Running head: SNP-BASED HERITABILITY OF NEUROTICISM

SNP-Based Heritability Estimates of Common and Specific Variance in Self- and Informant-Reported Neuroticism Scales

Anu Realo[1,2], Peter J. van der Most[3],
Jüri Allik[2,4], Tõnu Esko[5], Bertus F. Jeronimus[6,7], Liisi Kööts-Ausmees[2], René Mõttus[2,8]
Felix C. Tropf[9,10], Harold Snieder[3,5], Johan Ormel[6]

[1]Department of Psychology, University of Warwick, UK

[2]Department of Psychology, University of Tartu, Estonia

[3]Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

[4]The Estonian Academy of Sciences, Estonia

[5]Estonian Genome Centre of University of Tartu, Estonia

[6]Department of Psychiatry, Interdisciplinary Center of Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, The Netherlands

[7]Department of Developmental Psychology, University of Groningen, The Netherlands

[8]Department of Psychology, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK

[9]Department of Sociology, Interuniversity Center for Social Science Theory and Methodology, University of Groningen, The Netherlands

[10]Department of Sociology, Nuffield College, University of Oxford, UK

Corresponding author:

Anu Realo, Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom, E-mail: anu.realo@ut.ee

Abstract

Objective. Our study aims to estimate the proportion of the phenotypic variance of Neuroticism and its facet scales that can be attributed to common SNPs in two adult populations from Estonia (EGCUT; *N* = 3,292) and the Netherlands (Lifelines; *N* = 13,383).

Method. Genomic-Relatedness-Matrix Restricted Maximum Likelihood (GREML) using Genome-wide Complex Trait Analysis (GCTA) software was employed. To build upon previous research, we used self- and informant-reports of the 30-facet NEO personality inventories and analyzed both the usual sum scores and the residual facet scores of Neuroticism.

Results. In the EGCUT cohort, the proportion of phenotypic variance explained by the additive effects of common genetic variants in self- and informant-reported Neuroticism domain scores was 15.2% (p = .070, SE = .11) and 6.2% (p = .293, SE = .12), respectively. The SNP-based heritability estimates at the level of Neuroticism facet scales differed greatly across cohorts and modes of measurement but were generally higher (a) for self- than for informant-reports, and (b) for sum than for residual scores.

Conclusions. Our findings indicate that a large proportion of the heritability of Neuroticism is not captured by additive genetic effects of common SNPs with some evidence for gene-environment interaction across cohorts.

Keywords: Neuroticism, GREML-GCTA, SNP-based heritability, self- vs. informant-reports, sum- vs. residual scores

SNP-Based Heritability Estimates of Common and Specific Variance in Self- and Informant-Reported Neuroticism Scales

Neuroticism vs. emotional stability appears to be one of the most universal domains in personality theory and measurement (H. J. Eysenck, 1998/1947). The importance of Neuroticism – the frequent experience of negative affect – in personality research stems largely from its significance for public health. Overwhelmingly, research evidence has shown that Neuroticism is correlated with a wide range of mental and physical health problems and their development and that low scores of Neuroticism predict longevity after controlling for age, sex, education, smoking, alcohol consumption, physical activity, and baseline health (e.g., Friedman & Kern, 2014; Jeronimus, Kotov, Riese, & Ormel, 2016; Lahey, 2009; Ormel, Jeronimus, et al., 2013). It has even been argued that the economic burden of Neuroticism on society exceeds that of the common mental disorders combined (Cuijpers et al., 2010).

Despite the central importance of Neuroticism, there appears to be little consensus about its definition (Ormel, Riese, & Rosmalen, 2012). Like many other traits, Neuroticism is often viewed as a multifaceted construct consisting of various components that are highly correlated but partially distinct, including anxiety, dependence, sadness, worry, vulnerability, and impulsivity. In the current study, we use common genome-wide genetic variants to examine the heritability of Neuroticism and its facets within the context of the Five-Factor Model (FFM). According to the FFM, the core of the Neuroticism domain is ”the general tendency to experience negative affects such as fear, sadness, embarrassment, anger, guilt, and disgust” (Costa & McCrae, 1992, p. 14). The NEO family of instruments (i.e., NEO Personality Inventory, the Revised NEO PI, and the NEO PI-3), which are based on the FFM, distinguishes between six

facets of Neuroticism: N1: Anxiety, N2: Angry Hostility, N3: Depression, N4: Self-Consciousness, N5: Vulnerability, and N6: Impulsiveness (Costa & McCrae, 1992). As we argue below, there is enough evidence to suggest that heritability estimates for Neuroticism may differ across facets.

The Biological and Genetic Basis of Neuroticism

Despite Neuroticism being strongly associated with important health outcomes, its causal role in health as well as its biological basis is far from being fully understood (Friedman & Kern, 2014; Ormel, Bastiaansen, et al., 2013). The health relevance of Neuroticism as highlighted above, combined with the uncertainty of its biological basis and exact role in health and disease, has led to a growing interest in the genetic basis of Neuroticism.

Behavioral genetic (mostly twin and adoption) studies have shown that Neuroticism is substantially heritable: approximately 40-60% of the variance in Neuroticism scores is estimated to be attributable to genetic factors (Bouchard & Loehlin, 2001; Jang, Livesley, & Vernon, 1996; Viken, Rose, Kaprio, & Koskenvuo, 1994; Yamagata et al., 2006). A recent meta-analysis of twin data from six cohorts (total *N* = 29,496) found that the heritability of harmonized Neuroticism scores was 48% (van den Berg et al., 2014), whereas a meta-analysis of 62 behavior genetic studies representing more than 100,000 participants showed that the average effect of genetic contributions to individual differences in Neuroticism was 39% (Vukasović & Bratko, 2015). Although Turkheimer, Petterson, and Horn (2014) convincingly argue in their recent review that all personality traits are equally heritable, there are also several studies which have shown that heritability estimates of Neuroticism appear to be somewhat lower than those of the other Big Five factors (Kandler, Riemann, Spinath, & Angleitner, 2010; Riemann, Angleitner, & Strelau, 1997; Riemann & Kandler, 2010).

However, until fairly recently, the search for genes associated with Neuroticism has not been very successful. Some candidate gene studies reported associations between Neuroticism and specific genetic variants, such as the CNR1 gene (Juhasz et al., 2009) or a polymorphism (5-HTTLPR) of the serotonin transporter gene (Schinka, Busch, & Robichaux-Keene, 2004; Sen, Burmeister, & Ghosh, 2004), among others, but most of these findings failed to be replicated. Even when such associations were found, the polymorphisms seemed to have modest predictive value: 2% or even less (see Munafo, 2010; Munafo & Flint, 2011, for a discussion). Also, despite very promising progress in behavioral genetics, no significant results were found for Neuroticism in three large scale genome-wide association (GWA) studies (Calboli et al., 2010; de Moor et al., 2012; Service et al., 2012) where Neuroticism was measured by three different instruments (i.e., the NEO Five-Factor Inventory, Cloninger's Temperament and Character Inventory, and the Eysenck Personality Questionnaire). The only noteworthy findings come from three very recent studies. First, a meta-analysis of GWAS results across 29 cohorts by de Moor et al. (2015) identified a new locus for Neuroticism in the gene MAGI1, which had been associated with bipolar disorder and schizophrenia in earlier studies. Second, Smith and colleagues (2016) identified 9 novel loci associated with Neuroticism in a combined meta-analysis across three cohorts and nearly 100,000 participants using the short form of the Eysenck Personality Questionnaire-Revised (EPQ-R-S; S. B. G. Eysenck, Eysenck, & Barrett, 1985). Finally, Okbay and colleagues (2016) combined results of the above two studies (across heterogeneous measures of Neuroticism) and could identify a total of 11 variants associated with Neuroticism, including 2 inversion polymorphisms, in a total sample of 170,911

participants. However, none of the loci associated with Neuroticism by Smith et al. (2016) were replicated by Okbay and colleagues (2016).

These recent findings have discredited the idea that there is a relatively small set of genes that determines the level of either a personality factor or its facets. Instead, in light of the current findings, it seems more plausible that personality traits (including Neuroticism) are influenced by large sets of different common genetic variants, as assessed by single-nucleotide polymorphism (SNPs), and that the influence of these genes may differ across facets (McCrae, 2015).

Molecular Genetic Estimates of Heritability

In order to estimate the influence of common variants with small effects on the total heritability, a new method – Genomic-Relatedness-Matrix Restricted Maximum Likelihood (GREML) as implemented in the genome-wide complex trait analysis (GCTA) software – was proposed by Yang and colleagues (2010; 2011). The aim of GREML-GCTA is not to identify SNPs related to the target phenotype, but to estimate the total variance explained by all common SNPs for a given trait. As Yang et al. (2011) argued, it is possible that most of the heritability for Neuroticism or any other complex trait "is hiding rather than missing," because many SNPs contribute small effects. In essence, GREML-GCTA estimates the genetic influence on a certain trait or target phenotype by "predicting phenotypic similarity for each pair of individuals in the sample from their total SNP similarity" (Plomin, Haworth, Meaburn, Price, & Davis, 2013, p. 563).

GREML-GCTA has been successfully applied to a variety of complex traits, where it has demonstrated that the contribution of common additive genetic variants to phenotypic variation

is about 45% for height (Yang et al., 2010) and 30% for weight (Llewellyn, Trzaskowski, Plomin, & Wardle, 2013), a little less than 30% for cognitive abilities (Davies et al., 2015), 32% for major depressive disorder (Lubke et al., 2012), and 23% for liability to both schizophrenia (S. Hong Lee et al., 2012) and borderline personality disorder (Lubke et al., 2014). For most complex traits examined so far, GREML-GCTA or SNP-based heritability estimates appear to be about half of the heritability estimates from twin studies (Plomin & Deary, 2015).

In only a handful of studies so far, GREML-GCTA has been applied to personality traits, mostly to Neuroticism and Extraversion. To the best of our knowledge, a study by Power and Pluess (2015) in the British longitudinal cohort National Child Development Study (NCDS) is the first and only GREML-GCTA study including all five of the FFM. Common SNPs accounted for 15% of the variance in Neuroticism and for 8% of the variance in Extraversion, 21% in Openness, 0% in Agreeableness, and 1% in Conscientiousness, as measured by the International Personality Item Pool (Power & Pluess, 2015).

In a study by de Moor and colleagues (2015), the SNP-based genetic similarity across individuals accounted for 14.7% of the variance in Neuroticism in the Netherlands Twin Register (NTR) cohort and 15.7% in the Australian Berghofer Medical Research Institute adult cohort (QIMR), when using harmonized Neuroticism scores across different measurement instruments (van den Berg et al., 2014).[1] A very similar estimate was obtained by Mõttus and colleagues (2015) in the Generation Scotland cohort and by Smith et al. (2016) in the UK biobank cohort, using the short version of the EPQ-R (S. B. G. Eysenck et al., 1985). These SNP-based estimates, however, were nearly twice as high as those reported by Verweij and colleagues (2012) for Neuroticism measured by different versions of Cloninger's Harm Avoidance subscale (Cloninger,

Przybeck, & Svrakic, 1991; Cloninger, Svrakic, & Przybeck, 1993) in combined Australian and Finnish samples (7%) and by Vinkhuyzen et al. (2012) for harmonized Neuroticism scores[2] in combined samples from Australia, Sweden, the United Kingdom, and the United States (6%). In general, it appears that the SNP-based heritability estimates for Neuroticism (as well as for other main personality traits) are lower than those of other complex traits that have been subjected to the same analysis (see above).

The Present Study

The general aim of the present study is to estimate the proportion of the phenotypic variance of Neuroticism that can be attributed to common SNPs in two large-scale adult populations from Estonia and the Netherlands. Our study goes beyond earlier research in several important aspects.

First, as described above, the SNP-based heritability of Neuroticism has so far been examined using so-called harmonized Neuroticism scores (de Moor et al., 2015; Vinkhuyzen et al., 2012), the 10 items of the IPIP (Power & Pluess, 2015), the short version of the EPQ-R (Mõttus, Marioni, et al., 2015; Smith et al., 2016), or the Harm Avoidance scale from Cloninger's personality questionnaires (Verweij et al., 2012). Although Van den Berg and colleagues (2014) argued that harmonization across nine different personality inventories was very successful, and that Neuroticism and extraversion personality inventories were "largely measurement invariant across cohorts" (p. 296), such harmonized phenotype scores are, by definition, heterogeneous, because they are based on different personality scales measuring different facets or characteristics of Neuroticism. For instance, the Harm Avoidance subscale from Cloninger's personality scales, used by Verweij et al. (2012), cannot be taken as a pure measure of

Neuroticism, as it is a composite of both Neuroticism and extraversion (De Fruyt, Van De Wiele, & Van Heeringen, 2000). Several researchers have argued (e.g., Manolio et al., 2009) that in order to progress in our search for missing heritability, one important step is to reduce heterogeneity by refining our phenotypes. To achieve this, we used two versions of the same well-defined and well-validated FFM personality inventory: the NEO Personality Inventory-3 (NEO PI-3; McCrae, Costa, & Martin, 2005) and the Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992).

Second, we examined the variance in Neuroticism which could be accounted for by the additive effects of the SNPs, not only at the domain level as had been done in earlier studies (de Moor et al., 2015; Mõttus, Marioni, et al., 2015; Power & Pluess, 2015; Smith et al., 2016; Verweij et al., 2012; Vinkhuyzen et al., 2012), but also at the level of the more specific facet scales of the NEO PI-3. This was prompted by the growing number of studies showing that important variation in personality occurs at the level of facet scales (e.g., Mõttus, McCrae, Allik, & Realo, 2014; Mõttus, Realo, et al., 2015), or even at the level of more narrow personality characteristics/nuances (Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2016). A recent meta-analysis by Briley and Tucker-Drob (2014) showed that broad measures of personality (e.g., FFM domains) tended to be less heritable than narrow measures (e.g., facets), although the differences were rather small. On the other hand, several twin studies have shown that there are substantive differences between the heritability estimates of the facets of the FFM domains (Luciano, Wainwright, Wright, & Martin, 2006; Pincombe, Luciano, Martin, & Wright, 2007), including the facets of Neuroticism (Briley & Tucker-Drob, 2012; Jang et al., 1996; Jang, McCrae, Angleitner, Riemann, & Livesley, 1998; Kandler et al., 2010). For instance, in Jang and

colleagues' two studies (1996; 1998), the heritability estimates of Neuroticism facets ranged from .37 (N2: Hostility and N5: Impulsiveness) to .46 (N4: Self-Consciousness; Jang et al., 1996) and from .26 (N1: Anxiety) to .44 (N6: Vulnerability; Jang et al., 1998), respectively. In sum, there is enough evidence to suggest that (a) the SNP-based heritability estimates for Neuroticism may differ at the domain and the facet levels and that (b) there might be differences among facet scales because "aggregating facet scores to produce domain scores overshadows nuances of the genetic effects and renders the domain scores not genetically crisp" (Briley & Tucker-Drob, 2012; p. 755).

Third, as recently argued by McCrae (2015), it seems very likely that broad personality factors correspond to large sets of different genes, some of which may influence all facets, some several facets, and some only one or two facets. Thus, "a gene that affected only one facet, however, would not be part of the set [of genes that pertains to domain-level variance]; it would instead be one of the sources of specific variance in the facet" (p. 107). Consequently, and similarly to Jang and colleagues (1998), we decomposed the variability in Neuroticism into the common variance of its measured manifestations and their unique variances, and examined the SNP-based heritability separately in each of these components. There is evidence that facet-specific variance is agreed upon by different raters (e.g., self- vs. informant-reports; Kandler et al., 2010; McCrae & Costa, 1992; Mõttus et al., 2014) and that such residual scores contain valid and meaningful specific variance over and above the variance that they share with their respective domain scales (McCrae, 2015; Mõttus et al., 2016).

Fourth, in addition to self-reports, we employed informant-reports of personality by knowledgeable others. Informant-reports have become one of the most useful tools in

personality research (Allik, Borkenau, Hrebícková, Kuppens, & Realo, 2015), and it has even been shown that they may have substantially better validity than self-reports in predicting real-life outcomes, such as job performance and academic achievement (Connelly & Ones, 2010). Although substantial cross-rater agreement on personality traits has been demonstrated in numerous studies (see Connelly & Ones, 2010; Connolly, Kavanagh, & Viswesvaran, 2007, for reviews), with correlations between self- and informant-reports for Neuroticism typically ranging between .40 and .50, there are only a few twin studies which have examined the genetics of personality using informant-reports. Most of these studies have found a substantial genetic influence on informant-rated personality trait scores (including Neuroticism), similar to what has been found using self-reports (Borkenau, Riemann, Angleitner, & Spinath, 2001; Briley & Tucker-Drob, 2014; Heath, Neale, Kessler, Eaves, & Kendler, 1992; Riemann et al., 1997). Riemann and colleagues (1997) found, for instance, that additive genetic effects explained 52% and 61% of the variability in Neuroticism scales for self- and informant-reports, respectively. Our study is the first to examine whether SNP-based estimates of Neuroticism heritability generalize across self- and informant-reports, and the extent to which SNP-based estimates of heritability for Neuroticism are method-specific.

Fifth, we used two adult population samples, one from Estonia and the other from the Netherlands, to ensure that our findings were not restricted to a specific population. First, we calculated the SNP-based heritability estimates separately for the two samples. Next, we explored whether different genetic markers influence Neuroticism in different environments by testing for gene-environment interactions in a combined dataset, using the cohort of origin as a proxy for the different environments (cf. Tropf et al., 2016).

Method

Participants

The EGCUT sample. The Estonian participants for the present study came from the Estonian Biobank cohort (approximately 52,000 individuals), which is a volunteer-based sample of the Estonian resident adult population (Leitsalu et al., 2014). Participants were recruited by general practitioners (GPs), physicians, and other medical personnel in hospitals and private practices, as well as the recruitment offices of the Estonian Genome Centre of the University of Tartu (EGCUT). Each participant signed an informed consent form (available at www.biobank.ee) and underwent a standardized health examination by a physician. Participants also donated blood samples for DNA and completed a Computer Assisted Personal Interview (CAPI) on health-related topics and various clinical diagnoses listed in the WHO ICD-10 (Leitsalu et al., 2014).

The sample for the current study consists of 3,292 individuals (59.3% female) for whom both personality (self- and informant-reports) and genotype data were available.[3] The mean age of the participants was 47.2 years (*SD* = 17.0, ranging from 18 to 91 years). About 40 per cent of the respondents had higher education. The mean age of informants (*N* = 3,151, 71.3% female, gender was unknown for 64 participants) was 42.1 (*SD* = 16.0) years. On average, informants had known target subjects for 23.4 (*SD* = 15.1) years. Among informants, 47.5% were spouses or partners, 17.5% were parents, 15.3% were friends, and the remaining 19.7% were other relatives or acquaintances (for 60 respondents the relationship to the target was not known).

Genotyping was performed on multiple Illumina platforms (the Illumina CNV370-Duo BeadChip platforms for 1,126 samples; OmniExpress BeadChips for 1,079; HumanCoreExome-11 BeadChips for 569; HumanCoreExome-10 BeadChips for 174; and PsychChip arrays for 344

participants). Prior to imputation, samples with a call rate <0.95, excess heterozygosity, non-Caucasian ethnicity (as determined by principal component analysis), high relatedness (pi-hat > 0.4), or a gender mismatch were excluded, as were SNPs with a MAF < 1%, a HWE p-value ≤10-3, or a callrate < 95%. A total of 257,581 SNPs was used for imputation. Imputation was performed with IMPUTE v2 (Howie, Donnelly, & Marchini, 2009) using the 1000 Genomes Phase 1 v3 reference panel (The 1000 Genomes Project Consortium, 2012; The Genomes Project, 2015).

The Lifelines sample. The Dutch sample is part of the Lifelines Cohort Study (approximately 167,700 individuals), a large population-based cohort study and biobank in the Northern part of the Netherlands, which was established "as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing" (Scholtens et al., 2015; p. 1172). The Lifelines dataset contains information about participants' medical history, biochemistry, psychosocial characteristics, lifestyle, and other relevant characteristics.

The sample used in the current study includes 13,383 people (58.2% female), with a mean age of 48.8 years (*SD* = 11.4, ranging from 18 to 90 years) who completed a personality inventory and for whom genotype data were available. About 33 per cent of the respondents had higher education.

Genotyping was performed with the Illumina Cyto SNP12 v2 chip, and genotype calling was carried out with Illumina GenomeStudio. Prior to imputation, samples with a call rate < 0.8, excess heterozygosity, non-Caucasian ethnicity (as determined by principal component analysis), high relatedness (pi-hat > 0.4), or a gender mismatch were excluded, as were SNPs

with a MAF < 1%, a HWE p-value $\leq 10^{-3}$, or a callrate < 95%. A total of 257,581 SNPs was used for imputation. Imputation was performed in MiniMac (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) using the 1000 Genomes Phase 1 v3 reference panel (The 1000 Genomes Project Consortium, 2012; The Genomes Project, 2015).

Personality Measures

The EGCUT cohort. In the EGCUT cohort, personality was assessed with the Estonian version of the NEO PI-3 (McCrae et al., 2005), which is a slightly modified version of the NEO PI-R (Costa & McCrae, 1992; Kallasmaa, Allik, Realo, & McCrae, 2000). Like the NEO PI-R, the NEO PI-3 has 240 items that measure 30 personality facets, which in turn are grouped into the five FFM domains of six facet scores each. The NEO PI-3 has excellent psychometric properties in a wide range of countries (De Fruyt et al., 2009), including Estonia. Items are answered on a five-point scale (0 = *false/strongly disagree* - 4 = *true/strongly agree*).

Sum scores of the facet scales were calculated by adding up the scores of the 8 individual items (reverse scored if required) in each facet scale, separately for self- and informant-ratings. Higher scores reflect higher levels of anxiety, angry hostility, depression, and so on. The total score of the Neuroticism domain is the sum of all items from each subscale, with higher scores indicating greater levels of Neuroticism. Finally, to compare the findings with the Lifelines sample (see below), a sum score of 32 items was calculated (henceforth N-32) from the Angry Hostility (N2), Self-Consciousness (N4), Impulsiveness (N5), and Vulnerability (N6) facet scales. The descriptive statistics of the scales, including Cronbach alphas, are shown in Table 1.

Insert Table 1 about here

EGCUT participants completed the self-report form and informants the observer-report form of the NEO PI-3. Similar to findings from other studies (e.g., Connolly et al., 2007), the cross-rater correlation between the respective sum scores was .53 for the Neuroticism domain and ranged from .42 for Impulsiveness (N5) to .51 for Anxiety (N1) and Depression (N3) among the six facet scales (*median* = .47; all correlations significant at $p < 0.001$), see Table 1.

The Lifelines Sample. In the Lifelines sample, participants were asked to complete an abbreviated version[4] of the Dutch NEO PI-R (Hoekstra, Ormel, & De Fruyt, 1996). Therefore, self-reported data were only available for four facet scales of Neuroticism: Angry Hostility (N2), Self-Consciousness(N4), Impulsiveness (N5), and Vulnerability (N6). The items were answered on a five-point scale ranging from 1 (*false/strongly disagree*) to 5 (*true/strongly agree*). Sum scores of the four abovementioned facet scales were calculated in the same way as in the EGCUT sample. To compare the findings with the EGCUT sample, also a sum score of these 32 items (N-32) was calculated from the Angry Hostility (N2), Self-Consciousness (N4), Impulsiveness (N5), and Vulnerability (N6) facet scales. The means, *SDs,* and Cronbach alphas of the four facet scales and the N-32 score are given in Table 1.

Preliminary Analyses of the EGCUT Sample on the Personality Measure

Residual facet scores. The residual facet scores in the EGCUT sample were obtained using structural equation modelling, which allowed us to decompose the common variance of the Neuroticism domain (shared variance across all items of the domain) and its facets (shared variance of all items measuring a given facet, independently of what they shared with the common variance of the domain). First, the same bi-factor model as in Mõttus et al. (2014; see Figure 1) was fitted to self-reports. More specifically, the model was constructed in a way that

all 48 self-reported items defined the general domain of Neuroticism, and all 8 self-reported items of each facet defined the respective facet score. The correlations between the facet scores and the Neuroticism domain score were set to zero, resulting in an orthogonal domain and facet scores. Differently from Mõttus et al. (2014), the facets of Neuroticism were not allowed to correlate among themselves in the current study, which also made them orthogonal in relation to each other. Since this bi-factor model did not fit the data sufficiently well, the model was tweaked until it met the fit criteria, that is, until the Comparative Fit Index (CFI) reached a value of .90 and the root mean squared error of approximation (RMSEA) reached .06[5], by omitting loadings below < .20 and by allowing residual correlations among items (based on modification indices). The final model (shown in Figure 1) was then applied to informant-ratings; this fitted the data reasonably well with CFI = .894 and RMSEA = .065. As a result, we obtained a score of the general Neuroticism factor that underlies each of the 48 items and the residual scores of six facets, each of which accounts for the unique influence of the specific facet over and above the general factor. In other words, the six residual facet scores each represent variance not accounted for by the general factor of Neuroticism. The cross-rater correlations for the residual facet scores were .34, .42, .33, .40, .38, and .24 (*median* = .36) for Anxiety (N1) to Vulnerability (N6), respectively. These estimates are similar to what was reported by Costa and McCrae (2008; median = .33). The residual scores were – similarly to the sum-scores – analyzed with GCTA.

Preparing the Genotypes

The EGCUT sample. The EGCUT genotypes were delivered in 5 separate files (corresponding to the 5 genotyping platforms used). First, we selected only autosomal HapMap3 SNPs, as this

set was optimized to capture common genetic variation in the human genome (Altshuler et al., 2010). The files were then converted to the PLINK format with GTOOL (Freeman & Marchini, 2012), and merged with PLINK (Purcell, 2009; Purcell et al., 2007). We included only SNPs that were present on all five platforms. We also checked whether the remaining SNPs had large differences in allele frequencies between platforms, and excluded those with a difference > 15%. A number of SNPs with allele or position mismatches was also excluded. The final dataset contained 3,303 individuals and 1,234,312 markers. The first four principal components were calculated in PLINK for use as covariates.

The Lifelines sample. The Lifelines genotypes were available as PLINK files, with the first ten principal component values calculated with Eigenstrat (Price et al., 2006), so we only needed to select autosomal HapMap3 SNPs. The final Lifelines dataset contained 13,383 individuals and 1,401,138 SNPs.

The combined sample. For the combined analysis, we used the same merging procedure as for the EGCUT: we only included SNPs that appeared in both cohorts, and excluded SNPs with allele, allele-frequency (> 15%), or position mismatches between cohorts. When combining the dataset, 0.1% of SNPs in the EGCUT and 12.0% of SNPs in the Lifelines cohort were excluded due to reasons above. The final, combined dataset, consisted of 16,686 individuals and 1,233,075 SNPs.

GREML analyses in GCTA

GREML analyses were performed by calculating a genetic relationship matrix (GRM) in GCTA, and then using the GRM in a REML analysis of the phenotypes. We used a GRM cut-off value of 0.05 (Zaitlen et al., 2013), which excludes relationships that are approximately closer

than second-cousins. This removed around 460 samples in the EGCUT sample, 3,910 in the Lifelines sample, and 4,370 in the combined sample. In the EGCUT sample, we used sex, genotyping platform, age, and the first 4 principal components as covariates. In the Lifelines sample, we used sex, age, and the first 10 principal components as covariates. In order to run the analysis in the combined dataset, we merged the genetic data in PLINK, and then determined the first ten principal components using GCTA. Next, we calculated the residuals for the phenotypes separately for both cohorts (using age, sex, genotyping platform (EGCUT only) and the first ten PCs as covariates), and analyzed them in GCTA. We also included the cohort from which the sample originated as an environmental factor, to test for gene-environment interactions.

Power analysis (http://cnsgenomics.com/shiny/gcta) indicated a 99% chance of detecting a SNP-based heritability estimate (h^2) of 0.15 in Lifelines and the combined sample, and a mere 27% chance of detecting a h^2 of 0.15 in EGCUT sample.

Results

Genetic Correlations between SNP-based Estimates for Self- and Informant-Reports

Results of bivariate analyses (S. H. Lee, Yang, Goddard, Visscher, & Wray, 2012) showed that genetic correlations between the SNP-based estimates for self- and informant-reports of Neuroticism domain and facet sum scores were all 1, except for N6: Vulnerability (rG = .91).

Variance in Neuroticism Explained by Common SNPs

The SNP-based estimates of additive genetic effects of Neuroticism sum and residual facet scores, as well as for Neuroticism domain scores, are shown in Table 2. In Lifelines, all estimates, except for N5: Impulsivity, were significant at the level of $p < .05$. In EGCUT, on the contrary,

only the estimate for self-reported Impulsiveness (N5) sum scores was significant at $p < .05$, and the GREML-GCTA estimates for self-reported Depression (N3) and Vulnerability (N6) sum scores and for other-reported Anxiety (N1) sum scores were marginally non-significant at $p = .054$, .058, and .050, respectively. Due to the small sample size, the standard errors of the SNP-based heritability estimates in the EGCUT cohort were relatively large (11-12%), reducing the significance of our findings. In the Lifelines cohort, the standard errors were 4%.

Insert Table 2 about here

Self-Reports. In the EGCUT cohort, the SNP-based estimates for the Neuroticism domain score was 15.2% ($p = .070$). The SNP-based heritability estimate of the N-32 score (that is, a composite index of N2, N4, N5, and N6 items) was 13.4% ($p = .098$) in the EGCUT and 11.0% ($p = .002$) in the Lifelines cohort (Table 2).

At the level of facet scales, common SNPs explained between 8.3% (Angry Hostility (N2); $p = .210$) and 19.8% (Impulsiveness (N5); $p = .040$) of the variance in the EGCUT self-reported facet sum scores (*median* = 14.6%) and between 3.3% (Impulsiveness (N5); $p = .192$) and 16.1% (N2: Angry Hostility; $p < .001$) of the Lifelines facet sum scores (*median* = 11.2%).

The SNP-based heritability estimates in the EGCUT sample for self-reported residual facet scales – from which the common variance of Neuroticism had been statistically removed – ranged from 0% (Depression (N3), Self-Conscientiousness (N4) and Impulsivity (N5), all n.s.) to 14.7% (Anxiety (N1), $p = .106$), *median* = 3.6%.

Informant-reports. In the EGCUT cohort, common SNPs explained 6.2% of the variance (p = .293) in the informant-reported Neuroticism domain score and 1.5% of the variance (p = .448) in the N-32 score (see Table 2).

At the level of facet scales, common SNPs explained between 0% (Self-Consciousness (N4) and Impulsiveness (N5); n.s.) and 17.5% (Anxiety (N1); p = .050) of the variance in the EGCUT facet sum scores (*median* = 6.8%) of informant-ratings. The SNP-based heritability estimates for residual facet scales in informant-ratings ranged from 0% (Self-Consciousness (N4) and Impulsiveness (N5), n.s.) to 16.8% (Angry Hostility (N2); p = .073), *median* = 5.5%.

Gene-Environment Interactions in Neuroticism

Finally, we tested whether genetic influences on Neuroticism were shared across cohorts, or whether there were unique genetic effects to them. To this aim, we combined the EGCUT and Lifelines cohorts (using the cohort of origin as the environment) and estimated a gene-environment interaction model which gives two estimates, one for the genetic variance component which is shared across both cohorts and one that is specific to cohorts.

Insert Table 3 about here

A significant gene-environment interaction (p < .05) was found in self-reported Impulsivity (N5) sum scores (Table 3). At the same time, the genetic effects of Impulsivity (N5; 0.1%) and N-32 (6.8%) did not reach statistical significance (p < .05) in the combined sample.

Discussion

Despite strong research evidence from behavioral genetics showing that all personality traits, regardless of their type or breadth, are heritable (Turkheimer et al., 2014; Vukasović &

Bratko, 2015), candidate gene and GWA studies have provided little evidence for the existence of an involvement of specific genetic variants in personality traits (see Okbay et al., 2016 and Smith et al., 2016; for recent exceptions). It has been proposed that this may be due to the existence of many common variants with tiny effects which remain under the threshold of statistical significance (Manolio et al., 2009) and that it is the cumulative effect of these variants that makes people who share them similar in phenotype (Turkheimer et al., 2014).To overcome this problem, GREML-GCTA (Yang et al., 2010) has been developed and successfully applied to many complex phenotypes, providing considerably higher estimates of heritability than earlier GWA studies.

To complement the existing research on personality traits, we examined the SNP-based heritability of Neuroticism both at the domain and the facet level by using self- and other-reports that were obtained with well-defined and validated NEO personality instruments. Furthermore, in addition to the usual NEO sum scores of Neuroticism domain and facet scales, we also examined the SNP-based heritability of the residual facet scores of Neuroticism from which the common variance of Neuroticism had been statistically removed.

SNP-based Heritability of General Neuroticism

In the EGCUT cohort, the proportion of phenotypic variance explained by the additive effects of common genetic variants in the self-reported Neuroticism domain score (15.2%, p = .070) was very similar to what was reported earlier by De Moor and colleagues (2015; 14.7% and 15.7%, respectively), Power and Pluess (2015; 15%), Smith et al. (2016; 15%), and Mõttus and colleagues (2015; 16%).

In informant-reports, however, common SNPs explained only 6.2% (p = .293) of the variance in Neuroticism domain score. In the case of the N-32 score, the estimates ranged between 11% (Lifelines) and 14% (EGCUT) in self-reports vs. 1.5% in informant-reports (EGCUT). As Mõttus and colleagues (2016) recently showed, the heritability estimates of informant-ratings may indeed be smaller than those of self-ratings, even when based on aggregated ratings of multiple informants.

In sum, our findings indicate that the SNP-based heritability of general self-reported Neuroticism is typically 15%-16% which is less than half the average heritability estimate of Neuroticism (39%) obtained in a recent meta-analysis of behavior genetics studies in twins (Vukasović & Bratko, 2015). This estimate appears to be relatively robust across different measurement instruments and samples, suggesting that the phenotypic heterogeneity has a little effect on the heritability estimate. The difference between the heritability estimates from twin and GCTA-GREML studies may imply that the genetic contribution is primarily caused by (a) non-additive genetic variance due to genetic interactions within or across loci (Plomin, Corley, Caspi, Fulker, & Defries, 1998; Zuk, Hechter, Sunyaev, & Lander, 2012); (b) the existence of rare(r) variants with potentially large effects, which are not represented on genotype arrays (Yang et al., 2015); or (c) epigenetic influences that may substantially contribute to the transgenerational inheritance of complex traits (Trerotola, Relli, Simeone, & Alberti, 2015). Specifically, the finding that heritability estimates tend to be systematically lower in family/adoption studies than in twin studies (Vukasovic & Bratko, 2015), suggest the involvement of epistatic or dominance mechanisms (non-additive genetic variance).

Alternatively, (d) heritability estimates from twin studies may be biased upwards for different methodological reasons (Vinkhuyzen et al., 2012).

SNP-based Heritability Estimates of Neuroticism Facet Scales

The SNP-based heritability estimates at the level of facet scales differed greatly across the two samples, modes of measurement, and sum vs. residual scores, making it difficult to draw uniform general conclusions. There are, however, certain trends in our findings, which we describe below.

First, similarly to the Neuroticism domain level, the SNP-based heritability estimates were generally higher for self-reports – for both sum and residual scores – than for informant-reports at the level of facet scales. In general, this is in line with the findings of a study by Kandler and colleagues (2010), who showed that, while self-rater-specific variance was substantially influenced by genetic factors, peer-report method factors showed smaller genetic influences, with the non-shared environment explaining most of the remaining variance. Differences in heritability estimates between self- and informant-reports may also be due to the fact that ”genetic effects on the self-report method factor may reflect genetic effects on response distortions (e.g. response styles, self-enhancement)” or that “… self-reports may be partially based on information that is not accessible to peers or weighted less in peers’ personality judgments (like motives, mental states)“ (Riemann & Kandler, 2010; p. 263). For example, Mõttus et al. (2016) reported that 39% of individual differences in a specific response bias, acquiescent responding, could be explained by genetic variance; self-report method biases may be as heritable as trait scores themselves. Last but not least, it is also possible that the SNP-based heritability estimates were generally higher for self-reports than for other-reports simply

because people see themselves as more neurotic compared to how they are seen by other people (Allik et al., 2010) or that the self is more accurate than others in judging Neuroticism and its facet scales, as proposed by the self–other knowledge asymmetry (SOKA) model (Vazire, 2010).

Second, across both methods of measurement (self- and informant-reports) and cohorts (EGCUT and Lifelines, where applicable), SNP-based heritability estimates > 0% were observed for the Anxiety (N1), Angry Hostility (N2), Depression (N3), and Vulnerability (N6) facet sum scores. However, only three of these estimates (i.e., EGCUT informant-reported Anxiety (N1) and Lifelines self-reported Angry Hostility (N2) and Vulnerability (N6)) were statistically significant at $p = .05$, and there were also noticeable inconsistencies in our GREML-GCTA results, across cohorts and/or self- and informant-reports. When in the EGCUT self-reports, for instance, the SNP-based heritability of Impulsiveness (N5) sum scores was as high as 19.8% ($p = .040$), the same estimate was 0 (n.s.) in the EGCUT informant-reports, and a low 3.3% ($p = .193$) in the Lifelines self-reports. These inconsistent findings are particularly intriguing, as the genetic correlations between the SNP-based estimates for self- and informant-ratings for different Neuroticism sum scores (both domain and facet scores) in the EGCUT sample were all near unity (except for Vulnerability (N6), rG = .91), indicating that, by and large, the same genes seem to explain the genetic variance in self- and informant-ratings of Neuroticism. Overall, these results suggest that SNP-based heritability estimates are not very robust at a finer level of measurement of Neuroticism. Of course, it must be noted that the statistical power was very low for these analyses (less than 30% for a heritability of 15%).

Third, SNP-based heritability estimates were considerably lower for residual facet scales than for sum scores, for both self- and informant-reports. This means that, if you remove the "general Neuroticism genetic effect," there is little specific genetic effect left in facet-level scores. In other words, the heritability estimates of facet scores discussed in the paragraph above were largely due to the heritable variance they shared with other facets—something that could be interpreted as the general domain of Neuroticism. For instance, in both self- and informant-reports, a modest portion of the specific variance in the Neuroticism facet scales Anxiety (N1) and Angry Hostility (N2) was still attributable to genetic influence (8% to 17%), implying that there might be unique genetic influences underlying these facets (cf. McCrae, 2015).

Fourth, we found a significant gene-environment interaction effect ($p < .05$) in self-reported Impulsivity (N5) sum scores, suggesting that genetic effects on Impulsivity scores differ across cohorts.

Limitations and Conclusions

Despite several strengths, as outlined above, our study also has some notable limitations. The main limitation is restricted statistical power of some of the analyses conducted because of the relatively small sample size of the EGCUT cohort (as evinced by the large standard errors of the SNP-based heritability estimates), as GREML-GCTA requires very large samples to detect genetic effects (J. J. Lee & Chow, 2014; Visscher et al., 2014). Another major limitation is that the Lifelines cohort only had self-reported personality data for four Neuroticism facet scales, meaning that we could not repeat all analyses (self- vs. informant-reports, sum vs. residual scores) in both cohorts. As far as we are aware, this is the first study to estimate the heritability

of personality traits with both self- and informant-reports by using the GREML-GCTA approach, but as our findings are limited to a single cohort, it is impossible to generalize these effects.

In sum, if we accept that individual differences in Neuroticism are heritable to a larger extent than 15%—as demonstrated by behavioral genetics studies—our findings may indicate that a large proportion of the heritability of Neuroticism can be explained by other factors such as non-additive or due to rare genetic variants. In both cases, the genetic mechanisms of the trait are complex and possibly quite idiosyncratic: very different combinations of genes might be implicated, even in similar levels of the phenotypic traits.

Our results also show that, although there may be substantial differences in the heritability of Neuroticism facets—supporting a more "nuanced" approach to Neuroticism as recently advocated by McCrae (2015)—these differences are not robust, in that they do not replicate across different cohorts or modes of measurement. Even more, we found a significant gene-environment interaction effect for the Neuroticism facet scale Impulsiveness (N5), indicating that different genetic markers influence this Impulsiveness score in different populations. Thus, it is perhaps time to consider the possibility that "there actually is not a large set of neuroticism genes, each with small effect" but "there is merely a nonspecific genetic background to phenotypic neuroticism, and to its phenotypic causes and effects" (Turkheimer et al., 2014; p. 536). How this nonspecific genetic background becomes manifest in the phenotypic characteristics that we aggregate into the facets of Neuroticism, and the broad Neuroticism itself, may vary across people and the point of view from which the characteristics are judged.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article: Preparation of this manuscript was supported by the University of Tartu (SP1GVARENG) and by institutional research funding (IUT2-13) from the Estonian Ministry of Education and Science to Jüri Allik. Part of the statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is supported by funding from the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. EGCUT work was supported through the Estonian Genome Center of University of Tartu by a grant from the Estonian Ministry of Education and Science (SF0180142s08); the Development Fund of the University of Tartu (grant SP1GVARENG); the European Regional Development Fund to the Centre of Excellence in Genomics (EXCEGEN; grant 3.2.0304.11-0312); and through a FP7 grant no 313010.

Acknowledgements

Anu Realo was a Visiting Professor in the Health, Medical, and Neuropsychology Department of the Faculty of Social and Behavioral Sciences at Leiden University during the writing of this manuscript. She was also supported by a grant from the Netherlands Institute for Advanced Study (NIAS) during the preparation of the first draft of this manuscript.

We are grateful to the Estonian Genome Centre of the University of Tartu and its director, Andres Metspalu, for their help in collecting the data and for their kind permission to use the data in the current study. We also thank Delaney Michael Skerrett for his helpful comments on an earlier version of this manuscript. The authors also acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

This research was approved by the Research Ethics Committee of the University of Tartu (approvals: 236/M-29, 14 May 2014; 206/T-4, 22 Aug 2011; 170/T-38, 28 April 2008; 166/T-21, 17 Dec 2007).

Correspondence concerning this article should be addressed to Anu Realo, Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom. Electronic mail may be sent to a.realo@warwick.ac.uk.

References

Allik, J., Borkenau, P., Hrebícková, M., Kuppens, P., & Realo, A. (2015). How are personality trait and profile agreement related? *Frontiers in Psychology, 6(785)*. doi:10.3389/fpsyg.2015.00785

Allik, J., Realo, A., Mõttus, R., Borkenau, P., Kuppens, P., & Hrebickova, M. (2010). How people see others is different from how people see themselves: A replicable pattern across cultures. *Journal of Personality and Social Psychology, 99*(5), 870-882.

Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., . . . McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature, 467*(7311), 52-58.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001). Genetic and environmental influences on observed personality: Evidence from the German Observational Study of Adult Twins. *Journal of Personality and Social Psychology, 80*(4), 655-668.

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

Briley, D. A., & Tucker-Drob, E. M. (2012). Broad bandwidth or high fidelity? Evidence from the structure of genetic and environmental effects on the facets of the Five Factor Model. *Behav Genet, 42*(5), 743-763.

Briley, D. A., & Tucker-Drob, E. M. (2014). Genetic and environmental continuity in personality development: A meta-analysis. *Psychological Bulletin, 140*(5), 1303-1331.

Calboli, F. C., Tozzi, F., Galwey, N. W., Antoniades, A., Mooser, V., Preisig, M., . . . Balding, D. J. (2010). A genome-wide association study of neuroticism in a population-based sample. *PLoS One, 5*(7), e11504.

Cloninger, C. R., Przybeck, T. R., & Svrakic, D. M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. *Psychol Rep, 69*(3 Pt 1), 1047-1057.

Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Arch Gen Psychiatry, 50*, 975-990.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092-1122.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15*(1), 110-117.

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.

Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment* (Vol. 2: Personality Measurement and Testing, pp. 179-199). London: SAGE Publications Ltd.

Cuijpers, P., Smit, F., Penninx, B. W., de Graaf, R., ten Have, M., & Beekman, A. T. (2010). Economic costs of neuroticism: a population-based study. *Arch Gen Psychiatry, 67*(10), 1086-1093.

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., . . . Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry, 20*(2), 183-192.

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., Costa, P. T., Jr., & 43 Collaborators of the Adolescent Personality Profiles of Cultures Project. (2009). Assessing the universal structure of personality in early adolescence the NEO-PI-R and NEO-PI3 in 24 cultures. *Assessment, 16*(3), 301-311.

De Fruyt, F., Van De Wiele, L., & Van Heeringen, C. (2000). Cloninger's psychobiological model of temperament and character and the Five-Factor Model of personality. *Personality and Individual Differences, 29*, 441-452.

de Moor, M. H. M., Costa, P. T., Jr., Terracciano, A., Krueger, R. F., de Geus, E. J. C., Toshiko, T., . . . Boomsma, D. I. (2012). Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry, 17*(3), 337-349.

de Moor, M. H. M., van den Berg, S. M., Verweij, K. J., Krueger, R. F., Luciano, M., Arias Vasquez, A., . . . Boomsma, D. I. (2015). Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry, 72*(7), 642-650.

Eysenck, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: University of London Press.

Eysenck, H. J. (1998/1947). *Dimensions of personality*. New Brunswick: Transaction Publishers.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism Scale. *Personality and Individual Differences, 6*, 21-29.

Freeman, C., & Marchini, J. (2012). GTOOL (Version 0.7.5). Retrieved from http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html

Friedman, H. S., & Kern, M. L. (2014). Personality, well-being, and health. *Annual Review of Psychology, 65*, 719-742.

Heath, A. C., Neale, M. C., Kessler, R. C., Eaves, L. J., & Kendler, K. S. (1992). Evidence for genetic influences on personality from self-reports and informant ratings. *Journal of Personality and Social Psychology, 63*(1), 85-96.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *NEO Persoonlijkheidsvragenlisten NEO PI-R en NEO-FF-I. Handleigin. [Manual of the Dutch version of the NEO PI-R and NEO-FFI]*. Lisse, the Netherlands: Swets & Zeitlinger.

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of Genome-Wide Association Studies. *PLoS Genet, 5*(6), e1000529.

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

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

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*(3), 577-591.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*(6), 1556-1565.

Jeronimus, B. F., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history but the adjusted association hardly decays with time: A meta-analysis on 59 longitudinal/prospective studies with 443,313 participants. *Psychological Medicine, 46*(14), 2883-2906.

Juhasz, G., Chase, D., Pegg, E., Downey, D., Toth, Z. G., Stones, K., . . . Deakin, J. F. (2009). CNR1 gene is associated with high neuroticism and low agreeableness and interacts with recent negative life events to predict current depressive symptoms. *Neuropsychopharmacology, 34*(8), 2019-2027. doi:10.1038/npp.2009.19

Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the Five-Factor Model. *European Journal of Personality, 14*, 265-278.

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*(5), 1565-1594.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist, 64*, 241-256.

Lee, J. J., & Chow, C. C. (2014). Conditions for the validity of SNP-based heritability estimation. *Hum Genet, 133*(8), 1011-1022. doi:10.1007/s00439-014-1441-5

Lee, S. H., DeCandia, T. R., Ripke, S., Yang, J., The Schizophrenia Psychiatric Genome Wide Association Study, C., The International Schizophrenia, C., . . . Wray, N. R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet, 44*(3), 247-250. doi:10.1038/ng.1108

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics, 28*, 2540-2542.

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., . . . Metspalu, A. (2014). Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol*, 1137-1147. doi:10.1093/ije/dyt268

Llewellyn, C. H., Trzaskowski, M., Plomin, R., & Wardle, J. (2013). Finding the missing heritability in pediatric obesity: The contribution of genome-wide complex trait analysis. *Internal Journal of Obesity, 37*(11), 1506-1509. doi:10.1038/ijo.2013.30

Lubke, G. H., Hottenga, J. J., Walters, R., Laurin, C., de Geus, E. J. C., Willemsen, G., . . . Boomsma, D. I. (2012). Estimating the genetic variance of Major Depressive Disorder (MDD) due to all SNPs. *Biological psychiatry, 72*(8), 707-709. doi:10.1016/j.biopsych.2012.03.011

Lubke, G. H., Laurin, C., Amin, N., Hottenga, J. J., Willemsen, G., van Grootheest, G., . . . Boomsma, D. I. (2014). Genome-wide analyses of borderline personality features. *Moleular Psychiatry, 19*(8), 923-929. doi:10.1038/mp.2013.109

Luciano, M., Wainwright, M. A., Wright, M. J., & Martin, N. G. (2006). The heritability of conscientiousness facets and their relationship to IQ and academic achievement. *Personality and Individual Differences, 40*(6), 1189-1199.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753.

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*(2), 97-112. doi:10.1177/1088868314541857

McCrae, R. R., & Costa, P. T., Jr. (1992). Discriminant validity of NEOPI-R facet scales. *Educational and Psychological Measurement, 52*, 229-237.

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment, 84*, 261-270.

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2016). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, Epub ahead of print*.

Mõttus, R., Marioni, R., & Deary, I. J. (2015). Markers of psychological differences and social and health inequalities: Possible genetic and phenotypic overlaps. *Journal of Personality*, EPub ahead of print.

Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality, 52*, 47-54. d

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS One, 10*(3).

Munafo, M. R. (2010). Credible genetic associations? *Int J Mol Epidemiol Genet, 1*(1), 31-34.

Munafo, M. R., & Flint, J. (2011). Dissecting the genetic architecture of human personality. *Trends Cogn Sci, 15*(9), 395-400. doi:10.1016/j.tics.2011.07.007

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., . . . Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet, 48*(6), 624-633. doi:10.1038/ng.3552

Ormel, J., Bastiaansen, A., Riese, H., Bos, E. H., Servaas, M., Ellenbogen, M., . . . Aleman, A. (2013). The biological and psychological basis of neuroticism: Current status and future directions. *Neurosci Biobehav Rev, 37*(1), 59-72. doi:10.1016/j.neubiorev.2012.09.004

Ormel, J., Jeronimus, B. F., Kotov, R., Riese, H., Bos, E. H., Hankin, B., . . . Oldehinkel, A. J. (2013). Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clin Psychol Rev, 33*(5), 686-697. doi:10.1016/j.cpr.2013.04.003

Ormel, J., Riese, H., & Rosmalen, J. G. (2012). Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect? *Clin Psychol Rev, 32*(1), 71-79. doi:10.1016/j.cpr.2011.10.004

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*(3), 462-469. doi: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*(1), 211-218.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Mol Psychiatry, 20*(1), 98-108. doi:10.1038/mp.2014.105

Plomin, R., Haworth, C. M., Meaburn, E. L., Price, T. S., & Davis, O. S. (2013). Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol Sci, 24*(4), 562-568. doi:10.1177/0956797612457952

Power, R. A., & Pluess, M. (2015). Heritability estimates of the Big Five personality traits based on common genetic variants. *Transl Psychiatry, 5*, e604. doi:10.1038/tp.2015.96

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. d

Purcell, S. (2009). PLINK (Version v1.07). Retrieved from http://pngu.mgh.harvard.edu/purcell/plink/

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK: A Toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet, 81*, 559-575.

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.

Riemann, R., & Kandler, C. (2010). Construct validation using multitrait-multimethod-twin data: The case of a general factor of personality. *European Journal of Personality, 24*, 258-277.

Schinka, J. A., Busch, R. M., & Robichaux-Keene, N. (2004). A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Mol Psychiatry, 9*(2), 197-202. doi:10.1038/sj.mp.4001405

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., . . . Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol, 44*(4), 1172-1180. doi:10.1093/ije/dyu229

Sen, S., Burmeister, M., & Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *American Journal of Medical Genetics. Part B: Neuropsychiatric Genetics, 127b*(1), 85-89. doi:10.1002/ajmg.b.20158

Service, S. K., Verweij, K. J. H., Lahti, J., Congdon, E., Ekelund, J., Hintsanen, M., . . . Freimer, N. B. (2012). A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales. *Translational Psychiatry, 2*(5), e116. doi:10.1038/tp.2012.37

Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E., Colodro Conde, L., Ward, J., . . . O'Donovan, M. (2016). Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry, in press*. doi:10.1101/032417

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65.

The Genomes Project, C. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. doi:10.1038/nature15393

Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Hum Genomics, 9*, 17. doi:10.1186/s40246-015-0041-3

Tropf, F. C., Verweij, R. M., van der Most, P. J., Stulp, G., Bakshi, A., Briley, D. A., . . . Mills, M. C. (2016). Mega-analysis of 31,396 individuals from 6 countries uncovers strong gene-environment interaction for human fertility. *Manuscript under review, bioRxiv*. doi:http://dx.doi.org/10.1101/049163

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology, 65*(1), 515-540. doi:doi:10.1146/annurev-psych-113011-143752

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 Item Response Theory. *Behav Genet, 44*(4), 295-313. doi:10.1007/s10519-014-9654-x

Vazire, S. (2010). Who knows what about a person? The Self–Other Knowledge Asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300.

Verweij, K. J. H., Yang, J., Lahti, J., Veijola, J., Hintsanen, M., Pulkki-Råback, L., . . . Zietsch, B. P. (2012). Maintenance of genetic variation in human personality: Testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. *Evolution: International Journal of Organic Evolution, 66*(10), 3238-3251. doi:10.1111/j.1558-5646.2012.01679.x

Viken, R. J., Rose, R. J., Kaprio, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: extraversion and neuroticism from 18 to 59 years of age. *J Pers Soc Psychol, 66*(4), 722-730.

Vinkhuyzen, A. A. E., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K. E., Iacono, W. G., . . . Wray, N. R. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational Psychiatry, 2*. doi:10.1038/tp.2012.27

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A., Chen, G. B., Lee, S. H., Wray, N. R., . . . Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet, 10*(4), e1004269. doi:10.1371/journal.pgen.1004269

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin, 141*, 769–785.

Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K., . . . Jang, K. L. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of Personality and Social Psychology, 90*(6), 987-998. doi:10.1037/0022-3514.90.6.987

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A., Lee, S. H., . . . Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet, 47*(10), 1114-1120. doi:10.1038/ng.3390

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., . . . Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet, 42*(7), 565-569. doi:10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet, 88*(1), 76-82. doi:10.1016/j.ajhg.2010.11.011

Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., & Price, A. L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet, 9*(5), e1003520. doi:10.1371/journal.pgen.1003520

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A, 109*(4), 1193-1198. doi:10.1073/pnas.1119675109

Table 1

Descriptive Statistics of the NEO PI-3 (EGCUT) and the NEO PI-R (Lifelines) Neuroticism Scales (Sum Scores)

		EGCUT							Lifelines		
		Self-Ratings *N*=3,292 (59.3% female, mean age =47.2 (*SD* = 17.0)			Informant-Ratings *N*=3,151 (71.1% female, mean age = 42.1 (*SD* = 16.0)				Self-Ratings *N*=13,383, 58.2% female, mean age 48.8 (*SD* = 11.4)		
Scales	*n*	*M*	*SD*	*α*	*M*	*SD*	*α*	*r*	*M*	*SD*	*α*
N1: Anxiety	8	16.02	5.99	.81	15.38	5.92	.82	.51	-	-	-
N2: Angry Hostility	8	13.21	5.27	.78	13.37	6.04	.83	.49	10.74	4.26	.74
N3: Depression	8	14.20	5.61	.77	13.59	5.26	.77	.51	-	-	-
N4: Self-Consciousness	8	14.73	5.22	.72	13.25	4.90	.72	.45	11.51	4.64	.78
N5: Impulsiveness	8	16.64	4.86	.67	15.46	5.19	.70	.42	14.19	3.80	.61
N6: Vulnerability	8	10.71	4.68	.78	9.92	5.24	.83	.43	10.35	4.14	.80
Neuroticism domain	48	85.51	24.50	.93	80.98	25.28	.93	.53	-	-	-
N-32: N2, N4, N5, N6 items	32	55.29	15.57	.88	52.00	16.99	.90	.49	46.74	12.65	.87

Note. EGCUT = Estonian Genome Centre of the University of Tartu; Lifelines = The Lifelines Cohort Study; *n* = number of items; *r* = cross-rater correlation; *α* = Cronbach alpha; As Lifelines used an item scale of 1-5, whereas EGCUT 0-4, we subtracted the number of items (n) from the Lifelines averages to make them comparable to EGCUT.

Table 2

SNP-based Estimates of Heritability for Sum and Residual Facet Scores of Neuroticism in Self- and Informant-Ratings

	EGCUT								Lifelines	
	Self-Ratings[a]				Informant-Ratings[b]				Self-Ratings[c]	
	Sum (*N* = 2,836)		Residual (*N* = 2,833)		Sum (*N* = 2,721)		Residual (*N* = 2,717)		Sum (*N* = 9,469)	
Scales	h^2	*p*	h^2	*p*	h^2	*p*	h^2	*p*	h^2	*p*
N1: Anxiety	0.145	.091	0.147	.106	0.175	.050	0.078	.232		
N2: Angry Hostility	0.083	.210	0.121	.128	0.151	.095	0.168	.073	0.161	.000
N3: Depression	0.161	.054	0.000	.500	0.080	.236	0.129	.139		
N4: Self-Consciousness	0.120	.116	0.003	.489	0.000	.500	0.000	.500	0.071	.029
N5: Impulsiveness	0.198	.040	0.069	.279	0.000	.500	0.000	.500	0.033	.193
N6: Vulnerability	0.168	.058	0.000	.498	0.058	.313	0.033	.388	0.152	.000
Neuroticism domain	0.152	.070	0.118	.123	0.062	.293	0.079	.242		
N-32 (N2+N4+N5+N6)	0.134	.098			0.015	.448			0.110	.002

Note. EGCUT = Estonian Genome Centre of the University of Tartu; Lifelines = The Lifelines Cohort Study; Sum = sum scores of items; Residual = residual facet scores; h^2 = SNP-based heritability estimate, *p* = p-value; Standard errors (*SE*) of the SNP-based heritability estimates were 0.11-0.12 for the EGCUT sample and 0.04 for the Lifelines sample.

Table 3

SNP-based Estimates of Gene-Environment Interaction in the EGCUT and the Lifelines Combined Cohort

Scales	Interaction effect		Genetic effect	
	G x E	*p*	h^2	*p*
N2: Angry Hostility	0.087	.064	0.090	.023
N4: Self-Conscientiousness	0.000	.500	0.094	.015
N5: Impulsivity	0.107	.033	0.001	.491
N6: Vulnerability	0.049	.203	0.103	.014
N-32 (N2+N4+N5+N6)	0.082	.081	0.068	.069

Note. EGCUT = Estonian Genome Centre of the University of Tartu; Lifelines = The Lifelines Cohort Study; *G x E* = gene and environment interaction effect (the environment factor is the cohort of origin); h^2 = SNP-based heritability estimate, *p* = p-value. *p*-values are based on likelihood-ratio tests comparing the full model with a model constraining the particular effect to be zero. The Neuroticism scores of both cohorts were corrected separately (using sex, age, PC1-10 and genotyping platform (EGCUT only) as covariates). *SE* of the interaction effect was 0.058-0.060 and of the genetic effect 0.046-0.048.

Footnotes

[1] The Neuroticism score for the NTR cohort was composed of the items measuring psycho-neuroticism by the Amsterdamse Biografische Vragenlijst (which is mainly based on the Maudley Personality Inventory by Eysenck (1959) and by the NEO-FFI Neuroticism items). For the QIMR cohort, the harmonized Neuroticism score was obtained across the Neuroticism items of the Eysenck Personality Questionnaire, the NEO-FFI, the Temperament and Character Inventory, and the Multidimensional Personality Questionnaire.

[2] Neuroticism and Extraversion scores were derived from the Eysenck Personality Questionnaire, the NEO-FFI, the International Personality Item Pool, and the Multidimensional Personality Questionnaire.

[3] For calculating residual facet scores (see below), a sample of 3,345 individuals (59.3% female) was used with a mean age of 46.6 years (*SD* = 17.0, ranging from 18 to 91 years).

[4] This was done with the aim of reducing respondent burden and to limit overlap between the omitted Neuroticism facets and anxiety and depressive symptoms as assessed with the MINI International Neuropsychiatric Interview.

[5] Hu and Bentler (1999) suggested that a cut-off value close to or greater than .95 for CFI and a cut-off value close to .06 for RMSEA are needed in order to conclude that the model fits the observed data relatively well. However, as CFI tends to decrease in models with a large number of variables, even when the model is correctly specified (cf. Mõttus et al., 2014), we decided that a cut-off value of .90 for CFI (together with an RMSEA value close to .06) can be considered indicative of an acceptable model fit.

Figure 1

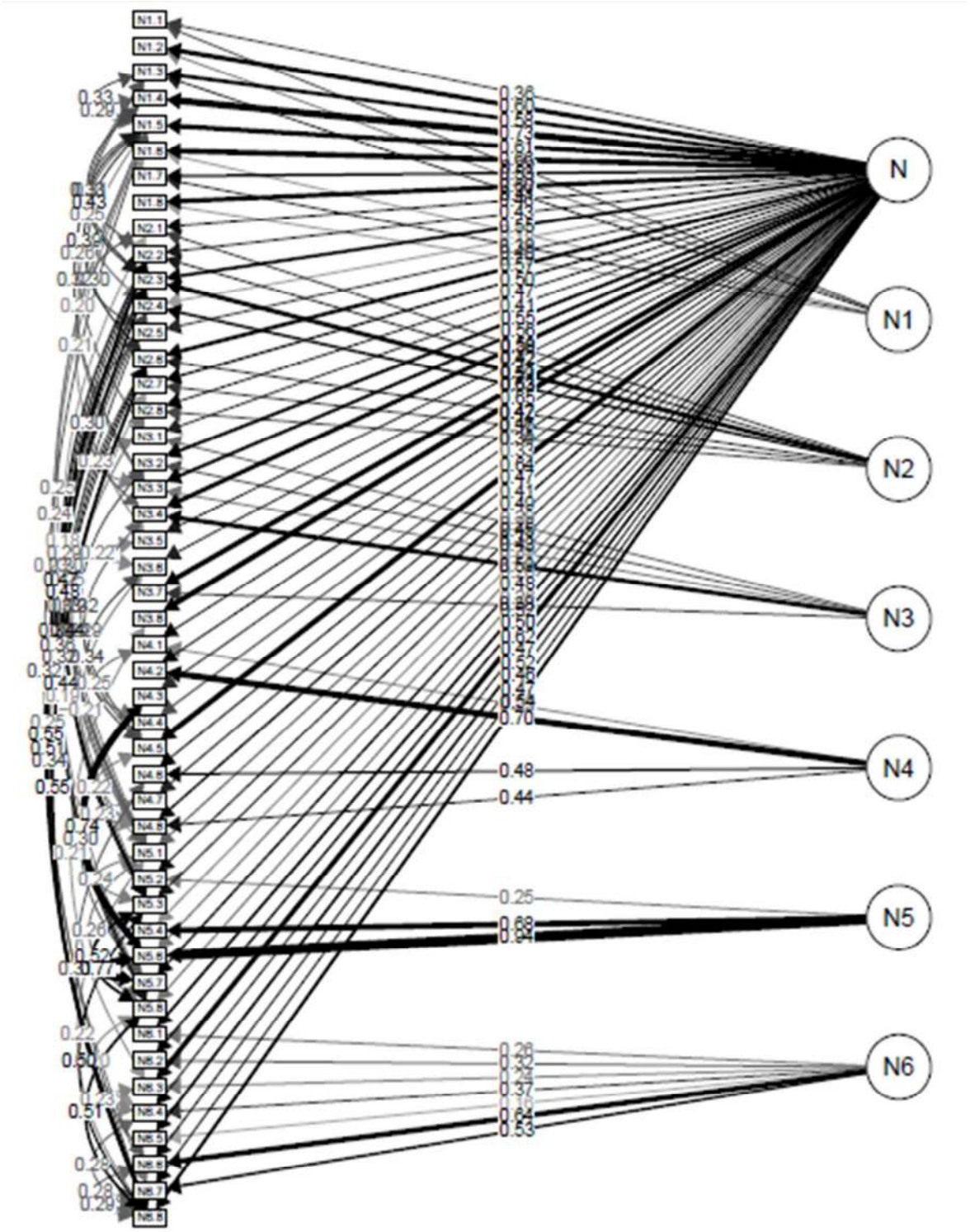

Figure 1. The bi-factor model for N in self-ratings. N1 = Anxiety; N2: Angry Hostility; N3 = Depression; N4 = Self-Consciousness; N5 = Impulsiveness; and N6 = Vulnerability.