Sex Differences in Variability in Personality: A Study in Four Samples

Journal of Personality 81:1, February 2013
© 2012 The Authors
Journal of Personality © 2012,
Wiley Periodicals, Inc.
DOI: 10.1111/j.1467-6494.2012.00784.x

Peter Borkenau,[1] Martina Hřebíčková,[2] Peter Kuppens,[3,4] Anu Realo,[5] and Jüri Allik[5,6]

[1]Martin-Luther-Universität Halle-Wittenberg
[2]Academy of Sciences of the Czech Republic
[3]University of Melbourne
[4]Katholieke Universiteit Leuven
[5]University of Tartu
[6]Estonian Academy of Sciences

Abstract

Objective: Men vary more than women in cognitive abilities and physical attributes, and we expected that men would vary more in personality too. That this has not been found previously may reflect that (a) personality was measured by self-reports that confound target sex with informant sex, and (b) men actually vary more but accentuate personality differences less than women.

Method: We analyzed informant reports and self-reports on the NEO Personality Inventory (NEO PI-R or NEO PI-3) collected for two community and two student samples from four countries: Czech Republic ($N=714$; age $M=36.1$, $SD=14.1$; 58% women), Estonia ($N=1{,}685$; age $M=42.6$, $SD=13.4$; 58% women), Belgium ($N=345$; age $M=18.4$, $SD=3.0$; 78% women), and Germany ($N=302$; age $M=23.4$, $SD=2.7$; 56% women).

Results: Higher male than female variability was found in each sample for informant reports of Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Men but not women were overrepresented in both tails of the distributions of several personality traits.

Conclusions: According to liability-threshold models of mental disorders, this may contribute to men's overrepresentation in some kinds of deviant groups.

Keywords: self-reports, informant reports, perceiver effects, sex differences, variability

Sex differences in cognitive abilities and in personality have been the subject of numerous studies and meta-analyses. For cognitive abilities, it was found that men tend to outperform women on tasks assessing the ability to manipulate visual images in working memory, whereas women tend to outperform men on tasks requiring retrieval from long-term memory and the acquisition and use of verbal information (Halpern, 2000). Altogether, however, sex differences in cognitive abilities seem to be small as far as differences in means are concerned (Hyde, 2005; Hyde, Lindberg, Linn, Ellis, & Williams, 2008).

But research on sex differences in cognitive abilities has not studied central tendencies exclusively. Rather, a recurring theme in this literature has been differences in intrasex variability, and a usual finding has been higher variability between men than between women: Men tend to be overrepresented in highly select samples as well as in some types of mental retardation (Arden & Plomin, 2006; Benbow, 1988; Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Feingold, 1995; Hedges & Friedman, 1993; Hedges & Nowell, 1995; Humphreys, 1988). Moreover, higher variability between men seems not to be limited to cognitive abilities, having also been found for birth weight, adult height, 60 meter dash times, and numerous blood parameters (Lehre, Lehre, Laake, & Danbolt, 2008).

Explanations of higher male variability have been suggested at various levels. Taking an evolutionary perspective, it has been argued that a larger variety of qualities is compatible with reproductive success among men than among women: Females but not males are restricted to high parental investment strategies, giving rise to greater male variability on sexually selected traits (Archer & Mehdikhani, 2003). Concerning

This research was supported by grants from the German Science Foundation to Peter Borkenau; by grant P407/10/2394 from the Czech Science Foundation related to research plan AV0Z70250504 of the Institute of Psychology, Academy of Sciences of the Czech Republic; and by grant SF0180029s08 from the Estonian Ministry of Science and Education to Jüri Allik. We thank the Estonian Genome Center of the University of Tartu and its director, Andres Metspalu, for their help in collecting the Estonian personality data and for their kind permission to use the data in the current study.

Correspondence concerning this article should be addressed to Peter Borkenau, Institut für Psychologie, Universität Halle-Wittenberg, D-06099 Halle, Germany. Email: p.borkenau@psych.uni-halle.de.

genetic explanations, candidate genes to account for larger male variability are located on the X chromosome on which many genes important for the nervous system have been identified (Check, 2005). As females have two X chromosomes whereas males have only one, the effects of two X-linked alleles may be averaged in females, resulting in less extreme phenotypes unless the female is homozygous for that gene. In contrast, effects of an X-linked gene will usually be fully expressed in males, resulting in more extreme phenotypes.

Social explanations are reasonable as well: It may be argued that various social norms are—or at least have been—stricter for women than for men, allowing men to choose between more diverse behavioral options. But as researchers on gender roles have focused mainly on sex differences in mean levels, not on sex differences in diversity of behavioral options, this argument is currently somewhat speculative.

Sex Differences in Variability in Personality

It is therefore reasonable that men vary more than women in personality. This is an important issue as, according to liability-threshold or continuity models of mental disorders, higher variance should result in a higher proportion of a population exceeding a "critical" threshold and being classified as abnormal (Hedges & Friedman, 1993). Thus, if there were no sex differences in mean levels but men varied more than women, the proportion of men exceeding the threshold should be higher than the proportion of women. Furthermore, effects of sex differences in mean levels and in variances on the proportion of individuals who exceed a threshold may add up (Feingold, 1995; Humphreys, 1988). For example, that there are substantially more men than women among prison inmates might reflect men's, as compared to women's, moderately higher mean level (Eagly & Steffen, 1986) as well as a higher variance in aggression. The latter, however, has not yet been established.

There are relatively few analyses on sex differences in variability in personality: In many studies, the intrasex standard deviations are not reported (Schmitt, Realo, Voracek, & Allik, 2008), or they are reported but not analyzed beyond obtaining effect size estimates or providing sex-specific norms for personality measures. For example, Mehl, Vazire, Ramírez-Esparza, Slatcher, and Pennebaker (2007) recorded utterances in six small student samples, using an electronically activated recorder. Women did not talk significantly more than men, but the standard deviation of words spoken per day was higher for men than women in each of the six samples (which is not discussed in that article). And as the measure under study was quite specific, it is not justified to conclude from this finding that men's personality varies more than women's in general.

Earlier meta-analyses on sex differences in personality (e.g., Feingold, 1994) did not analyze differences in variances either. This is changing, however: In a more recent meta-analysis, Else-Quest, Hyde, Goldsmith, and Van Hulle (2006) analyzed sex differences in temperament among children between the ages of 3 months and 13 years, taking means and variances into account. The data consisted mostly of mother and teacher ratings, and no higher variability among boys than among girls was found. But it is not clear how far that generalizes to adults, as "patterns of gender differences and similarities in temperament bear little resemblance to patterns of gender differences and similarities in adult personality" (Else-Quest et al., 2006, p. 63).

A recent meta-analysis by Cross, Copping, and Campbell (2011) studied sex differences in impulsivity in adults, including mean levels as well as variances. These authors expected more variability between men than between women but did not find it, except for the Disinhibition facet of the Sensation Seeking Scale. The personality measures in the studies reviewed were mostly self-reports, but some studies used behavioral measures. Among the latter, a male:female variance ratio of 1.37 was found for the Balloon Analogue Risk Task, but being based on only three studies with modest sample sizes, this variance ratio was not significant. A limitation of this meta-analysis is that the authors excluded clinical and incarcerated samples, and given the overrepresentation of men in pathological and criminal behavior in which risk taking is a factor, this constraint may have reduced the male variance more than the female variance. For that reason, the authors conclude that "our observation of equal variance is therefore inconclusive" (Cross et al., 2011, p. 121).

Thus far, the major analyses of sex differences in variability in adult personality relied on self-reports, the most widespread measurement tool in this field, and they did not yield higher variances between men than between women. In self-reports, however, the sex of the person being described (the target) and of the person providing the description (the perceiver) are entirely confounded. And target and perceiver sex might have opposite effects on intrasex variances, thereby cancelling each other out and resulting in similar self-reported variability among men and among women. Specifically, men might actually vary more, whereas women might accentuate individual differences more than men in descriptions of personality, including their own personality. Currently, that is speculative but it can be tested by analyzing personality descriptions by knowledgeable informants that allow one to separate target sex and perceiver sex, and therefore to study sex differences in variability in personality more thoroughly. That personality descriptions by men and by women differ has been found in studies showing that women tend to be more accurate judges of personality than men (Chan, Rogers, Parisotto, & Biesanz, 2011; Vogt & Colvin, 2003).

As a first check on this hypothesis of opposite effects of target and perceiver sex, we compared intrasex variances in self-reports and informant reports of personality, relying on published data. Given the findings of the meta-analyses by Else-Quest et al. (2006) and Cross et al. (2011), we expected similar variances for women and men in self-reports. If such a finding reflected opposite effects of target sex and perceiver sex on variability in personality descriptions, however, higher

variances for male than for female targets should be observed in informant reports in which a lower variance in the personality of women would not be obscured by their tendency to accentuate individual differences in personality more than men.

In a non-exhaustive search, we identified seven samples described in five publications in which standard deviations for self-reports on the facet scales of Costa and McCrae's (1992) Revised NEO Personality Inventory (NEO PI-R) were reported separately for women and men. Moreover, we identified two samples in which means and standard deviations of descriptions by informants on the same instrument were reported separately for male and female targets. We quantified sex differences in variability by a variance ratio (VR), the intrasex variance for men divided by that for women. Thus, VRs greater than 1.0 indicate higher variability in men. Before averaging the VRs, they were base-10 log transformed, which results in an approximately normal distribution (Hedges & Friedman, 1993). Finally, the antilog of the average VR was computed and will be referred to as the antilog VR here.

Self-Reports

In the manual of the American NEO PI-R (Costa & McCrae, 1992), sex-specific standard deviations for Form S are reported for two samples, one adult (Table B-1) and the other college-aged (Table B-3). The antilog VRs are 0.999 for the adult sample and 1.07 for the college-aged sample. Second, according to the manual of the German NEO PI-R (Ostendorf & Angleitner, 2004) the antilog VR for a nonclinical sample (Table 20) is 1.01. Third, intrasex standard deviations for two Turkish samples, one adult and the other students, are reported by Gülgöz (2002), with antilog VRs of 1.04 for the adult sample and 0.95 for the student sample. Fourth, in a sample from India (Lodhi, Deo, & Belhekar, 2002), the antilog VR is 1.07. Finally, in a sample from Zimbabwe (Piedmont, Bain, McCrae, & Costa, 2002), the antilog VR is 0.89. Thus, four antilog VRs are greater than 1.0 and three are less than 1.0, implying no systematic sex differences in variability in self-reported personality.

Informant Reports

Sex-specific standard deviations for informant reports are available in the manuals of the American (Costa & McCrae, 1992, Table B-2) and the German (Ostendorf & Angleitner, 2004, Table 43) NEO PI-R. The antilog VRs are 1.14 in both samples, higher than any of the seven antilog VRs for self-reports reported above. But it is desirable to replicate this finding in more than just two samples. This is one purpose of the present research.

Second, finding higher VRs in informant reports than in self-reports is a quite indirect test of the hypothesis that higher variability between men is not found in self-reports because descriptions by male perceivers vary less than descriptions by female perceivers. A more direct test would be comparing the variances in descriptions of personality by informants for the four combinations of perceiver sex and target sex: women describing women, women describing men, men describing women, and men describing men. This might also reveal possible Perceiver Sex × Target Sex interactions on perceived variability in personality. But despite these advantages of informant reports, it is useful to analyze self-reports as well: Whereas informant reports allow separating effects of perceiver sex and target sex on descriptions of personality, self-reports are needed to check whether opposite effects of target sex and perceiver sex (if found) are sufficient to explain why no systematic sex differences in variability are found in self-reports of personality.

The Present Study

We analyzed four datasets comprising self-reports and informant reports on the NEO PI-R or NEO PI-3. First, we compared the variance ratios in self-reports and informant reports to replicate the pattern of findings just reported. Second, we compared the variances in informant reports on the NEO facet scales for the four combinations of target sex and perceiver sex to identify main effects of target sex, main effects of perceiver sex, and Target Sex × Perceiver Sex interactions. Third, we studied the effect of sex on the difference between the variance in self-reports and the variance in descriptions by same-sex informants to clarify whether opposite effects of target sex and of perceiver sex are sufficient to explain the lower VRs in self-reports than in informant reports of personality.

Given the lack of systematic previous research on these issues, it was desirable to study more than one sample. Thus, we included datasets from four countries: Czech Republic, Estonia, Belgium (Flanders), and Germany. In each of these countries, self-reports and descriptions by knowledgeable informants on Costa and McCrae's (1992) NEO PI-R or NEO PI-3 (McCrae, Costa, & Martin, 2005) had been collected for mixed-sex samples.

METHOD

Samples

Czech sample. This sample included 714 targets (416 women) recruited in a series of studies (McCrae et al., 2004). Their ages ranged from 15 to 81 years, with a mean of 36.1 (SD = 14.1) years. Dyads of participants described themselves and each other mutually. Thus, age and sex distributions of targets and informants were identical. In the descriptions by informants, 152 women described a woman, 264 women described a man, 264 men described a woman, and 34 men described a man. The Czech self-report and informant report versions of the NEO PI-R (Hřebíčková, 2002) were used.

Estonian sample. Participants of this study (N = 1,685; 969 women) were recruited by the Estonian Genome Center, University of Tartu, and aged 18–89 (M = 42.6, SD = 13.4) years.

The whole project was conducted in accordance with the Estonian Gene Research Act, and all participants had given broad informed consent. Subjects were randomly selected from individuals visiting general practitioners' offices and hospitals and were then recruited by the general practitioners or hospital physicians. In addition to donating blood samples and answering a medical questionnaire, participants were asked to complete the Estonian self-report version of the NEO PI-3 (McCrae et al., 2005) and to ask somebody who knew them well to fill in the informant report version of this instrument. The knowledgeable informants were aged 13 to 89 (M=43.3, SD=16.4) years, and 562 women described a woman, 575 women described a man, 386 men described a woman, and 130 men described a man.

Flemish sample. Flemish data were collected from 345 target participants (270 women) who were psychology students at the Katholieke Universiteit Leuven and who, as a course requirement, rated their own personality with the Dutch version of the NEO PI-R (Hoekstra, Ormel, & DeFruyt, 1996). They also recruited a well-acquainted person (n=345; 190 women, 112 men, 43 not reporting sex), either a relative or a friend, who rated the participant's personality using the informant report form of the same instrument. In the descriptions by informants, 165 women described a woman, 25 women described a man, 79 men described a woman, and 33 men described a man (33 informants did not report gender). The mean age of the targets was 18.4 (SD=3.0) years, and the mean age of the informants was 29.5 (SD=13.7) years.

German sample. The German participants were 302 students (169 women) enrolled at Martin-Luther University in Halle, Germany. Only three of them studied psychology, and they were aged 18 to 35 (M=23.4, SD=2.7) years. They were recruited in 76 groups, each comprising four persons who knew each other well (Borkenau, Zaltauskas, & Leising, 2009). Participants first described the three other group members on 30 adjective scales that are not relevant in the present context. Next, each four-person group was split into two dyads, and all participants described themselves and the other dyad member on several personality inventories, including the NEO PI-R in German (Ostendorf & Angleitner, 2004). As informants, 136 women described a woman, 33 women described a man, 33 men described a woman, and 100 men described a man.

Thus, the four samples were quite different, comprising psychology students (Flemish), non-psychology students (German), and community samples with diverse social and educational backgrounds (Czech and Estonian). Second, the age distributions were quite different, with means ranging from 18.4 years (Flemish) to 42.6 years (Estonian). Third, the four samples represented three different language families: Germanic (German, Flemish), Slavic (Czech), and Finnic (Estonian). Finally, same-sex target-perceiver dyads were overrepresented in the Flemish and German samples, whereas opposite-sex target-perceiver dyads were overrepresented in the Czech and Estonian samples: Finding similar sex differences in variability in all four samples would therefore suggest a robust phenomenon.

Measures

Self-reports were provided on Form S of the NEO PI-R (Costa & McCrae, 1992) in the Czech, Flemish, and German samples, and of the NEO PI-3 (McCrae et al., 2005) in the Estonian sample (Allik et al., 2010). These two inventories are very similar to each other. The NEO Personality Inventories are well-established instruments to assess personality in the normal range, and measure the domains of the Five-Factor Model of personality, distinguishing between six facets within each domain. Responses are given on 5-point scales with the endpoints 0 (*disagree strongly*) and 4 (*agree strongly*). Reports by knowledgeable informants are provided on Form R of these instruments, in which the same items are worded in the third- instead of the first-person singular.

Data Analyses

For descriptive purposes, variance ratios were calculated as described. Variance ratios for single scales were tested for significance using Levene's test for variance homogeneity. To test for differences between intrasex variances in women and men across multiple scales, we first calculated the squared deviation of each participant's score on each NEO scale from the mean of that scale in that particular (sex- and country-specific) sample. These squared deviation scores for individual participants (their mean is the variance) were then analyzed by hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002; Raudenbush, Bryk, & Congdon, 2010), with the multiple scales nested within subjects and sex varying between subjects. As squared deviation scores are not distributed normally, robust standard errors were used for statistical inference. At the within-subject level, we estimated the average of the squared deviation scores for each participant across scales, and at the between-subject level we estimated the effect of the participants' sex on this average. Sex was grand-mean centered. Thus, the intercept for the outcome variable was the mean of the squared deviation scores across all scales and all participants in that particular analysis. And as male and female had been coded 0 and 1, respectively, coefficients for sex indicate the difference of the average variance in women minus that in men.

RESULTS

The variance ratios for self-descriptions and descriptions by informants on the 30 NEO facet scales are reported separately for the four samples in Table 1. Variance ratios for informant reports are variances between male targets divided by variances between female targets independent of the sex of the

Table 1 Male:Female Variance Ratios for Self-Reports and Informant Reports in the Four Samples

	Czech Sample		Estonian Sample		Flemish Sample		German Sample	
	Self-Reports	Informant Reports	Self-Reports	Informant Reports	Self-Reports	Informant Reports	Self-Reports	Informant Reports
N1	.86*	.89	.77*	.88*	.83	1.12	.83	.95
N2	.89	1.24	.93	1.16	1.25	1.27	.96	.95
N3	.67*	.90	.84*	.98	.84	.84	.99	1.13
N4	.94	1.08	.92	1.05	.86	.98	.74	1.06
N5	.86*	.92	.90*	.99	1.48	1.21	.91	1.22
N6	.78*	.93	.85*	.92	.85	1.04	.78	.78
E1	1.06	1.58*	1.15*	1.24*	.94	1.06	.86	1.26
E2	.97	1.33*	.93	1.25*	1.16	1.45	1.16	.97
E3	.84*	1.03	.84*	1.11	.96	1.27	.96	1.13
E4	1.08	1.13	.79*	1.11*	1.18	1.39	1.34	1.49*
E5	.95	1.10	.87*	1.12	.96	1.10	1.15	1.09
E6	.82*	1.34*	.89	1.16	.99	1.13	1.50*	1.56*
O1	.80*	1.27*	.86*	1.08	.84	1.14	1.01	1.20
O2	1.27*	1.33*	.89	.93	1.00	.99	1.49*	1.41*
O3	.97	1.43*	.79*	1.15	1.23	.99	1.44*	1.69*
O4	1.04	1.06	.91*	1.08	1.20	1.51	1.44*	1.48*
O5	.92	1.13	1.08	1.18*	1.16	1.16	1.19	1.51*
O6	.96	.96	.94	1.07	1.38	1.30	1.05	1.37
A1	.98	1.21*	.90	1.18	1.04	1.41	.92	1.32
A2	1.07	1.12	.96	1.09	1.08	1.12	.93	1.21
A3	1.11	1.43*	1.08	1.19*	1.46	1.44*	1.04	.91
A4	.71*	1.27*	.78*	.97	.76*	1.18	.70*	.93
A5	1.00	1.20*	.99	1.25*	1.10	1.11	1.20	1.36
A6	.89	1.11	1.10	.99	1.74*	1.37	1.41*	1.63*
C1	.81*	1.25*	.91	1.17*	.94	1.10	1.05	1.21
C2	.92	1.09	.99	1.13	.68*	.98	.79	1.20
C3	.94	1.24	.98	1.35*	.92	1.33	.91	1.37*
C4	.91	1.19	.92	1.21*	1.05	1.29	1.13	1.34
C5	.85	1.02	.98	1.13*	.93	1.32	1.03	1.11
C6	.80*	1.27*	.88	1.14*	.93	1.29	.89	1.16

Note. Variance ratios greater than 1.0 indicate higher variances for men than for women. N1 = Anxiety; N2 = Angry hostility; N3 = Depression; N4 = Self-consciousness; N5 = Impulsiveness; N6 = Vulnerability; E1 = Warmth; E2 = Gregariousness; E3 = Assertiveness; E4 = Activity; E5 = Excitement seeking; E6 = Positive emotions; O1 = Fantasy; O2 = Aesthetics; O3 = Feelings; O4 = Actions; O5 = Ideas; O6 = Values; A1 = Trust; A2 = Straightforwardness; A3 = Altruism; A4 = Compliance; A5 = Modesty; A6 = Tender-mindedness; C1 = Competence; C2 = Order; C3 = Dutifulness; C4 = Achievement striving; C5 = Self-discipline; C6 = Deliberation.
*$p < .05$.

informant. Asterisks indicate significant ($p < .05$) heterogeneity of the intrasex variances according to Levene's test. Whereas the self-reports did not vary more between men than between women, variances in informant reports were systematically higher for male than for female targets, resulting in numerous VRs larger than 1.0: Of the 33 significant VRs for informant reports in Table 1, only one was less than 1.0 (N1 in the Estonian sample), whereas 32 indicated more variance in descriptions of men.

Separate Analyses for Personality Domains

To check whether the VRs differed between personality domains, we averaged the VRs in Table 1 across the four samples and the six facets in each domain (that is, across 24 VRs). For self-reports, this resulted in antilog VRs of 0.89 for Neuroticism, 1.00 for Extraversion, 1.06 for Openness to Experience, 1.02 for Agreeableness, and 0.92 for Conscientiousness. For statistical inference, the squared deviation scores were submitted to separate HLM analyses for each domain, run across the combined sample of 3,046 participants, with the six relevant facets nested within participants. The results are reported in the upper part of Table 2. Self-reports of Neuroticism varied significantly more between women, whereas sex differences were not significant for the four other personality domains.

These analyses were repeated for reports by informants. Here, the antilog VRs were 1.01 for Neuroticism, 1.21 for Extraversion, 1.21 for Openness to Experience, 1.20 for Agreeableness, and 1.21 for Conscientiousness. Figure 1 illustrates the antilog VRs separately for self- and informant reports and the five personality domains.

HLM analyses of informant reports were run as for self-reports, except that three between-subjects predictors were included: (a) sex of target, (b) sex of perceiver, and (c) the Target Sex × Perceiver Sex interaction. The latter was modeled

Table 2 Hierarchical Linear Modeling of the Variances Separately for Personality Domains

Personality Domain		Coefficient[a]	Standard Error[b]	*t*-ratio	*p*-value
Self-Reports					
Neuroticism	Intercept	26.89			
	Sex	**3.47**	0.87	4.00	<.001
Extraversion	Intercept	29.85			
	Sex	1.26	0.92	1.38	.17
Openness to Experience	Intercept	26.63			
	Sex	–.07	0.78	−0.09	.93
Agreeableness	Intercept	21.86			
	Sex	0.21	0.70	0.30	.76
Conscientiousness	Intercept	23.72			
	Sex	1.26	0.78	1.61	.11
Informant Reports					
Neuroticism	Intercept	28.40			
	Target sex	−0.67	0.94	−0.72	.46
	Perceiver sex	1.65	0.96	1.72	.07
	T×I	**−5.63**	2.01	−2.79	.003
Extraversion	Intercept	31.52			
	Target sex	**−5.81**	1.06	−5.50	<.001
	Perceiver sex	**4.27**	1.01	4.22	<.001
	T×I	−1.21	2.22	−0.55	.59
Openness to Experience	Intercept	26.17			
	Target sex	**−4.50**	0.85	−5.29	<.001
	Perceiver sex	1.27	0.84	1.51	.13
	T×I	−1.69	1.81	−0.93	.35
Agreeableness	Intercept	25.77			
	Target sex	**−4.17**	0.89	−4.70	<.001
	Perceiver sex	**2.19**	0.87	2.51	.01
	T×I	−1.17	1.88	−0.62	.53
Conscientiousness	Intercept	28.31			
	Target sex	**−5.91**	1.11	−5.32	<.001
	Perceiver sex	0.97	1.06	0.91	.36
	T×I	0.60	2.33	0.26	.80

Note. Statistically significant coefficients are in boldface. T×I=Target Sex×Perceiver Sex interaction.
[a]Coefficients for target sex or perceiver sex with a positive sign indicate higher variances for female targets or informants. [b]Robust standard errors.

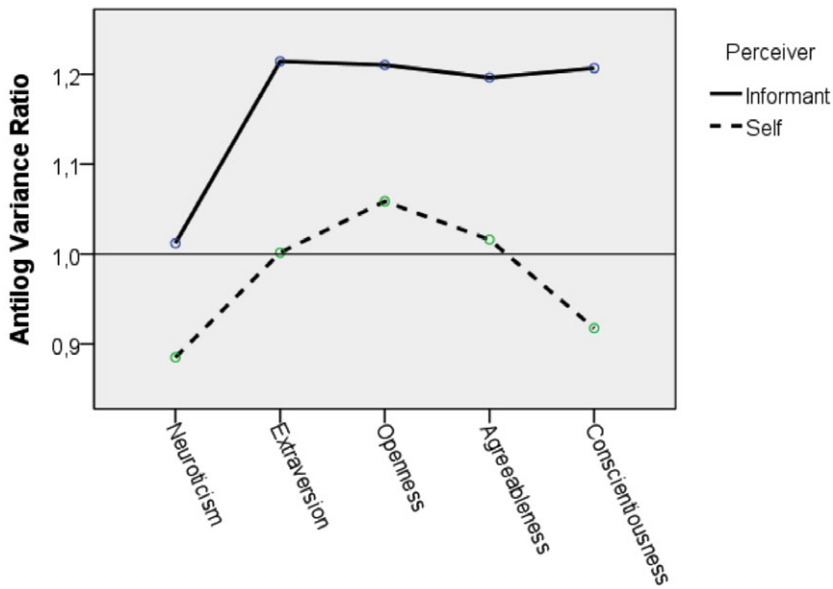

Figure 1 Male:female antilog variance ratios separately for the domains of the Five-Factor Model and self- and informant reports. Variance ratios are averaged across the six facets within each personality domain and across the four samples.

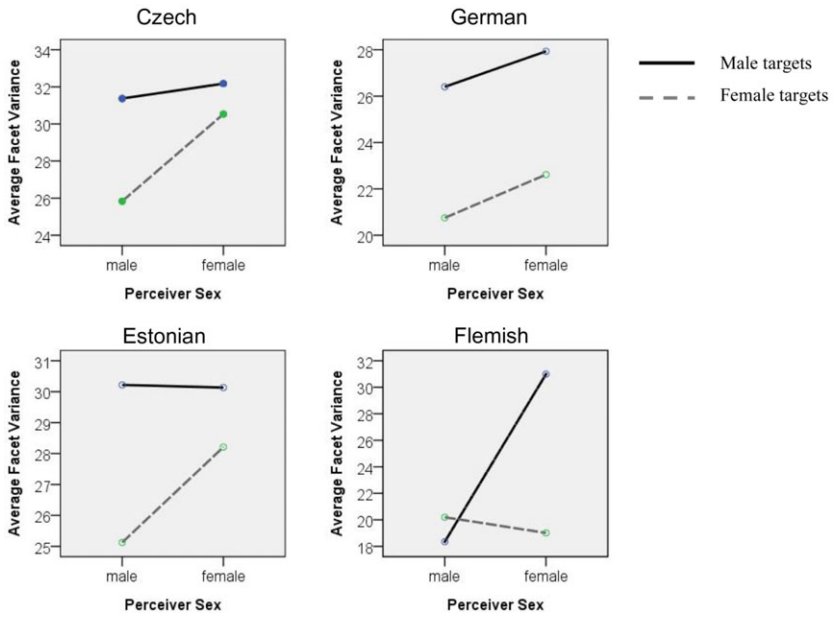

Figure 2 Average facet variances in the four samples separately for male and female targets and perceivers.

by first centering and then multiplying the codes for target sex and perceiver sex, following suggestions by West, Aiken, and Krull (1996). Results are reported in the lower part of Table 2. Male targets varied significantly more than female targets on Extraversion, Openness to Experience, Agreeableness, and Conscientiousness (all $ps < .001$), but not on Neuroticism. Moreover, reports by female informants varied more than reports by male informants for all five personality domains, but that difference was significant for Extraversion and Agreeableness (and marginally significant for Neuroticism) only. Thus, there was strong evidence of more variability between male than female targets on all traits except Neuroticism, and there was weaker evidence of more variability in descriptions by female than by male perceivers.

Separate Analyses for Samples

In an additional set of analyses, we analyzed the similarities and differences between the four datasets to check how well our findings replicated across the four quite diverse samples. For descriptive purposes, the variances in the descriptions of female targets by female perceivers, female targets by male perceivers, male targets by female perceivers, and male targets by male perceivers were separately averaged within samples across all 30 NEO scales. These averages are reported in Figure 2 and illustrate the opposite effects of target sex and perceiver sex on the variances in informant reports of personality.

For purposes of statistical inference, we used HLM analyses again, but this time with the squared deviation scores for all 30 NEO facets nested within participants (implying that we aggregated across personality domains), and separate analyses for the four samples. Otherwise, the analytic procedure was the same as in the trait-specific analyses. The results are reported in Table 3. Again, the upper part refers to self-reports and the lower part refers to informant reports. There was significantly *more* variance between women in the self-reports by the Czech sample, whereas there was insignificantly *less* variance between women in the three other samples. Thus, the sex differences in self-reported variability were inconsistent.

By contrast, significantly more variability for male than for female targets was reported by the informants in each sample, showing that more variability between men than between women is a recurring finding in informant reports. Moreover, more variability in descriptions by female than by male perceivers was found in each of the four samples, with this difference being significant in the Czech data and marginally significant ($p = .07$) in the Estonian data. Finally, a significant Perceiver Sex × Target Sex interaction was found in the Flemish data, but it did not replicate across samples.

Comparison of Self-Reports With Descriptions by Same-Sex Informants

The finding of opposite effects of target sex and perceiver sex on the variances in informant reports, in combination with the confounding of target sex and perceiver sex in self-reports, might account for the lack of systematic sex differences in the variability of self-descriptions. If that were the case, self-reports by women should vary as much as descriptions of women by other women, and self-reports by men should vary as much as descriptions of men by other men. Therefore, we ran additional analyses for those participants ($n = 1{,}312$) who had been described by a same-sex informant. To compare the variances in their self-reports to those in their informant

Table 3 Hierarchical Linear Modeling of Variances Separately for Samples Combined for Personality Domains

Sample		Coefficient[a]	Standard Error[b]	*t*-ratio	*p*-value
Self-Reports					
Czech	Intercept	27.39			
	Sex	**2.50**	1.01	2.48	.01
Estonian	Intercept	27.10			
	Sex	−0.84	0.78	−1.07	.28
Flemish	Intercept	18.86			
	Sex	−1.65	1.71	−0.96	.34
German	Intercept	21.71			
	Sex	−0.94	1.22	−0.77	.44
Informant Reports					
Czech	Intercept	30.51	0.59		
	Target sex	**−3.80**	1.40	−2.72	.007
	Perceiver sex	**3.14**	1.36	2.31	.02
	T×I	3.48	2.88	1.21	.23
Estonian	Intercept	29.01			
	Target sex	**−3.15**	0.87	−3.64	<.001
	Perceiver sex	1.60	0.90	1.79	.07
	T×I	3.08	1.91	1.61	.11
Flemish	Intercept	20.55			
	Target sex	**−7.07**	2.82	−2.50	.01
	Perceiver sex	1.22	1.76	0.69	.49
	T×I	**−12.13**	4.93	−2.46	.02
German	Intercept	24.14			
	Target sex	**−5.40**	2.29	−2.35	.02
	Perceiver sex	1.65	2.04	0.81	.42
	T×I	−0.04	4.35	−0.01	.99

Note. Significant coefficients are in boldface. T×I=Target Sex×Perceiver Sex interaction.
[a]Coefficients for sex with a positive sign indicate higher variances between female targets or informants. [b]Robust standard errors.

reports, we subtracted—separately for each participant and each NEO scale—the squared deviation score for their self-report from the squared deviation score for their description by an informant. These differences were then submitted to HLM analyses, with the difference scores for the 30 NEO scales nested within participants and sex as a between-subjects factor. Separate analyses were run for the Czech, Estonian, Flemish, and German samples. The findings are reported in Table 4.

The intercept coefficients suggest a tendency in each sample for somewhat more variance in informant reports than in self-reports that was significant in the German sample but not in any other sample. And the more interesting coefficients for sex do not suggest a clear trend either: Whereas the differences of the variances in informant reports minus those in self-reports were significantly larger for men than for women in the Estonian sample, an insignificant trend in the opposite direction was found in the Czech and Flemish samples. Thus, if targets and informants were of the same sex, the VRs in informant reports did not systematically exceed those in self-reports. This finding is consistent with the hypothesis that opposite effects of target sex and perceiver sex on the intrasex variances cancel each other out in self-descriptions.

Table 4 Hierarchical Linear Modeling of Differences Between Variability in Informant Reports Minus Variability in Self-Reports

Sample		Coefficient[a]	Standard Error[b]	*t*-ratio	*p*-value
Self-Reports					
Czech	Intercept	1.10	1.21	0.91	.36
	Sex	1.93	3.06	0.63	.53
Estonian	Intercept	0.14	0.71	0.19	.85
	Sex	**−3.95**	1.78	−2.23	.03
Flemish	Intercept	0.72	0.88	0.82	.41
	Sex	2.44	2.57	0.95	.34
German	Intercept	**2.37**	1.05	2.26	.03
	Sex	−2.42	2.20	−1.10	.27

Note. Only participants described by same-sex informants were included. Significant coefficients are in boldface.
[a]Coefficients for sex with a positive sign indicate that the difference of variability in informant reports minus that in self-reports is higher for females. [b]Robust standard errors.

Sex Ratios in the Tails of the Distributions

Given the higher intrasex variances between male than between female targets in informant reports of personality (unless targets and informants were of the same sex), we expected that men but not women should be overrepresented in both tails of the distributions of informant-reported trait levels. As women outnumbered men in all samples, comparing the absolute number of men to the absolute number of women in the tails of each distribution would have been misleading. Therefore, we chose a common cut-off score and then counted which percentage of all women and which percentage of all men exceeded that threshold. Specifically, we first identified the 5th and 95th percentiles in the mixed-sex sample, and then the percentage of men among all men, and the percentage of women among all women, with scores more extreme than these percentiles. For example, if 4% of the total sample had a score exceeding 28, and 2% had a score of exactly 28, we counted which percentages of all men and all women had scores higher than 28. These analyses were run for the Czech and Estonian but not for the Flemish and German data because, for the latter samples, too few men (fewer than seven) were expected beyond the 5th and 95th percentiles.

The percentages of Czech and Estonian men and women in the tails of each distribution are reported in Table 5. For six traits (Czech sample) or seven traits (Estonian sample), a higher proportion of men than women was found in both tails. In the Czech sample, this occurred for E2 (Gregariousness), O1 (Openness to Fantasy), O5 (Openness to Ideas), A2 (Straightforwardness), C1 (Competence), and C6 (Deliberation). In the Estonian sample, that pattern was found for N2 (Angry Hostility), E1 (Warmth), E2 (Gregariousness), O6 (Openness to Values), A1 (Trust), C1 (Competence), and C4 (Achievement Striving). It is remarkable that for none of the scales in none of the samples a higher proportion of women was observed in both tails of the same distribution. This illustrates nicely that, according to informant reports, extreme

Table 5 Percentages of Women and Men in the Czech and Estonian Samples Scoring Below the 5th or Above the 95th Percentile in the Mixed-Sex Distribution of Informant Reports

	Czech Sample				Estonian Sample			
	<5th Percentile		>95th Percentile		<5th Percentile		>95th Percentile	
	Men	Women	Men	Women	Men	Women	Men	Women
N1	8.1	1.0	0.3	6.0	6.3	1.4	1.8	5.9
N2	6.4	1.7	3.7	4.1	**5.4**	**2.3**	**3.5**	**3.2**
N3	7.4	2.4	2.3	5.8	4.7	2.5	2.8	4.8
N4	6.0	1.4	2.7	6.2	7.8	2.9	3.2	5.8
N5	3.7	4.8	4.0	4.3	3.4	3.7	4.6	4.2
N6	8.1	2.6	2.3	5.8	8.5	2.5	2.8	4.6
E1	8.1	2.4	1.3	2.2	**4.9**	**3.1**	**3.9**	**3.4**
E2	**5.4**	**1.4**	**4.0**	**3.8**	**5.4**	**2.8**	**5.6**	**3.5**
E3	3.0	4.6	4.4	2.6	5.0	5.0	5.3	3.0
E4	7.4	3.1	2.7	3.8	4.9	2.8	3.8	5.0
E5	2.7	2.9	6.0	3.1	4.3	4.5	6.6	2.7
E6	5.4	2.9	3.7	5.3	6.7	2.2	2.7	5.0
O1	**5.4**	**2.6**	**4.0**	**3.6**	4.2	2.4	4.2	5.2
O2	6.7	1.9	2.0	4.6	7.7	2.8	2.1	5.3
O3	8.1	0.7	1.3	2.2	5.9	1.2	1.7	5.3
O4	5.4	3.1	3.0	5.0	4.9	4.6	2.9	3.1
O5	**4.0**	**3.8**	**6.7**	**2.6**	3.6	3.6	7.1	3.3
O6	3.0	3.4	3.4	4.1	**4.9**	**4.0**	**5.3**	**3.3**
A1	7.4	2.9	3.4	4.3	**5.2**	**3.4**	**4.2**	**2.9**
A2	**6.0**	**2.9**	**2.3**	**2.2**	5.6	3.4	3.9	4.2
A3	7.0	2.2	0.7	4.1	6.3	2.7	3.8	4.9
A4	5.7	2.9	4.0	4.3	4.7	3.4	3.4	3.4
A5	6.0	2.2	1.7	4.8	8.5	3.4	3.2	3.6
A6	7.7	1.0	0.7	5.0	5.0	2.7	2.9	6.0
C1	**6.4**	**3.4**	**4.0**	**2.4**	**3.9**	**3.7**	**5.9**	**4.2**
C2	6.4	0.7	2.0	5.0	6.4	2.8	2.7	4.5
C3	6.7	2.4	2.0	6.5	5.7	2.2	1.5	2.6
C4	3.7	1.9	2.7	3.4	**5.3**	**3.0**	**3.4**	**2.8**
C5	6.4	3.1	2.3	6.7	5.9	3.1	4.1	4.2
C6	**6.7**	**3.4**	**6.0**	**3.4**	6.1	3.3	3.6	4.6

Note. Boldface indicates that men are overrepresented in both tails of a distribution.

levels of personality traits occur more frequently among men than among women.

Sex Differences in Means and Variances

Usually, there are curvilinear relations between means and variances of personality variables: the more extreme their mean, the lower their variance (Wood & Wortman, 2012). Thus, sex differences in variances might be by-products of sex differences in means. To check whether this might explain our findings, we calculated the absolute value of the deviation of the sample mean from the midpoint of the scale (i.e., 16, the sum of eight items, each coded from 0 to 4) separately for (a) the four samples, (b) self-reports and informant reports, (c) male and female targets, and (d) the 30 NEO facet scales and averaged them across samples and facets. For self-reports, these averages were 3.25 for female participants and 2.82 for male participants, whereas for informant reports they were 3.27 for female targets and 2.99 for male targets. Thus, the sample means were more extreme for women than for men. But that cannot explain our findings because this higher extremity in women was more pronounced for self-reports (an average difference of 0.43) than for informant reports (an average difference of 0.28), whereas a higher variability for men was found in descriptions by informants only.

Contributions of Item Variances and Item Covariances to Variability

There may be two distinguishable reasons why reports by women yielded more variability than reports by men, and reports on male targets yielded more variability than reports on female targets. First, the item responses may be more extreme, resulting in higher item variances. Second, the item variances may be the same, but responses may covary more strongly within scales, that is, the scales may be more consistent. Effects of perceiver sex on consistency could reflect sex differences in conscientiousness of reporting, whereas effects of target sex on consistency could reflect sex differences in the traitedness (Church, 2009) of behavior.

Table 6 Averaged Alpha Coefficients Separately for Samples and Sex of Targets and Informants

	Sample			
Type of Judgments	Estonian	Czech	Flemish	German
Informant reports by female perceivers	.79	.77	.75	.79
Informant reports by male perceivers	.76	.74	.73	.77
Informant reports on female targets	.77	.74	.74	.77
Informant reports on male targets	.78	.77	.75	.79

A useful measure of consistency in this context is Cronbach's coefficient alpha, the ratio of all item covariances divided by the sum of all item variances plus all item covariances (counting the covariances twice). We calculated the alphas for the 30 NEO facet scales separately in the four samples for (a) informant reports by female perceivers, (b) informant reports by male perceivers, (c) informant reports on female targets, and (d) informant reports on male targets. These alpha coefficients were then averaged across the 30 NEO scales, using Fisher's *z*-transformation for correlations.

The back-transformed average alphas are reported in Table 6, and they are quite similar. Nevertheless, there are systematic differences in that in each of the four samples they are higher for female perceivers than for male perceivers and for male targets than for female targets. This is consistent with the hypotheses that (a) men are less careful than women in providing descriptions of personality, and (b) the behavior of men is more traited than that of women. But these differences are tiny and therefore not particularly conclusive. And they do not imply that men do not also vary more in their specific behaviors. Rather, given the very small differences in the alpha coefficients, and the quite large VRs of about 1.20 for all traits except Neuroticism (Figure 1), men seem to vary more on the single items as well.

DISCUSSION

The present study shows that although higher variability between men than between women is not found in self-reports of personality, informant reports vary significantly more for male than for female targets. That was found for all traits except Neuroticism (Table 2), and it was found in each of the four samples (Table 3).

Furthermore, there seems to be an opposite effect of the sex of the perceiver: Personality descriptions by female informants vary more than those by male informants. This effect is weaker, however, and does not always reach conventional levels of significance. But it is unlikely to reflect sampling error only because the sign of the coefficient for perceiver sex was positive for each trait (Table 2) and in all four samples (Table 3). The latter finding is particularly telling, as observations in different samples are independent. Thus, in cases of no effects of perceiver sex, the probability of four coefficients having a positive sign is $p = .06$.

There were also significant Target Sex × Perceiver Sex interactions, but they seem to not be reliable. We checked whether the interactive effect on Neuroticism in the combined sample (Table 2) replicated across the four individual samples, but it did not. And the sample-specific interactions suggested by Figure 2 are insignificant for three samples, whereby the sign of the significant interaction in the Flemish sample is opposite to that of the nonsignificant interactions in the Czech and Estonian samples (Table 3).

The finding of opposite effects of target sex and perceiver sex on the variability in informant reports of personality suggests that there is actually more variability in the observable behavior of men, but it is obscured in self-reports by men's tendency to accentuate individual differences less than women. Consequently, if informants and targets are of the same sex, the VRs in self-reports and informant reports do not differ systematically (Table 4).

Neuroticism, however, is an exception: Informant reports of Neuroticism vary equally, whereas self-reports of Neuroticism vary more between women than men. That a lower male:female VR is found for Neuroticism than for the other personality domains independent of the source of the descriptions suggests that this lower VR does not reflect reporting bias only. Rather, it seems that men actually do not vary more in Neuroticism than women. In combination with lower Neuroticism means for men than for women (Costa & McCrae, 1992), this might reflect that emotional instability is not particularly well accepted in men, resulting in a truncation of the upper tail of the distribution of men's Neuroticism levels. Indeed, the highest informant-reported Neuroticism scores in our four samples were 7 to 20 points lower for men than for women.

Accuracy of Self-Reports and Informant Reports

That only informant reports but not self-reports indicate more variability between men than women raises the question of whether self-reports or informant reports are more trustworthy. If self-reports are more dependable, that discrepancy would reflect that informants describe men as *more* extreme than they actually are. Conversely, if the informant reports are more dependable, that discrepancy would reflect that men perceive themselves as *less* extreme than they actually are. So which source should we trust more?

There are several reasons to trust the informant reports more: Some of these reasons are general, whereas others are specific to research on sex differences. General reasons are that in current personality research, informant reports are regarded as a very useful complement to self-reports (Connelly & Ones, 2010; Kolar, Funder, & Colvin, 1996; McCrae et al., 2005; Vazire, 2006; Vazire & Mehl, 2008), having the

advantage that descriptions by more than one outside observer are possible (Hofstee, 1994). Moreover, several studies show that informant reports are frequently more valid (John & Robins, 1993; Kolar et al., 1996; Vazire, 2010; Vazire & Mehl, 2008; Connelly & Ones, 2010). Reasons specific to research on sex differences are that informant reports allow for separating the effects of target sex and perceiver sex that are necessarily confounded in self-reports. Obviously, that advantage of informant reports is also relevant for research on sex differences in means.

Limitations and Future Directions

As with every study, this one has limitations. One limitation is that each participant was described by one informant only. Aggregated descriptions by multiple judges are more reliable and less subjective than descriptions by a single judge (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Funder, 1995; Hofstee, 1994; Kenny, 1994, 2004; Kolar et al., 1996). We expect, however, that similar sex differences in variability as in the present study would be obtained for descriptions aggregated across judges, as the variance of an aggregate is the sum of the variances and covariances of the single measures. Therefore, our finding of more variability in men would not replicate in aggregated descriptions of personality only if higher variances in descriptions of men were compensated by higher covariances (between informants) in descriptions of women. We view such a pattern as not particularly likely (Allik et al., 2010).

A second limitation is that only one measurement instrument, the NEO PI-R (or NEO PI-3), was included. However, this is also a strength, as different findings for different instruments would be difficult to interpret. It has yet to be clarified, though, whether similar findings would be obtained for traits not represented in the Five-Factor Model of personality, or for the same traits measured with different instruments. Given the widespread use of the Five-Factor Model in descriptions of personality, as well as the widespread use of the NEO PI-R to measure the dimensions of this model, we believe that the NEO PI-R is a good starting point.

A third limitation is the range of cultures represented in our collaborative effort. Although the four samples differ in numerous respects, all data were collected in Europe. Thus, there are cultural similarities. It would therefore be interesting to see whether our main findings replicate in non-European cultures. Such studies could not only provide replications, but also help clarify whether the sources of higher male variability are mainly biological or cultural. If the findings were systematically different in samples stemming from other continents, cultural explanations would be supported, whereas similar findings as in the present study would be more consistent with biological explanations. We trust that several datasets exist in non-European cultures that include self-reports and informant reports of personality and allow for such analyses without gathering new data.

This research was motivated by findings that men vary more than women in cognitive abilities. We identified similar phenomena in informant reports of Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Future research should investigate how these findings generalize across personality traits, across cultures, and across methods for the assessment of personality.

References

Allik, J., Realo, A., Mõttus, R., Esko, T., Pullat, J., & Metspalu, A. (2010). Variance determines self-observer agreement on the Big Five personality traits. *Journal of Research in Personality*, **44**, 421–426.

Archer, J., & Mehdikhani, M. (2003). Variability among males in sexually selected attributes. Review of General *Psychology*, **7**, 219–236.

Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, **41**, 39–48.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, **11**, 169–183.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, **86**, 599–614.

Borkenau, P., Zaltauskas, K., & Leising, D. (2009). More may be better but there may be too much: Optimal trait level and self-enhancement bias. *Journal of Personality*, **77**, 825–858.

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. S. (2011). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, **45**, 117–120.

Check, E. (2005). Genetics: The X factor. *Nature*, **434**, 266–267.

Church, A. T. (2009). Prospects for an integrated trait and cultural psychology. *European Journal of Personality*, **23**, 153–182.

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

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Orlando, FL: Psychological Assessment Resources.

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, **137**, 97–130.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey 1932. *Intelligence*, **31**, 533–542.

Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, **100**, 309–330.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, **132**, 33–72.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, **116**, 429–456.

Feingold, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist*, **50**, 5–13.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, **102**, 652–670.

Gülgöz, S. (2002). Five-factor model and NEO-PI-R in Turkey. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 175–196). New York: Kluwer.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Erlbaum.

Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research*, **63**, 94–105.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, **269**, 41–45.

Hoekstra, H. A., Ormel, J., & DeFruyt, F. (1996). *NEO Persoonlijkheidsvragenlisten, NEO-PI-R en NEO-FF-I*. Handleigin. Lisse, the Netherlands: Swets & Zeitlinger.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality*, **8**, 149–162.

Hřebíčková, M. (2002). Internal consistency of the Czech version of the NEO Personality Inventory (NEO-PI-R). *Ceskoslovenska Psychologie*, **46**, 521–535.

Humphreys, L. G. (1988). Sex differences in variability may be more important than sex differences in means. *Behavioral and Brain Sciences*, **11**, 195–196.

Hyde, J. S. (2005). The gender similarity hypothesis. *American Psychologist*, **60**, 581–592.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, **321**, 494–495.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, **61**, 523–551.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, **8**, 265–280.

Kolar, D. W., Funder, D. C., & Colvin, R. C. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, **64**, 311–337.

Lehre, A. C., Lehre, K. P., Laake, P., & Danbolt, N. C. (2008). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology*, **51**, 198–206.

Lodhi, P. H., Deo, S., & Belhekar, V. M. (2002). The five-factor model of personality: Measurement and correlates in the Indian context. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 227–248). New York: Kluwer.

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

McCrae, R. R., Costa, P. T., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., et al. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality*, **38**, 179–201.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, **317**, 82.

Ostendorf, F., & Angleitner, A. (2004). *NEO-PI-R: NEO-Persönlichkeitsinventar nach Costa und McCrae* [NEO Personality Inventory according to Costa and McCrae]. Göttingen, Germany: Hogrefe.

Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T. (2002). The applicability of the five-factor model in a sub-Saharan culture: The NEO-PI-R in Shona. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 155–173). New York: Kluwer.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2010). *HLM7: Hierarchical linear and nonlinear modelling* [Computer software]. Lincolnwood, IL: Scientific Software International.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, **94**, 168–182.

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, **40**, 472–481.

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.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, **95**, 1202–1216.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, **71**, 267–295.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, **64**, 1–48.

Wood, D., & Wortman, J. (2012). Trait means and desirabilities as artifactual and real sources of differential stability of personality traits. *Journal of Personality*, **80**, 665–701.