatsoft.org/v36/i03/paper

*Vukasović, T., Bratko, D., & Butković, A. (2009). From specific reactions to aggregated dimensions of personality: Familial aggregation estimate in the function of construct width. *Contemporary Psychology, 12,* 323–339.

*Weiss, A., Bates, T. C., & Luciano, M. (2008). Happiness is a personal-(ity) thing: The genetics of personality and well-being in a representative sample. *Psychological Science, 19,* 205–210. http://dx.doi.org/10.1111/j.1467-9280.2008.02068.x

*Wray, N. R., Birley, A. J., Sullivan, P. F., Visscher, P. M., & Martin, N. G. (2007). Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Research and Human Genetics, 10,* 695–702. http://dx.doi.org/10.1375/twin.10.5.695

Received November 12, 2014
Revision received February 19, 2015
Accepted March 25, 2015 ■