

Individual Differences in Personality Masculinity-Femininity: Examining the Effects of Genes, Environment, and Prenatal Hormone Transfer

Karin J. H. Verweij,¹ Miriam A. Mosing,^{1,2} Fredrik Ullén,¹ and Guy Madison³

¹Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

²Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

³Department of Psychology, Umeå University, Umeå, Sweden

Males and females score differently on some personality traits, but the underlying etiology of these differences is not well understood. This study examined genetic, environmental, and prenatal hormonal influences on individual differences in personality masculinity–femininity (M-F). We used Big-Five personality inventory data of 9,520 Swedish twins (aged 27 to 54) to create a bipolar M-F personality scale. Using biometrical twin modeling, we estimated the influence of genetic and environmental factors on individual differences in a M-F personality score. Furthermore, we tested whether prenatal hormone transfer may influence individuals' M-F scores by comparing the scores of twins with a same-sex versus those with an opposite-sex co-twin. On average, males scored 1.09 standard deviations higher than females on the created M-F scale. Around a third of the variation in M-F personality score was attributable to genetic factors, while family environmental factors had no influence. Males and females from opposite-sex pairs scored significantly more masculine (both approximately 0.1 *SD*) than those from same-sex pairs. In conclusion, genetic influences explain part of the individual differences in personality M-F, and hormone transfer from the male to the female twin during pregnancy may increase the level of masculinization in females. Additional well-powered studies are needed to clarify this association and determine the underlying mechanisms in both sexes.

■ **Keywords:** personality, twin study, masculinity, hormone transfer, heritability, genes, sex-differences

Personality has been defined as ‘those characteristics of the person that account for consistent patterns of feeling, thinking, and behaving’ (Pervin et al., 2005, p. 6). Human personality research focuses predominantly on dimensions of variation between individuals that predict behavior in various situations and domains of life and that are relatively stable over time. One of the most influential models that is used to describe human personality is the Five-Factor Model, which structures personality traits into five broad factors: Agreeableness, Conscientiousness, Extraversion, Openness to Experience, and Neuroticism (Costa & McCrae, 1985; Digman, 1990). Previous research has shown that males and females score differently on some of these and other personality scales, but sex differences are generally small to medium; in a cross-national meta-analysis of effect sizes in gender differences in the Big Five facets, the sex differences ranged between $D = 0.02$ and 0.56 (Feingold, 1994). Females tend to score higher on traits associated with anxi-

ety, neuroticism, extraversion, agreeableness, and warmth, whereas males generally score higher on assertiveness and self-esteem, as summarized by various literature reviews and meta-analyses (Costa et al., 2001; Feingold, 1994; Gentile et al., 2009; Kling et al., 1999; Lippa, 2010).

Although there are well-established mean sex differences in these particular traits, there is no consensus whether sex differences in personality arise due to different environmental influences, such as social and cultural expectations regarding male and female's behavior, or whether they are due to underlying biological differences between sexes. An important socio-cultural explanation is the social role model

RECEIVED 21 January 2016; ACCEPTED 27 January 2016.

ADDRESS FOR CORRESPONDENCE: Karin Verweij, Department of Neuroscience, Karolinska Institutet, Retzius v 8, SE-171 77 Stockholm, Sweden. E-mail: karin.verweij@ki.se/karin.verweij@vu.nl

(Eagly, 1987; Eagly & Wood, 1991) that proposes that most differences between the sexes result from the adoption of gender roles that arise from different expectations for men's and women's social roles, particularly in relation to family and occupation. The social role model would therefore predict cultural variation in personality sex differences across countries with more or less traditional sex roles. However, the social role model is contradicted by the finding that sex differences are consistent across 53 nations (Lippa, 2010) and that the sex differences appear more pronounced in Western countries (Costa et al., 2001), in which differences in traditional sex roles are generally minimized and there is more formal and economic equality between the sexes.

Other socio-cultural explanations of sex differences are the expectancy model and the artifact model. The expectancy model (Deaux & Major, 1987) suggests that sex differences in behavior are partly the result of stereotype-based expectations of perceivers that result in self-fulfilling prophecies. However, self-fulfilling prophecy effects are very small, with meta-analytical estimates at around 0.10 (Jussim, 2012). The artifact model, finally, proposes that social desirability may cause males and females to endorse gender-appropriate personality scores (Feingold, 1990, 1991, 1992). According to this model, males and females place different values on the importance of certain personality traits and these differences differentially bias self-reports of personality characteristics. As such, according to this model, sex differences in personality scores reflect differences in social desirability responding rather than real differences in personality.

In contrast to socio-cultural models, biological and evolutionary psychology theories propose that sex differences in personality are partly due to innate temperamental differences that can be expected for all behaviors for which males and females have faced different adaptive pressures over the evolutionary past (Buss, 1995; Buss & Hawley, 2010; Geary, 2010; Trivers, 1972). Sex differences in personality traits could be explained by the parental investment theory (Trivers, 1972). This theory predicts that women have been under selection to invest more time and resources in their offspring, which makes them choosier in mate selection and more cautious and careful in social relations than males. In turn, males have faced greater pressure to compete for and attract mates, which favors assertiveness, aggressiveness, and competitiveness. Thus, personality traits linked to these behaviors have come under sex-specific selection, and this may explain why females are more agreeable, warm and nurturing, and less assertive and competitive than males. Such a common underlying selection pressure for male-female sex differences would suggest a common mechanism for their implementation, even if they manifest themselves in several personality dimensions.

In this perspective, sex differences in personality could be due to hormonal influences. Testosterone levels are 2–4 times higher in the amniotic fluid of male versus female

fetuses (Auyeung et al., 2009) and during puberty levels are 20- to 30-fold higher in males than females (Fechner, 2003). Prenatal exposure to high levels of testosterone have been found to cause masculinization of activity and occupational interests (Berenbaum & Beltz, 2011), and circulating testosterone levels during adolescence seem to play a role in aggression (e.g., Pajer et al., 2006).

One approach by which prenatal hormonal exposure can be examined is by studying twin pairs. It has been hypothesized that hormone transfer from one twin to the other during pregnancy may influence the level of masculinization (Miller, 1994); that is, individuals with a male co-twin may develop more masculine behaviors than individuals with a female co-twin and the other way around. Two mechanisms of hormone transfer have been proposed: hormones from one twin could be transferred to the other twin either through maternal circulation (Miller, 1994) or directly from one twin to the other through diffusion across fetal membranes (Even & vom Saal, 1992). Note that research has predominantly focused on testosterone transfer, because this is believed to be the most potent androgen and has shown strong effects in non-human species (Ryan & Vandenbergh, 2002), while other hormones remain understudied in this context. Hence, studies examining the effect of hormone transfer on males are limited.

While in humans relatively consistent evidence has been found supporting the hormone transfer hypothesis for perception and cognition (especially for females), support for the influence of hormone transfer on personality is inconsistent (for a review, see Tapp et al., 2011). For example, Resnick et al. (1993) and Slutske et al. (2011) found that females with a male co-twin scored on average higher on sensation seeking than females with a female co-twin. Other studies, however, did not find evidence for prenatal hormone transfer effects on behavioral traits, including personality/temperament (Cohen-Bendahan et al., 2005; Loehlin & Martin, 2000) and sex-typed childhood play (Henderson & Berenbaum, 1997; Rodgers et al., 1998).

Berenbaum and Beltz (2011) reviewed the literature of different types of studies on the influence of prenatal and postnatal sex-hormone exposure on behavior. They concluded that prenatal exposure to high levels of androgens is associated with masculinization of occupational and activity interests, sexual orientation, and spatial abilities, whereas evidence for an influence of postnatal hormone exposure is not as strong. However, they did not find much evidence of hormone influences on sex differences in personality traits and social behaviors (Berenbaum & Beltz, 2011).

Because of the confounding effects of genetic and societal influences, it is hard to identify the source of the between-sex differences in personality. However, we can determine the source(s) of within-sex variation in the sex-differentiating dimension of personality by using a genetically informative sample, and we can determine whether the same or different genes or environmental factors influence level of

masculinization in males versus females. Accordingly, the aim of the current study was to examine the etiology of individual differences in personality M-F. To this end, we created a M-F personality scale by performing a discriminant analysis on the 44-item Big-Five personality inventory on 9,520 Swedish twin individuals. Because of our large sample we were able to use all single personality items in one discriminant analysis without the risk of overfitting, which can result from a high ratio of predictors to participants. Using single item scores instead of the overarching dimensions is preferable as aggregating at the level of the dimensions can mute or even annihilate sex differences as an effect of component traits cancelling each other out (Del Giudice et al., 2012). For example, extraversion loads on 'warmth', which is higher in females, as well as 'dominance', which is higher in males. Accordingly, accurate assessment of sex differences therefore requires analysis at the primary trait level or lower (Del Giudice et al., 2012).

We applied biometrical twin modeling to estimate the influence of genetic and environmental factors on individual differences in the derived M-F score. Based on previous twin studies using various personality inventories and indices of M-F, we expect the M-F personality score to be moderately heritable (previous heritability estimates were in the range of 30–60%, Lippa & Hershberger, 1999; Loehlin & Martin, 2000; Loehlin et al., 2005), similarly to the heritability of personality itself (Johnson et al., 2008). However, previous studies had substantially smaller sample sizes (Lippa & Hershberger, 1999; Mitchell et al., 1989) or used less ideal methods to create the personality M-F scales (Loehlin & Martin, 2000; Loehlin et al., 2005). The large sample size enabled us to estimate the variance components separately for males and females, and provided power to detect non-additive genetic effects.

Second, we examined potential influence of prenatal hormone transfer on individuals' M-F personality scores by comparing the scores of twins with a same-sex co-twin with the scores of twins with an opposite-sex co-twin. If hormone transfer would play a role we would expect twins with a male co-twin to score more masculine than twins with a female co-twin.

Methods

Participants

Between 2012 and 2013 a large cohort of approximately 32,000 Swedish twins born between 1959 and 1985 (the STAGE Cohort, see Lichtenstein et al., 2006) were invited to complete a web-based survey designed to collect data on music-related traits. In total, 11,543 twins participated in the web survey, and their age was between 27 and 54 (mean = 40.7; *SD* = 7.8). However, due to missing data and zygosity, the effective study sample used for this study is slightly lower, as described in the results section. Zygosity determination was based on a questionnaire about intra-

pair resemblance. In the Swedish Twin Registry (STR), this method has been confirmed in 27% of the twins using genotyping and showed an accuracy of more than 98%. For further information on this survey, the STAGE cohort, or zygosity determination in the STR, see Lichtenstein et al. (2002), Lichtenstein et al. (2006) and Mosing et al. (2014). All participants gave informed consent and the study received approval from the Regional Ethics Review Board in Stockholm (Diary Numbers 2011/570-31/5, 2011/1425-31, and 2012/1107/32).

Measures

As part of the web survey, participants filled out the Swedish translation of the Big Five Inventory, a 44-item self-report inventory measuring the Big-Five dimensions of personality, that is, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (BFI; John et al., 1991; John et al., 2008). The participants were asked to indicate on a 5-point Likert scale (ranging from *disagree strongly* to *agree strongly*) to what extent certain characteristics applied to them. The BFI is a commonly used personality questionnaire and previous research has shown that the 44-item version has a test-retest reliability of more than 0.80 and an acceptable external validity of 0.56 (Rammstedt & John, 2007). Comparisons between paper-and-pencil personality questionnaires and online surveys like the present showed that both types of assessment are highly comparable (Lang et al., 2011; Pettit, 2002; Rammstedt et al., 2004). For the current study, single item scores for all items were used to obtain a measure of M-F personality for each individual, as described below.

Statistical Analysis

Discriminant-Function Analysis

To compute a data-driven single measure of M-F personality, we conducted a discriminant-function analysis (DFA) in SPSS (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp) with sex as the grouping variable (male = 0, female = 1). Based on the responses to the personality questionnaire, this analysis generated a discriminant function that was the linear weighted combination of personality items that optimally classified individuals as male or female. The discriminant function from this analysis represents the bipolar M-F dimension. Accordingly, for each individual, a M-F personality score was derived by multiplying the standardized coefficients by the standardized item scores and adding across all items. The positive pole of the scale represents the masculine end and the negative pole the feminine end. To determine whether hormone transfer may play a role in individuals' M-F personality score we tested for mean differences between twins with an opposite-sex co-twin versus those with a same-sex co-twin (correcting for age effects and relatedness of the sample).

Genetic Analysis

With the classical twin design we determined the extent to which individual differences in the M-F personality scores were due to genetic and environmental influences, by decomposing the variance into additive genetic (A), non-additive genetic (D), shared environmental (C), and residual (E) influences. Additive genetic variance is the influence of the summed allelic effects, while non-additive genetic effects include allelic interactions within and across genes (dominance and epistasis). Shared environmental variance results from environmental influences shared within pairs, which make them more similar to each other, such as family environment. Residual variance results from influences not shared by twin pairs, including environmental influences not shared between twins, stochastic biological effects, as well as measurement error.

The classical twin design makes use of the fact that identical (monozygotic, MZ) twins share 100% of their genes, whereas non-identical (dizygotic, DZ) twins on average only share 50% of their segregating genes. If A were the only source of variance in a trait we would expect a twin pair correlation of 1 for MZ pairs, while for DZ pairs the twin correlation would be 0.5. If non-additive genetic influences were the sole source of variance in a trait, we would expect a twin pair correlation of 1.0 for MZ pairs and, at most, 0.25 for DZ pairs (for an explanation, see Posthuma et al., 2003). In contrast, if C were the only source of variance in a trait, by definition we would expect a twin pair correlation of 1 for both MZ and DZ twin pairs. Finally, if all variance were due to E we would expect a twin pair correlation of 0 for both MZ and DZ twin pairs. Hence, A, C, D, and E influences predict different patterns of MZ and DZ twin pair correlations, and we used structural equation modeling to determine which combination best matched our observed data.

It is important to note that it is not possible to estimate C and D simultaneously when including only twins reared together, as C and D are negatively confounded: C decreases the MZ-DZ correlation ratio, while D increases it. Only three of the four sources of variance can therefore be estimated at a time, the choice of which (i.e., an ACE or ADE model) depends on the pattern of MZ and DZ correlations. When DZ twin correlations are at least half the MZ correlation, shared environmental influences are implied and so an ACE model is fitted. If DZ twin correlations are less than half the MZ correlations, non-additive genetic influences are implied and an ADE model is more suitable. A second limitation is that the classical twin design provides little statistical power to disentangle non-additive from additive genetic effects, because they are partly confounded as they predict similar but not identical patterns of MZ versus DZ twin pair correlations. However, as shown by Keller et al. (2010) the broad sense heritability of a trait (H^2 ; i.e., the total proportion of variance accounted for by genetic factors (i.e., A + D) is quite robustly estimated using only

twins reared together. Additional information on the classical twin design can be found in Neale and Cardon (1992) and Posthuma et al. (2003).

Twin analyses were conducted using maximum likelihood procedures in the statistical package Mx (Neale et al., 2006). In maximum-likelihood modeling, the goodness-of-fit of a model to the observed data is distributed as chi-square (χ^2). To test whether dropping model parameters or constraining parameters to be equal significantly worsened the model fit, we tested the change in chi-square ($\Delta\chi^2$) against the change in degrees of freedom (Δdf). Variance components were estimated separately for males and females, and for all analyses age effects on the means were accounted for by including age as a covariate.

Results

Of the 11,543 twins who participated in the survey, those who had an unknown zygosity (368), did not fill out the personality questionnaire (1,555), or only partly filled out the personality questionnaire (100) were excluded from further analyses. Accordingly, the final sample consisted of 9,520 participants, including 2,245 complete twin pairs — 695 MZ female, 374 MZ male, 392 DZ female, 248 DZ male, and 536 DZ opposite-sex pairs, and 5,030 single twins without a participating co-twin. Single twins were included as they contribute to the estimation of mean and variance effects. Participants were aged between 27 and 54 (mean 40.8, SD 7.8).

Discriminant Analysis

Table 1 shows the item loadings for each of the Big Five Inventory items as obtained from the Discriminant Function Analysis; the standardized coefficients (ranging from -0.56 to 0.30) indicate which of the items have the highest predictor capability of predicting male versus female group membership. Multiplying the standardized coefficients by the standardized variables and adding across all items results in the discriminant score for each participant, with positive values indicating the masculine side of the scale and negative values the feminine side. As shown in Table 1, examples of items that differentiate relatively strongly between males and females are, for example: ‘I am someone who can be cold and aloof’, ‘I am someone who values artistic, aesthetic experiences’, ‘I am someone who is emotionally stable, not easily upset’, and ‘I am someone who is inventive’.

Figure 1 shows the distribution of the discriminant scores for males and females. On average, males scored 1.09 standard deviations higher on the discriminant score than females, indicating a large effect size (males Mean = 0.64 [SD = 1.01], females Mean = -0.45 [SD = 0.99]; Cohen’s d = 1.09, 95% confidence intervals: 1.07–1.11). The canonical correlation between participants’ sex and the discriminant score was 0.47, indicating that the personality items combined can explain 22% of the group membership (Wilks’

TABLE 1
Standardized Canonical Discriminant Function Coefficients for Each Item of the Big-Five Inventory as Obtained from the Discriminant Function Analysis

BFI item number	I am someone who....	Standardized coefficient
1	Is talkative	-0.142
2	Tends to find fault with others	0.104
3	Does a thorough job	-0.044
4	Is depressed, blue	0.078
5	Is original, comes up with new ideas	0.144
6	Is reserved	-0.040
7	Is helpful and unselfish with others	0.052
8	Can be somewhat careless	0.011
9	Is relaxed, handles stress well	0.193
10	Is curious about many different things	-0.029
11	Is full of energy	0.038
12	Starts quarrels with others	-0.064
13	Is a reliable worker	-0.166
14	Can be tense	-0.036
15	Is ingenious, a deep thinker	0.176
16	Generates a lot of enthusiasm	-0.200
17	Has a forgiving nature	0.025
18	Tends to be disorganized	0.123
19	Worries a lot	-0.231
20	Has an active imagination	0.064
21	Tends to be quiet	0.083
22	Is generally trusting	-0.016
23	Tends to be lazy	0.047
24	Is emotionally stable, not easily upset	0.289
25	Is inventive	0.298
26	Has an assertive personality	0.033
27	Can be cold and aloof	0.228
28	Perseveres until the task is finished	0.205
29	Can be moody	-0.076
30	Values artistic, aesthetic experiences	-0.559
31	Is sometimes shy, inhibited	-0.160
32	Is considerate and kind to almost everyone	-0.038
33	Does things efficiently	-0.292
34	Remains calm in tense situations	0.029
35	Prefers work that is routine	-0.065
36	Is outgoing, sociable	-0.031
37	Is sometimes rude to others	0.178
38	Makes plans and follows through with them	-0.205
39	Gets nervous easily	0.002
40	Likes to reflect, play with ideas	-0.011
41	Has few artistic interests	0.128
42	Likes to cooperate with others	-0.090
43	Is easily distracted	0.016
44	Is sophisticated in art, music, or literature	0.204

Note: On the overall discriminant function score males score higher than females.

Lambda = 0.78, $p < .001$). Based on the discriminant function, the sex of 72% of participants could be correctly classified.

To cross-validate our methods, we also ran a Discriminant Function Analysis on half of our sample and used the obtained standardized coefficients to estimate the sex of the other half of the sample. This yielded a canonical correlation between participants' sex and the discriminant score of 0.46 and a correct classification rate of 70%.

Prenatal Hormone Influences on M-F Personality Scores

Males with a male co-twin scored on average 0.61 ($SD = 1.00$) and males with a female-co-twin 0.70 ($SD = 1.01$) on the derived M-F personality scale. Females with a female co-twin scored on average -0.49 ($SD = 0.99$) and females with a male-co-twin -0.38 ($SD = 1.01$). So, both male and female participants from opposite-sex pairs scored signifi-

cantly higher on the M-F personality scale (both approximately 0.1 SD more masculine) than did males and females from same-sex pairs ($\Delta\chi^2_1 = 8.56$, $p = .003$ and $\Delta\chi^2_1 = 9.32$, $p = .002$ for males and females, respectively).

Genetic Analysis

Before modeling the variance components, we tested the effects of age and zygosity on the derived M-F personality scale using an α level of 0.01. The mean scores on the M-F personality scale did not differ significantly between MZ and DZ twins of the same sex ($\Delta\chi^2_2 = 6.40$, $p = .04$), whereas, as mentioned above, opposite sex twins scored significantly higher than same-sex twins. Variances in the M-F scale did not differ significantly between MZ versus DZ twins ($\Delta\chi^2_2 = 0.49$, $p = .78$), or between twins of opposite sex versus same-sex pairs ($\Delta\chi^2_2 = 1.10$, $p = .58$). We found a significant effect of age for both sexes, such

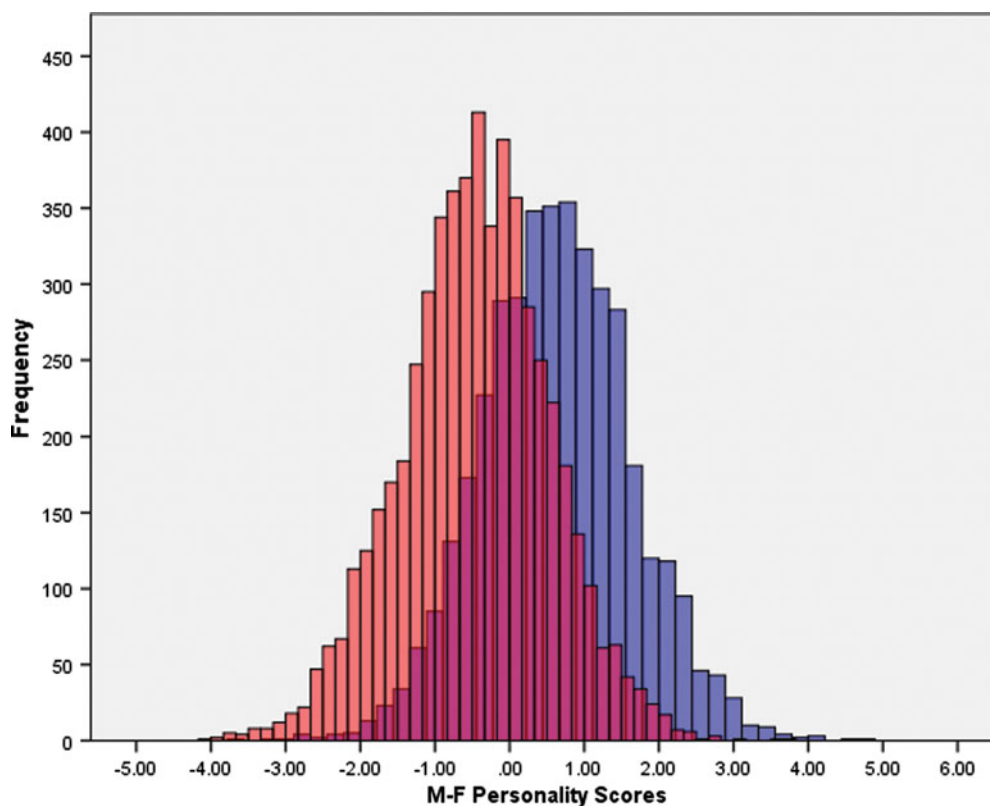


FIGURE 1

(Colour online) Frequency of masculinity-femininity (M-F) personality scores as obtained from the discriminant-function analysis. For males (blue) $M = 0.64$ ($SD = 1.01$) and for females (pink) $M = -0.45$ ($SD = 0.99$). The purple portions of the bars represent overlapping distributions between sexes.

TABLE 2

Twin Pair Correlations (and 95% Confidence Intervals) for the Obtained M-F Personality Score by Zygosity

Zygosity	N pairs	Twin pair correlation
MZ males	374	0.34 [0.25; 0.43]
MZ females	695	0.34 [0.27; 0.40]
DZ males	248	0.18 [0.06; 0.30]
DZ females	392	0.03 [-0.07; 0.13]
DZ opposite sex	536	0.11 [0.02; 0.19]
MZ males and females	1,069	0.34 [0.29; 0.39]
DZ males and females	1,176	0.10 [0.04; 0.16]

that older males scored more feminine than younger males ($\Delta\chi^2_1 = 49.84$, $p < .001$) and older females scored slightly more masculine than younger females ($\Delta\chi^2_1 = 6.93$, $p = .008$). Note that age and sex were included as covariates in subsequent modeling, and means were estimated separately for same-sex versus opposite-sex twins.

Table 2 shows the twin pair correlations for each zygosity group taking age and sex effects into account as obtained by maximum likelihood procedures in Mx. MZ twin pair correlations were higher than DZ twin pair correlations in both sexes, suggesting the influence of genetic factors — this is formally tested below. The DZ opposite-sex twin pair

correlation is not significantly lower than the DZ same-sex twin pair correlations ($\Delta\chi^2_1 = 0.07$, $p = .79$), indicating there are no qualitative sex-differences in sources of familial aggregation between males and females. This means there is no evidence that different genes or different environmental factors influence the M-F scores in males versus females.

Because the DZ twin pair correlation is less than half the MZ twin pair correlation we fitted a non-additive genetic latent variable instead of a shared environmental component. Table 3 shows the A, D, and E parameter estimates, as well as an estimate of the broad-sense heritability (H^2). Results are shown for males and females separately as well as for both sexes combined. Equating the male and female parameter estimates did not result in a significant deterioration of model fit ($\Delta\chi^2_2 = 1.81$, $p = .41$), indicating that the relative influence of genes and environment does not differ significantly between sexes. Estimates indicate that individual differences in M-F personality scores are moderately heritable for both sexes; broad-sense heritability estimates are 35% for males and 33% for females. Also, we found some evidence for non-additive genetic influences, especially for females. The majority of the variance in the derived personality score can be explained by residual

TABLE 3

Estimates of the Proportions of Variance (95 % Confidence Intervals Between Brackets) in Masculinity-Femininity (M-F) Personality Scores Explained by A (Additive Genetic), D (Non-Additive Genetic), and E (Residual) Influences

	A	D	E	H ²
Males	0.29 [0.00; 0.42]	0.07 [0.00; 0.42]	0.65 [0.57; 0.74]	0.35 [0.26; 0.43]
Females	0.04 [0.00; 0.27]	0.29 [0.05; 0.39]	0.67 [0.61; 0.73]	0.33 [0.26; 0.39]
Males and females combined	0.06 [0.00; 0.29]	0.28 [0.03; 0.39]	0.66 [0.61; 0.71]	0.34 [0.29; 0.39]

Note: Note: H² represents the broad sense heritability.

influences, including non-shared environmental influences and measurement error.

Discussion

In the current study, we used data on 9,520 twins to examine the etiology of individual differences in personality M-F. We computed a bipolar M-F personality scale, and examined the extent to which individual differences in this scale could be attributed to genetic and environmental influences. We also tested whether prenatal hormone transfer may influence individuals' M-F personality score.

Males scored on average 1.09 standard deviations higher than females on the derived bipolar M-F personality score. This is a large effect size and indicates an overlap in distributions of approximately 41% between the sexes and means that approximately 86% of males score higher than the average female. This derived sex difference is larger than those reported in other studies (using different methodologies; Costa et al., 2001; Lippa, 2010; Loehlin et al., 1999), but substantially smaller than the multivariate effect size of $D = 2.71$ found by Del Giudice et al. (2012), who used multigroup latent variable modeling to estimate sex differences on 16 individual personality dimensions, which were then aggregated to yield a multivariate effect size taking intercorrelations between the dimensions into account.

We found that individual differences in the M-F score were moderately heritable; broad-sense heritability estimates were 35% for males and 33% for females, with some evidence for a role of non-additive genetic effects, especially for females. Testing for sex-differences in the genetic architecture showed no evidence that different genes or environmental factors influence the M-F scores in males versus females, and also the relative influences of genes and environment do not differ significantly between sexes. Shared environmental factors do not seem to influence individual differences in M-F score, implying factors such as parenting style, socioeconomic status, familial attitudes and values, home environment, and other family environmental factors have very little influence on M-F personality development. Residual influences have the strongest impact on individual differences in M-F score ($E = 65\%$ for males and 67% for females); while part of E will be due to measurement error, this finding also suggests that unique experiences and unique social interactions may play a role.

The estimated broad-sense heritability estimates for M-F scores are similar to those found previously for various indices of masculinity/femininity, which ranged approximately between 30% and 60% (Lippa & Hershberger, 1999; Loehlin & Martin, 2000; Loehlin et al., 2005). These studies generally also found zero or very low influences of shared environment, and Loehlin and Martin (2000) — but not Lippa and Hershberger (1999) or Loehlin et al. (2005) — also found some evidence for non-additive genetic influences. Our estimates of the genetic and environmental influences are also comparable to those of the various Big-Five and other personality traits, for which heritability estimates are in the range of 30–60% and do not indicate much shared environmental influences (Johnson et al., 2008). Moreover, non-additive genetic influences were also detected for several personality traits (Keller et al., 2005).

The presence of heritable variation shows that it may be of interest to identify the specific genes that are involved in individual differences in personality M-F, to gain insight into the underlying biological mechanisms. However, the identification of genes for the typical personality scales has proven difficult (de Moor et al., 2015; Service et al., 2012; van den Berg et al., 2016; Verweij et al., 2010), so very large sample sizes are expected to be required to identify specific variants.

The prenatal hormone transfer hypothesis was addressed by comparing personality scores of twins with a same-sex co-twin with scores of twins with an opposite-sex co-twin. Both male and female twins from opposite-sex pairs scored significantly more masculine (approximately 0.1 *SD*) than males and females from same-sex pairs. For females, this finding is consistent with the hypothesis that testosterone transfer from the male co-twin to the female twin during pregnancy increases the female's level of masculinization. Several previous studies also found that females with a male co-twin exhibited more masculine behavior for disinhibition, experience seeking and overall sensation seeking (Resnick et al., 1993), experience-seeking and thrill-and-adventure-seeking (Slutske et al., 2011), rule-breaking behavior (for one of the two subsamples; Loehlin & Martin, 2000), social conservatism (Miller & Martin, 1995), and aggression (Cohen-Bendahan et al., 2005). The effect size we found is comparable with the effect sizes found by Slutske et al. (2011), and Loehlin and Martin (2000), while the effect sizes reported by Resnick et al. (1993, Cohen's

d ranging between 0.18 and 0.38) and Cohen-Bendahan et al. (2005, Cohen's *d* of 0.34 and 0.49) were substantially larger. There are also studies that did not find support for the hormone transfer hypothesis in females; for instance, for sensation seeking or various temperament subscales (Cohen-Bendahan et al., 2005), toy preference (Henderson & Berenbaum, 1997; Rodgers et al., 1998), feminine interest (Rose et al., 2002), and for Worried and Reserved subscales (Loehlin & Martin, 2000). In some cases, these null results may be due to the much smaller sample sizes; the samples used by Cohen-Bendahan et al. (2005, *N* = 129 twins), Henderson and Berenbaum (1997, *N* = 71 twins), and Rodgers et al. (1998, *N* = 70 female twins) provided low statistical power to detect an effect. Moreover, contrary to expectations, Koopmans et al. (1995) found that female opposite-sex twins scored lower on Experience Seeking than female same-sex twins.

The finding that males with a female co-twin scored more masculine than males with a male co-twin is puzzling and not accounted for by the hormone transfer theory. There are few studies on prenatal hormone transfer in males and findings are inconsistent. Similar to our results, Koopmans et al. (1995) found that male opposite-sex twins scored higher on Experience Seeking than male same-sex twins. Loehlin and Martin (2000) found that males with a female co-twin scored more feminine than males from same-sex pairs on the Worried subscale (in one of two subsamples), but that they also scored more masculine for the Breaks Rules subscale (in both samples), and a trend in this direction for the Worried subscale in the second of the two samples. No differences were found between males with a female versus male co-twin for the Reserved subscale (Loehlin & Martin, 2000), various sensation seeking scores (Resnick et al., 1993), toy preference (Rodgers et al., 1998), feminine interest (Rose et al., 2002), and social conservatism (Miller & Martin, 1995); again, in some cases, these null results may be due to the small sample sizes. Overall, the evidence for prenatal hormone transfer influences on personality is not strong, with study findings especially inconsistent for males.

It is important to note that with the present design it is impossible to differentiate between prenatal hormone and postnatal socialization influences of having a sibling of a certain sex. The higher scores on the M-F scale for males and females with an opposite sex co-twin might therefore be caused by some kind of social interaction after birth or because of a comparison effect (albeit differing for males and females). Slutske et al. (2011) included comparisons between female same-sex twins with and without a non-twin brother (close in age) to distinguish between prenatal hormone transfer versus postnatal socialization influences. Their findings suggested that the masculinization effect of having a male co-twin on females' level of experience seeking and thrill and adventure seeking could not be explained by postnatal socialization effects of having a brother, and

that it therefore had to be attributed to prenatal hormone transfer. Unfortunately, Slutske et al. (2011) did not perform the same analyses for males, so there is no indication of whether higher masculinization for male twins might be an effect of interacting with a female co-twin.

Our study showed that within-sex differences can be explained by genetic, unshared environmental influences and potentially male-to-female hormonal transfer. While the sources of within-sex individual differences are not necessarily the same as the sources of between-sex differences in personality, our findings may inform future work to that end. In that respect, it is of interest that we did not find evidence for qualitative or quantitative sex differences in the sources of variation, indicating that the same genes and environmental factors influence masculinity of personality in both sexes to the same extent. An important strength of the present study is the very large sample size that enabled us to be the first twin study to perform a discriminant analysis using all single personality items without the risk of overfitting. The large sample also provided power to detect subtle hormone-transfer effects and non-additive genetic influences. Another strength of this study is that we also looked for prenatal hormone effects in males, whereas prior studies often focused on prenatal effects in females only.

The main limitation of the current study is that we relied on self-report data, which is subject to response biases such as social desirable responding. To reduce influences of social desirability and measurement error as well as to differentiate between actual sexual dimorphism in personality versus the artefact or expectation theories, future studies should employ different types of personality measurement, such as observational studies or parental, peer, or teacher ratings.

Overall, by means of a discriminant analysis on the single items of the Big Five personality inventory, we created a M-F personality scale with large sex differences, and showed that around one third of the variance on this scale was attributable to genetic factors, while shared environmental factors have no influence. Prenatal hormone transfer may also play a role, but additional well-powered studies are needed to clarify the association and determine the underlying mechanisms in both sexes. By including non-twin siblings these studies should aim to differentiate between prenatal influences and postnatal socialization influences.

Acknowledgments

This work was supported by the Bank of Sweden Tercentenary Foundation (M11-0451:1), the Swedish Scientific Council (521-2010-3195), and the Sven and Dagmar Salén Foundation.

References

- Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., & Hines, M. (2009). Fetal testosterone predicts sexually differentiated childhood

- behavior in girls and in boys. *Psychological Science*, 20, 144–148.
- Berenbaum, S. A., & Beltz, A. M. (2011). Sexual differentiation of human behavior: Effects of prenatal and pubertal organizational hormones. *Frontiers in Neuroendocrinology*, 32, 183–200.
- Buss, D. M. (1995). Psychological sex differences. Origins through sexual selection. *American Psychologist*, 50, 164–168.
- Buss, D. M., & Hawley, P. H. (2010). *The evolution of personality and individual differences*. Oxford: Oxford University Press.
- Cohen-Bendahan, C. C., Buitelaar, J. K., van Goozen, S. H., Orlebeke, J. F., & Cohen-Kettenis, P. T. (2005). Is there an effect of prenatal testosterone on aggression and other behavioral traits? A study comparing same-sex and opposite-sex twin girls. *Hormones and Behavior*, 47, 230–237.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322–331.
- de Moor, M. H., 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, 642–650.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behaviour. *Psychological Review*, 94, 369–389.
- Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS One*, 7, e29265.
- Digman, J. M. (1990). Personality structure — emergence of the 5-factor model. *Annual Review of Psychology*, 41, 417–440.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum.
- Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic study. *Personality and Social Psychology Bulletin*, 17, 306–315.
- Even, M. D., & vom Saal, F. S. (1992). Seminal vesicle and preputial gland response to steroids in adult male mice is influenced by prior intrauterine position. *Physiology & Behavior*, 51, 11–16.
- Fechner, P. Y. (2003). The biology of puberty: New developments in sex differences. In C. Hayward (Ed.), *Gender differences at puberty* (pp. 17–28). Cambridge, UK: Cambridge University Press.
- Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology*, 59, 981–993.
- Feingold, A. (1991). Sex differences in the effects of similarity and physical attractiveness on opposite-sex attraction. *Basic and Applied Social Psychology*, 12, 357–367.
- Feingold, A. (1992). Gender differences in mate selection preferences: a test of the parental investment model. *Psychological Bulletin*, 112, 125–139.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116, 429–456.
- Geary, D. C. (2010). *Male, female: The evolution of human sex differences* (2nd ed.). Washington, DC: American Psychological Association.
- Gentile, B., Grabe, S., Dolan-Pascoe, B., Twenge, J. M., Wells, B. E., & Maitino, A. (2009). Gender differences in domain-specific self-esteem: A meta-analysis. *Review of General Psychology*, 13, 34–45.
- Henderson, B. A., & Berenbaum, S. A. (1997). Sex-typed play in opposite-sex twins. *Developmental Psychobiology*, 31, 115–123.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory — versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- 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* (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. Boyle, G. Matthews & D. Saklofske (Eds.), *Handbook of personality theory and assessment* (vol. 1, pp. 145–173). London: Sage.
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. Oxford: Oxford University Press.
- 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.
- Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*, 40, 377–393.
- Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, 125, 470–500.
- Koopmans, J. R., Boomsma, D. I., Heath, A. C., & van Doornen, L. J. (1995). A multivariate genetic analysis of sensation seeking. *Behavior Genetics*, 25, 349–356.
- Lang, F. R., John, D., Ludtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the big five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, 43, 548–567.
- Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., & Pedersen, N. L. (2002). The Swedish Twin Registry: A unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine*, 252, 184–205.

- Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlstrom, E., . . . Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: An update. *Twin Research and Human Genetics*, 9, 875–882.
- Lippa, R. A. (2010). Sex differences in personality traits and gender-related occupational preferences across 53 nations: Testing evolutionary and social-environmental theories. *Archives of Sexual Behavior*, 39, 619–636.
- Lippa, R. A., & Hershberger, S. (1999). Genetic and environmental influences on individual differences in masculinity, femininity, and gender diagnosticity: Analyzing data from a classic twin study. *Journal of Personality*, 67, 127–155.
- Loehlin, J. C., Jonsson, E. G., Gustavsson, J. P., Stallings, M. C., Gillespie, N. A., Wright, M. J., & Martin, N. G. (2005). Psychological masculinity-femininity via the gender diagnosticity approach: Heritability and consistency across ages and populations. *Journal of Personality*, 73, 1295–1319.
- Loehlin, J. C., & Martin, N. G. (2000). Dimensions of psychological masculinity-femininity in adult twins from opposite-sex and same-sex pairs. *Behavior Genetics*, 30, 19–28.
- Loehlin, J. C., Spurdle, A., Treloar, S. A., & Martin, N. G. (1999). Number of X-linked androgen receptor gene CAG repeats and femininity in women. *Personality and Individual Differences*, 27, 887–899.
- Miller, E. M. (1994). Prenatal sex hormone transfer: A reason to study opposite-sex twins. *Personality and Individual Differences*, 17, 511–529.
- Miller, E. M., & Martin, N. (1995). Analysis of the effect of hormones on opposite-sex twin attitudes. *Acta Geneticae Medicae et Gemellologiae (Roma)*, 44, 41–52.
- Mitchell, J. E., Baker, L. A., & Jacklin, C. N. (1989). Masculinity and femininity in twin children: Genetic and environmental factors. *Child Development*, 60, 1475–1485.
- Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of musical practice on musical ability. *Psychological Science*, 25, 1795–1803.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.
- Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer.
- Pajer, K., Tabbah, R., Gardner, W., Rubin, R. T., Czambel, R. K., & Wang, Y. (2006). Adrenal androgen and gonadal hormone levels in adolescent girls with conduct disorder. *Psychoneuroendocrinology*, 31, 1245–1256.
- Pervin, L. A., Cervone, D., & John, O. P. (2005). *Personality: Theory and research* (9th ed.). Hoboken, NJ: Wiley.
- Pettit, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers*, 34, 50–54.
- Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Hjelmborg, J. B., Lachine, I., & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research*, 6, 361–376.
- Rammstedt, B., Holzinger, B., & Rammesayer, T. (2004). Comparison of paper-and-pencil vs. computerized administration of the NEO-Five-Factor-Inventory (NEO-FFI). [Zur Äquivalenz der Papier-Bleistift- und einer computergestützten Version des NEO-Fünf-Faktoren-Inventars (NEO-FFI)]. *Diagnostica*, 50, 88–97.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality*, 41, 203–212.
- Resnick, S. M., Gottesman, II, & McGue, M. (1993). Sensation seeking in opposite-sex twins: an effect of prenatal hormones? *Behavior Genetics*, 23, 323–329.
- Rodgers, C. S., Fagot, B. I., & Winebarger, A. (1998). Gender-typed toy play in dizygotic twins: A test of hormone transfer theory. *Sex Roles*, 39, 173–184.
- Rose, R. J., Kaprio, J., Winter, T., Dick, D. M., Viken, R. J., Pulkkinen, L., & Koskenvuo, M. (2002). Femininity and fertility in sisters with twin brothers: Prenatal androgenization? Cross-sex socialization? *Psychological Science*, 13, 263–267.
- Ryan, B. C., & Vandenberg, J. G. (2002). Intrauterine position effects. *Neuroscience & Biobehavioral Reviews*, 26, 665–678.
- Service, S. K., Verweij, K. J., 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, e116.
- Slutske, W. S., Bascom, E. N., Meier, M. H., Medland, S. E., & Martin, N. G. (2011). Sensation seeking in females from opposite- versus same-sex twin pairs: Hormone transfer or sibling imitation? *Behavior Genetics*, 41, 533–542.
- Tapp, A. L., Maybery, M. T., & Whitehouse, A. J. (2011). Evaluating the twin testosterone transfer hypothesis: A review of the empirical evidence. *Hormones and Behavior*, 60, 713–722.
- Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man 1871–1971* (pp. 136–179). Chicago, IL: Aldine.
- van den Berg, S. M., de Moor, M. H., Verweij, K. J., Krueger, R. F., Luciano, M., Arias Vasquez, A., . . . Boomsma, D. I. (2016). Meta-analysis of genome-wide association studies for extraversion: Findings from the genetics of personality consortium. *Behavior Genetics*, 46 (2), 170–182.
- Verweij, K. J. H., Zietsch, B. P., Medland, S. E., Gordon, S. D., Benyamin, B., Nyholt, D. R., . . . Wray, N. R. (2010). A genome-wide association study of Cloninger's temperament scales: Implications for the evolutionary genetics of personality. *Biological Psychology*, 85, 306–317.