2014

Gender Stereotypes of Personality: Universal and Accurate?

C. E. Löckenhoff

W. Chan

R. R. McCrae

F. De Fruyt

L. Jussim

See next page for additional authors

Journal of Cross-Cultural Psychology
2014, Vol. 45(5) 675–694

DOI: 10.1177/0022022113520075
jccp.sagepub.com
SAGE

Gender Stereotypes of Personality: Universal and Accurate?

Corinna E. Löckenhoff[1], Wayne Chan[2], Robert R. McCrae[3], Filip De Fruyt[4], Lee Jussim[5], Marleen De Bolle[4], Paul T. Costa Jr.[6], Angelina R. Sutin[7], Anu Realo[8], Jüri Allik[8,9], Katsuharu Nakazato[10], Yoshiko Shimonaka[11], Martina Hřebíčková[12], Sylvie Graf[12], Michelle Yik[13], Emília Ficková[14], Marina Brunner-Sciarra[15], Nora Leibovich de Figueora[16], Vanina Schmidt[16], Chang-kyu Ahn[17], Hyun-nie Ahn[18], Maria E. Aguilar-Vafaie[19], Jerzy Siuta[20], Barbara Szmigielska[21], Thomas R. Cain[21], Jarret T. Crawford[22], Khairul Anwar Mastor[23], Jean-Pierre Rolland[24], Florence Nansubuga[25], Daniel R. Miramontez[26], Veronica Benet-Martínez[27], Jérôme Rossier[28], Denis Bratko[29], Iris Marušić[30], Jamin Halberstadt[31], Mami Yamaguchi[31], Goran Knežević[32], Thomas A. Martin[33], Mirona Gheorghiu[34], Peter B. Smith[35], Claudio Barbaranelli[36], Lei Wang[37], Jane Shakespeare-Finch[38], Margarida P. Lima[39], Waldemar Klinkosz[40], Andrzej Sekowski[40], Lidia Alcalay[41], Franco Simonetti[41], Tatyana V. Avdeyeva[42], V. S. Pramila[43], and Antonio Terracciano[7]

Abstract

Numerous studies have documented subtle but consistent sex differences in self-reports and observer-ratings of five-factor personality traits, and such effects were found to show well-defined developmental trajectories and remarkable similarity across nations. In contrast, very little is known about *perceived* gender differences in five-factor traits in spite of their potential implications for gender biases at the interpersonal and societal level. In particular, it is not clear how perceived gender differences in five-factor personality vary across age groups and national contexts and to what extent they accurately reflect assessed sex differences in personality. To address these questions, we analyzed responses from 3,323 individuals across 26 nations (mean age = 22.3 years, 31% male) who were asked to rate the five-factor personality traits of typical men or women in three age groups (adolescent, adult, and older adult) in their respective nations. Raters perceived women as slightly higher in openness, agreeableness, and conscientiousness as well as some aspects of extraversion and neuroticism. Perceived gender differences were fairly consistent across nations and target age groups and mapped closely onto assessed sex differences in self- and observer-rated personality. Associations between the average size of perceived gender differences and national variations in sociodemographic characteristics, value systems, or gender equality did not reach statistical significance. Findings contribute to our understanding of the underlying mechanisms of gender stereotypes of personality and suggest that perceptions of actual sex differences may play a more important role than culturally based gender roles and socialization processes.

Keywords
personality, gender/sex roles, developmental: child/adolescent, developmental: elderly

Sex differences[1] in personality traits have captured scientific and popular interest for decades (Maccoby & Jacklin, 1974). While initial inquiries used a wide variety of conceptual models, which hampered integration across studies (Feingold, 1994), the emergence of the five-factor model (FFM) of personality in the 1990s (Goldberg, 1990; McCrae & John, 1992) has provided an overarching framework to systematically aggregate such findings. The resulting body of research suggests that sex differences in personality are fairly small in size. Also, although

[1]Cornell University, Ithaca, NY, USA
[2]Rutgers University, Camden, NJ, USA
[3]Baltimore, MD, USA
[4]Ghent University, Belgium
[5]Rutgers University, Piscataway, NJ, USA
[6]Duke University, Durham, NC, USA
[7]Florida State University, Tallahassee, FL, USA
[8]University of Tartu, Estonia
[9]Estonian Academy of Sciences, Estonia
[10]Tokyo University of Social Welfare, Isesaki City, Japan
[11]Bunkyo Gakuin University, Japan
[12]Academy of Sciences of the Czech Republic, Brno
[13]Hong Kong University of Science and Technology, China
[14]Slovak Academy of Sciences, Bratislava
[15]Universidad Peruana Cayetano Heredia, Lima, Peru
[16]University of Buenos Aires, Argentina
[17]Pusan National University, Busan, Korea
[18]Ewha Womans University, Seoul, Korea
[19]Tarbiat Modares University, Tehran, Islamic Republic of Iran
[20]Jagiellonian University, Krakow, Poland
[21]Hampshire College, Amherst, NH, USA
[22]The College of New Jersey, Ewing, NJ, USA
[23]Universiti Kebangsaan Malaysia, Bangi
[24]Université Paris Ouest Nanterre La Défense, France
[25]Makerere University, Kampala, Uganda
[26]San Diego Community College District, Riverside, CA, USA
[27]ICREA and Universitat Pompeu Fabra, Barcelona, Spain
[28]University of Lausanne, Switzerland
[29]University of Zagreb, Croatia
[30]Institute for Social Research in Zagreb, Croatia
[31]University of Otago, Dunedin, New Zealand
[32]Belgrade University, Serbia
[33]Susquehanna University, Selinsgrove, PA, USA
[34]The Queen's University Belfast, Northern Ireland
[35]University of Sussex, Brighton, UK
[36]La Sapienza University, Rome, Italy
[37]Peking University, Beijing, China
[38]Queensland University of Technology, Brisbane, Australia
[39]University of Coimbra, Portugal
[40]The John Paul II Catholic University of Lublin, Poland
[41]Pontificia Universidad Católica de Chile, Santiago
[42]University of St. Thomas, St Paul, MN, USA
[43]Visakhapatnam, AP, India

Corresponding Author:
Corinna E. Löckenhoff, Cornell University, G60B MVR Hall, Ithaca, NY 14853, USA.
Email: CEL72@cornell.edu

systematic developmental trends and cross-cultural differences have been found, there is remarkable consistency across age groups, assessment methods, and cultural contexts (Costa, Terracciano, & McCrae, 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt, Realo, Voracek, & Allik, 2008).

In spite of their small size, sex differences in personality have important practical implications because five-factor traits are linked to consequential outcomes in the interpersonal, work, and health domains (Ozer & Benet-Martinez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). When considering practical consequences, perceived gender differences may be just as relevant because—even in the absence of actual sex differences—they may influence the range of experiences that men and women have access to (e.g., whether they are offered leadership positions) and the way they are treated in their social environment (e.g., whether emotional expressiveness is rewarded or sanctioned).

Although there is a rich literature on gender stereotypes (e.g., Diekman & Eagly, 2000), they are rarely examined from a five-factor perspective (Williams, Satterwhite, & Best, 1999). Thus, it is not clear to what extent gender stereotypes of five-factor personality map onto assessed sex differences in terms of direction, size, developmental trends, and cross-national patterns. A better understanding of such associations would provide insights into the underlying mechanisms and practical implications of sex differences, and the present study represents a step in this direction. To put our research into context, we now review existing evidence for assessed sex differences and gender stereotype differences (GSDs) of personality traits with particular emphasis on developmental and cross-national patterns. In doing so, we adopt the hierarchical FFM of personality (McCrae & Costa, 2008), in which each of the higher order factors—Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C)—is composed of lower order facets that capture specific subcomponents.

Assessed Sex Differences

Previous research examining sex differences in assessed five-factor traits (Costa et al., 2001; Lippa, 2010; Lynn & Martin, 1997; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008; Soto, John, Gosling, & Potter, 2011) suggests that relative to their male counterparts, women score consistently higher on N and A. Women also tend to score higher on E, O, and C, but these findings are somewhat mixed across studies (e.g., Lynn & Martin, 1997; Schmitt et al., 2008; Soto et al., 2011). The discrepancy across studies may be due, in part, to divergent sex effects at the facet level. Compared with men, women tend to score higher on all facets of O except Openness to Ideas (O5) and higher on all facets of C except C1: Competence and C6: Deliberation (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005). For facets of E, men tend to score higher on E3: Assertiveness and E5: Excitement Seeking, whereas women score higher on E1: Warmth, E2: Gregariousness, and E6: Positive Emotions (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005). Thus, observed sex differences for E, O, and C may depend on the emphasis a given personality measures places on various subcomponents of the higher order factor. The present study, therefore, assessed GSDs at both the factor and the facet level.

Sex differences are not limited to self-reports (Costa et al., 2001); observer-ratings reveal similar patterns (Allik et al., 2009; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005). Moreover, effects for N and A have been found across a range of measurement instruments including those based on the five-factor taxonomy (e.g., Revised NEO Personality Inventory [NEO-PI-R], NEO Personality Inventory-3 [NEO-PI-3], and Big Five Inventory; Benet-Martinez & John, 1998; Costa & McCrae, 1992; De Bolle et al., 2013; Goodwin & Gotlib, 2004; Lehmann, Denissen, Allemand, & Penke, 2013; McCrae, Terracciano, & 78

Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008) as well as alternative models of personality (e.g., Eysenck Personality Questionnaire, Temperament and Character Inventory, and International Personality Item Pool; Lippa, 2010; Lynn & Martin, 1997; Miettunen, Veijola, Lauronen, Kantojarvi, & Joukamaa, 2007). Across assessment methods and studies, sex effects in assessed personality were found to be fairly small and rarely exceed .5 standard deviations (Chapman, Duberstein, Soerensen, & Lyness, 2007; Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008).

Assessed sex differences also show remarkable similarity across nations and cultural contexts. Large-scale studies comparing sex differences in personality in up to 55 cultures (Costa et al., 2001; Lippa, 2010; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008; Soto et al., 2011) found high agreement in the direction of sex effects. Despite this consistency, there are systematic cross-national differences in effect size: Sex differences are larger in Western, independent, and more developed nations (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008).

Much of the work on assessed sex differences in personality has focused on college-aged and adult samples. College-aged samples resemble adult samples in both the direction and magnitude of effects, although McCrae and colleagues (McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005) found somewhat larger sex differences for ratings of adults than for college student targets. Research on older samples is limited but in a study examining self-rated personality among those 65 and over, the direction of sex differences generally matched the patterns observed in younger samples (Chapman et al., 2007), although the effects for A were somewhat attenuated. Consistent with these findings, a large-scale longitudinal study (Terracciano, McCrae, Brant, & Costa, 2005) and a comprehensive meta-analysis (Roberts, Walton, & Viechtbauer, 2006) suggested that mean-level changes in personality across the adult life span do not vary by sex.

Somewhat more variability is seen in adolescence—in part because personality matures at different rates for girls and boys (Costa, McCrae, & Martin, 2008; De Bolle et al., 2013; Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009). Sex differences in N, E, and—in some studies—A are tend to be less pronounced among younger adolescents than among college-aged and adult samples (Branje, Van Lieshout, & Gerris, 2007; De Bolle et al., 2013; Klimstra et al., 2009; Vecchione, Alessandri, Barbaranelli, & Caprara, 2012). For C, several studies reported that sex differences (i.e., females scoring higher) are stronger among school-aged children than among adult respondents (De Bolle et al., 2013; De Fruyt, Van Leeuwen, De Bolle, & De Clercq, 2008; Klimstra et al., 2009), although other studies have not found this pattern (McCrae et al., 2002). For O, sex differences at the factor level tend to remain stable from adolescence to adulthood (Klimstra et al., 2009; Vecchione et al., 2012), although facet-level effects may show some increases (Costa et al., 2008; De Bolle et al., 2013). In summary, sex differences in personality show some variation over the course of adolescence but remain relatively stable from the college years and beyond.

Gender Stereotypes of Personality

In contrast to assessed sex differences, GSDs have rarely been examined from a FFM perspective. Prior research suggests that perceived gender differences in abilities and behavioral tendencies are fairly accurate depictions of actual effect sizes (Swim, 1994), but it is not clear whether such effects extend to FFM traits. There is, of course, a large body of research examining perceived gender differences in personal characteristics: Men are seen as higher in agentic and instrumental traits, whereas women are seen as higher in nurturing and communal traits (for reviews, see Diekman & Eagly, 2000; Eagly, Mladinic, & Otto, 1991; Kite, Deaux, & Miele, 1991). However, previous work has primarily relied on specialized measures like the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978) or the Bem Sex-Role Inventory

(BSRI; Bem, 1974), which were specifically developed to assess sex-typed personality traits and gender roles. These measures can be mapped onto the FFM via the interpersonal circumplex (Wiggins & Broughton, 1985). Masculinity in both instruments is strongly related to the Ambitious/Dominant pole (i.e., high E, low A, and specifically high E3: Assertiveness), whereas Femininity is related to the Warm-Agreeable pole (i.e., high E, high A, and specifically E1: Warmth). Although these patterns match assessed sex differences in A and facets of E, much less is known about the remaining FFM traits.

In the most comprehensive study to date, Williams et al. (1999) assessed GSDs in five-factor traits across 25 nations by asking respondents to rate whether men or women scored higher on a list of adjectives. Consistent with assessed sex differences, women were perceived to score higher than men on A and N. However, women were also perceived to score lower on E, O, and C, which does not match the typical patterns of assessed sex effects. Another study (Marcus & Lehman, 2002) asked respondents to rate the personality of unknown men and women on five items representing each of the five factors. Women were perceived to score higher than men on A, C, and O, but they were rated as similar to men on N and E—findings that are only partially consistent with the results by Williams et al. (1999). Thus, evidence for the direction of GSDs in five-factor traits is scarce and findings are inconsistent. Moreover, GSDs have only been assessed for the five higher order factors, and facet-level data are not available, which further limits comparisons with assessed sex differences. Finally, although the findings by Williams et al. (1999) suggest that GSDs of personality show similar levels of cross-national consistency as assessed sex differences, it is not clear whether stereotypes also mirror age trajectories of assessed personality in males and females. This is a critical gap because societies may vary in the ages at which boys and girls are expected to assume adult gender roles.

Underlying Mechanisms

A comprehensive examination of GSDs in FFM traits addresses an important gap in the research record by integrating the literature on assessed sex differences in five-factor traits with the literature on gender stereotypes, which has rarely employed a five-factor perspective. Moreover, examining developmental and cross-national patterns in GSDs and their convergence with actual sex differences can provide key insights about underlying mechanisms. With regard to assessed sex differences, theoretical perspectives have traditionally pitted *biological explanations* that emphasize the adaptive benefits of sex-differentiated behaviors and the role of evolutionary selection pressures (Baron-Cohen, 2003; Buss, 1997) against *sociocultural explanations* that emphasize gender-specific socialization processes (Ruble, Martin, & Berenbaum, 2006) and the influence of culture-based gender roles (Eagly, 1987; Ruble & Martin, 1998; for a review, see Best & Thomas, 2004). As noted previously, cross-cultural research on assessed sex differences suggests that they are remarkably consistent across a diverse range of nations (Costa et al., 2001; Lippa, 2010; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008; Soto et al., 2011). Together with growing evidence for genetically based hormonal mechanisms in the development of sex-specific behaviors and personality characteristics (for a review, see Hines, 2011), such findings suggest that biological factors are a key contributor to sex differences in personality, although their effects may be amplified by culturally based role expectations (Eagly, 1987), cognitive biases (Campbell, 1967; Stangor & McMillan, 1992), and social and economic conditions (Schmitt et al., 2008).

Much less is known about the mechanisms behind GSDs of personality. On one hand, they may reflect valid person perception and social judgment processes (Brunswik, 1952; Funder, 1995) that draw on personal experiences with individual men and women as well as aggregate information about sex differences at the group level. From this perspective, one would expect that GSDs are fairly consistent with assessed sex differences in terms of direction, size, and

developmental trajectories and relatively invariable across diverse cultural backgrounds. On the other hand, GSDs may be influenced by culture-specific role expectations (Eagly, 1987), socialization processes (Ruble et al., 2006), and gender inequities in the distribution of political and economic power. Such factors would likely increase cross-national variability and promote associations with nation-level indicators of gender equality, value systems, and social norms. Independent of the strength of gender roles, cultural constraints in social norms may also limit the range of acceptable behaviors for everyone and thus curtail the expression of intrinsic sex differences in personality traits (Gelfand et al., 2011). Of course, these pathways are not mutually exclusive. Thus, rather than pitting one against the other, the present study aimed to examine the relative association of empirically observed GSDs with assessed sex differences as compared with sociocultural variables. Although this correlational approach cannot explicitly address the underlying causal mechanisms, it can provide initial evidence for the differential importance of biological versus cultural explanations.

The Present Study

The present study examined perceived gender differences in FFM factors and facets for three different age groups in a large cross-national sample. Specifically, we focused on consensual stereotypes of men and women of different ages that were aggregated across individual respondents within a given national context. This approach was chosen because it simplifies comparisons with prior results on assessed sex differences in five-factor traits, which were also reported at the aggregate level.

Our work extended prior research in several respects. First, we assessed the direction and magnitude of GSDs at both the factor and facet levels. Second, we examined the accuracy of GSDs relative to assessed sex differences reported in the prior literature. Third, we explored the extent to which GSDs are consistent across targets of different ages and compared age trends in stereotypes with age patterns in assessed sex differences. Fourth, we examined the degree of universality in GSDs across 26 nations, with particular emphasis on contrasts between Asian and Western nations, which were previously found to differ in the size of assessed sex differences (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008). Finally, we explored associations between the magnitude of GSDs and nation-level indicators of socioeconomic development, gender equality, value systems, and the strength of social norms, which may suggest reasons for cultural variation in gender differentiation (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; Schmitt et al., 2008).

Method

Participants and Procedure

Participants ($N = 3{,}470$) were recruited from 26 nations.[2] After excluding 147 participants due to poor data quality (i.e., incomplete or random responding), 3,323 participants were included in the present study. Table 1 reports sample sizes for each nation (median n/site = 100, range 49-283). The combined sample was 31% male with a mean age of 22.3 years. Most participants were university students in their early 20s, but for three nations (Italy, South Korea, and the United States), subsamples included community-dwelling adults.[3] For a detailed description of recruitment procedures, data collection, and sample characteristics by country, refer to Löckenhoff et al. (2009) and Chan et al. (2012).

Participants were asked to rate the personality traits of the typical adolescent, adult, and old person in their nation on the National Character Survey (NCS; Terracciano, Abdel-Khalek, et al., 2005). They were randomly assigned to rate either male or female targets, and the order of target

Table 1. Mean Differences Between Stereotypes of Females and Males in 26 Nations on Trait Factors or Composites.

			Mean difference by domain					
Nation	*N*	Item/total[a]	N factor	A factor	F-Ex/In	F-Op/Cl	F-Co/Un	M[b]
India	49	.06	.01	−.24	−.14	.18	−.08	−.06
Switzerland	97	.31**	.04	.19	−.04	.06	.09	.07
Peru	136	.47**	.02	.09	.05	.21	.16	.11
Italy	151	.33**	.20	.09	.07	.14	.14	.13
Iran	113	.57**	.01	.27	.10	.08	.26	.14
Argentina	128	.48**	.09	.23	.08	.22	.16	.16
Uganda	98	.49**	.17	.28	.12	.08	.14	.16
Chile	87	.40**	.23	.21	.07	.15	.15	.16
Japan	271	.52**	.10	.31	.17	.11	.10	.16
P.R. China	91	.31**	.29	.23	.16	−.01	.19	.17
Hong Kong	162	.65**	.07	.36	.20	.15	.15	.19
France	100	.64**	.29	.33	.06	.17	.18	.21
South Korea	118	.43**	.15	.47	.19	.15	.10	.21
Serbia	94	.59**	.21	.29	.13	.19	.27	.22
Russia	94	.66**	.30	.47	.24	.11	.27	.28
Poland	193	.71**	.40	.46	.12	.23	.19	.28
New Zealand	94	.79**	.33	.53	.16	.25	.13	.28
Portugal	89	.49**	.36	.44	.15	.24	.24	.29
Australia	90	.77**	.51	.41	.14	.26	.12	.29
Croatia	96	.76**	.34	.49	.12	.28	.22	.29
Malaysia	100	.64**	.31	.54	.20	.16	.26	.29
United Kingdom	93	.77**	.40	.56	.22	.29	.10	.32
United States	316	.87**	.38	.64	.23	.30	.16	.34
Estonia	110	.80**	.43	.59	.17	.30	.23	.34
Czech Republic	215	.73**	.47	.61	.16	.27	.22	.35
Slovakia	138	.84**	.50	.58	.13	.23	.44	.38

Note. Nations are listed in the increasing order of total gender differentiation. N = Neuroticism; A = Agreeableness; F-Ex-In = Feminine Extraversion/Introversion; F-Op/CL = Feminine Openness/Closedness; F-Co/Un = Feminine Conscientiousness/Unconscientiousness.

[a]Correlation ($N = 90$) of national means for each facet and age group with means for all other nations.

[b]Mean across domains = total stereotype differentiation score.

**$p < .01$.

age group was counterbalanced within participants. Participants used the official/primary language of their respective country, and the majority responded in group settings.

Measures

NCS. The NCS (Terracciano, Abdel-Khalek, et al., 2005) comprises 30 bipolar items that correspond to the facets of the NEO-PI-R (Costa & McCrae, 1992), with six facets representing each of the five higher order traits. For instance, to obtain ratings of the typical Swiss adolescent boy on the anxiety facet of neuroticism, Swiss participants were presented with the prompt "In Switzerland, adolescent boys are likely to be . . . " and asked to provide their answer on a 5-point scale between the anchors of *anxious, nervous, worrying* and *at ease, calm, relaxed* (for a complete list of individual items, see the online supplement to Terracciano, Abdel-Khalek, et al., 2005). For

non-English-speaking nations, an initial translation of the scale was provided by local collaborators who were bilingual psychologists. Back-translations were performed by individuals other than the initial translator and reviewed by two of the authors (F.D.F. and R.R.M.). If needed, further modifications were made to ensure equivalence.

Cronbach's alpha for the scales corresponding to each factor ranged from .62 to .77 and when rotated toward the U.S. structure of the NEO-PI-R, factor congruence coefficients ranged from .74 to .91 suggesting that the FFM is adequately represented in the NCS (for more detailed analyses, see Chan et al., 2012). For the present analyses, factor scores were calculated as weighted sums of the individual items using NEO-PI-R scoring weights (Costa & McCrae, 1992). Compared with simply summing the six items representing each factor, factor scores are more nearly orthogonal and have better psychometric characteristics (Costa & McCrae, 1992).

We calculated GSDs by *z*-standardizing ratings for each facet and factor in the full sample and subtracting scores for male targets from those for female targets within age groups and nations, across age groups within each nation, and for the sample as a whole.[4] Because participants described either male or female targets, difference scores are reported at the group level, not the individual level. Note, however, that in previous analyses, male and female raters did not differ significantly in their NCS ratings (see Chan et al., 2012, for details).

Comparison Samples

To estimate the accuracy of GSDs of personality, we drew on the prior literature to obtain sex difference scores based on self-ratings as well as observer-ratings of actual males and females provided by individuals who knew them well.

Self-reports for college-aged and adult men and women were drawn from Costa and colleagues (2001), who compiled NEO-PI-R data from 11,690 adults and 10,952 college-aged respondents[5] from 26 nations. Fourteen of these nations overlap with the present sample.

Observer-ratings for college-aged and adult targets were drawn from the Personality Profiles of Cultures Project (PPOC; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005), which provides NEO-PI-R observer-ratings for 5,095 college-aged individuals ($M = 19.8$ years) and 6,128 adults ($M = 49.9$ years) from 50 nations. Twenty-five of these nations overlap with the ones considered in the present study.

Observer-ratings for individuals in early and middle-adolescence were drawn from the Adolescent Personality Profiles of Cultures Project (APPOC; De Bolle et al., 2013; De Fruyt et al., 2009), which provides ratings on the NEO-PI-3 (McCrae, Costa, & Martin, 2005), a more readable version of the NEO-PI-R. Observer-ratings are available for 5,109 adolescents in 24 nations, 20 of which overlap with the present study.

Nation-Level Indicators

Socioeconomic development. Levels of socioeconomic development for each nation were assessed with the *Human Development Index* (HDI; $n = 26$), a composite of three aspects of development: life expectancy at birth, access to education (adult literacy and gross enrollment ratio across grade levels), and standard of living (gross domestic product per capita). Higher scores indicate higher development. Data for 2005 were retrieved from http://hdr.undp.org/en/media/HDR05_complete.pdf

Gender equality. Gender equality was assessed with two indicators: The *Gender-Related Development Index* (GDI; $n = 25$), which captures equalities in the HDI dimensions of life expectancy (male vs. female life expectancy at birth), education (male vs. female adult literacy and gross

enrollment ratio), and standard of living (male vs. female estimate earned income), and the *Gender Empowerment Measure* (GEM; $n = 20$), which captures women's involvement in political decision making (male vs. female parliamentary seats) and economic decision making (male vs. female shares in legislative, managerial, professional, and technical positions). Higher scores indicate greater equality and empowerment. Data for 2005 were retrieved from http://hdr.undp.org/en/media/HDR05_complete.pdf

Value systems. Cultural differences in value systems were captured along the dimensions proposed by Hofstede (1980, 2001; that is, power distance, uncertainty avoidance, individualism, and masculinity, $n = 26$), Schwartz (2005; that is, embeddedness, affective autonomy, intellectual autonomy, hierarchy, mastery, egalitarian commitment, and harmony, $n = 26$), and Inglehart and Baker (2000; that is, traditional values vs. self-expression values, $n = 24$). For a detailed description of each indicator, see Löckenhoff et al. (2009).

Social norms. Cultural differences in the strength of social norms were quantified with Gelfand and colleagues' (2011) indicator of societal tightness–looseness, which captures the degree of tolerance toward deviant behaviors as compared with pressure to conform to social norms ($n = 14$).

Results

Patterns and Magnitude of GSDs

Table 2 (first column) reports *z*-score differences between stereotypes of men and women (collapsed across age groups and nations) with positive values indicating higher scores for women. To account for the large sample size and the calculation of multiple comparisons at the facet level, we limit our interpretation to difference scores that are larger than .1 ($p < .001$).

As seen in the table, women were rated higher than men on all of the higher order factors except E. This pattern was most pronounced for A, followed by C, O, and N. At the facet level, the direction of effects for the facets of O, A, and C was consistent with the effects for the corresponding factor. N and E, in contrast, showed divergent effects for individual facets. Among the facets of N, women were rated higher than men on N1: Anxiety and N6: Vulnerability, lower than men on N5: Impulsiveness, and about the same on the remaining facets. Among the facets of E, women were rated higher than men on E1: Warmth and E6: Positive Emotions, lower than men on E3: Assertiveness and E5: Excitement Seeking, and about the same on the remaining facets.

Convergence With Assessed Sex Differences

To examine the convergence between GSDs and assessed sex differences, the right two columns of Table 2 present data from comparison samples based on self-reports and observer-ratings of actual individuals. A comparison of the mean absolute value of effects summed across the facets within each column (bottom row) indicated that GSDs (.17) were similar in size to assessed sex differences in both self- (.20) and observer-ratings (.17). These data suggested that GSDs did not exaggerate sex differences in actual self- and observer-ratings. Also note that none of these average effect sizes exceeded one fifth of a standard deviation.

To examine whether GSDs showed similar profiles across the 30 facets as assessed sex differences, we computed Pearson correlations between the three columns in Table 2. The observed associations were substantial and statistically significant.[6] GSDs appeared to be more closely associated with assessed sex differences in observer-ratings ($r = .67$, $p < .001$) as compared with self-ratings ($r = .47$, $p < .01$), although the difference in correlation coefficients did not reach

Table 2. Mean z-Score Differences in Gender Stereotypes and Assessed Sex Differences in Self- and Observer-Ratings.

Factor or facet	Stereotype	Assessed self-reports	Assessed observer-ratings
N: Neuroticism	.26	.48	.43
E: Extraversion	.06	.18	.12
O: Openness	.26	.09	.15
A: Agreeableness	.40	.45	.25
C: Conscientiousness	.28	.11	.27
N1: Anxiety	.20	.38	.38
N2: Angry hostility	−.06	.15	.09
N3: Depression	.01	.23	.20
N4: Self-consciousness	.04	.25	.20
N5: Impulsiveness	−.14	.17	−.05
N6: Vulnerability	.12	.36	.23
E1: Warmth	.25	.27	.18
E2: Gregariousness	.03	.18	.23
E3: Assertiveness	−.16	−.19	−.03
E4: Activity	.05	.09	.10
E5: Excitement seeking	−.12	−.29	−.23
E6: Positive emotions	.14	.24	.21
O1: Fantasy	.05	.01	.06
O2: Aesthetics	.37	.36	.40
O3: Feelings	.45	.31	.36
O4: Actions	.05	.16	.12
O5: Ideas	.10	−.22	−.16
O6: Values	.11	.03	.02
A1: Trust	.24	.15	.11
A2: Straightforwardness	.05	.36	.10
A3: Altruism	.27	.31	.16
A4: Compliance	.28	.19	.05
A5: Modesty	.23	.27	.16
A6: Tender-mindedness	.39	.28	.28
C1: Competence	.09	−.13	.00
C2: Order	.41	.08	.27
C3: Dutifulness	.22	.10	.20
C4: Achievement striving	.13	.03	.12
C5: Self-discipline	.12	.04	.17
C6: Deliberation	.14	−.07	.12
M absolute value across facets	.17	.20	.17

Note. Positive values indicate higher scores in women. Facet scores for assessed self-reports are unweighted means across college-aged and adult men and women from Costa, Terracciano, and McCrae (2001); scores for assessed observer-ratings are unweighted means from boy and girls from De Bolle et al. (2013) and college-aged and adult men and women from McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project (2005). Factor difference scores are calculated by applying factor scoring weights (Costa & McCrae, 1992) to facet difference scores in each column. The bottom row shows the mean absolute value of gender stereotype and sex differences across facets.

statistical significance. However, the strongest association was found between sex differences in assessed self- and observer-ratings ($r = .78$, $p < .001$).

Figure 1 visually illustrates associations between GSDs and assessed sex differences in self-reports (top) and observer-ratings (bottom). In general, there appeared to be considerable

agreement between GSDs and sex differences. For instance, assessments and stereotypes agreed that women were higher in O2: Aesthetics, O3: Feelings, as well as A6: Tender-Mindedness, whereas men were higher in E3: Assertiveness. Nevertheless, there were some notable discrepancies. Across all facets of N (and for N1: Anxiety in particular), assessed sex differences appeared to be more pronounced than GSDs, and this was true for both self-reports and observer-ratings. In contrast, GSDs in the facets of Conscientiousness tended to be stronger than assessed sex differences—particularly in self-reports. Another notable outlier was O5: Ideas: Stereotypes suggested that women scored higher on this facet, whereas self-reports and observer assessments showed the reverse. In summary, GSDs appeared to be remarkably consistent with assessed sex differences, although some deviations were seen, especially at the facet level.

Age Differences

The analyses reported so far collapsed GSDs across target age groups. In a next step, we examined whether the developmental patterns that have been found for assessed sex differences would extend to GSDs. Figure 2 plots age trends in GSDs for the five higher order factors. Three types of patterns emerged: Perceived gender differences in N and C were somewhat higher for adolescent as compared with adult targets but dropped precipitously for old targets. For O and A, there was a slight increase in GSDs from adolescence to adulthood and a steep drop toward old age. GSDs in E, finally, were close to zero for adolescent and adult targets; only in the old group, we found a small gender effect.

These age trends in GSDs show only limited convergence with age patterns in assessed sex effects. From adolescence to adulthood, both GSDs and sex differences showed consistent increases for O and A but decreases for C. For N and E, in contrast, assessed sex effects increased over the course of adolescence, whereas GSDs showed stability for E and a slight decline for N. With regard to old age, prior research on assessed differences has found a selective attenuation of sex differences in A (Chapman et al., 2007). GSDs, in contrast, showed a marked attenuation for all factors except E. This trend was also reflected in the mean absolute value of GSDs computed across the 30 facets, which was almost twice as large for adolescents (.19) and adults (.21) relative to older adults (.11).

Conceivably, the college-aged respondents in our sample may have had more difficulty in accessing information about typical older adults as compared with targets closer to their own age. This could have led to more random responding, resulting in an age-related attenuation of GSDs. However, if random responding played a role, one would have expected that GSDs for older targets were not only smaller but also less accurate, resulting in lower concordance with actual self-reports and observer-ratings. To examine this possibility, Table 3 presents correlations—across the 30 facets—between age-specific GSDs and age-specific assessed sex differences. All correlations were significant, and the degree of concordance between stereotypes and assessed sex differences was comparable in size across age groups, speaking against an explanation based on random responding. To sum up, college-aged raters attributed a similar pattern of gender differences to old targets as to younger targets, although they perceived them to be generally smaller.

There was also a subtler and perhaps more interesting finding (Table 3). Among observer-ratings, there was a pattern indicating higher agreement when GSDs of a given age group were (roughly) matched with assessed sex differences for that group. In other words, age trends in assessed sex differences across the 30 facets were paralleled by developmental trends in perceived gender stereotypes, suggesting sensitivity to the nuances of life span development.

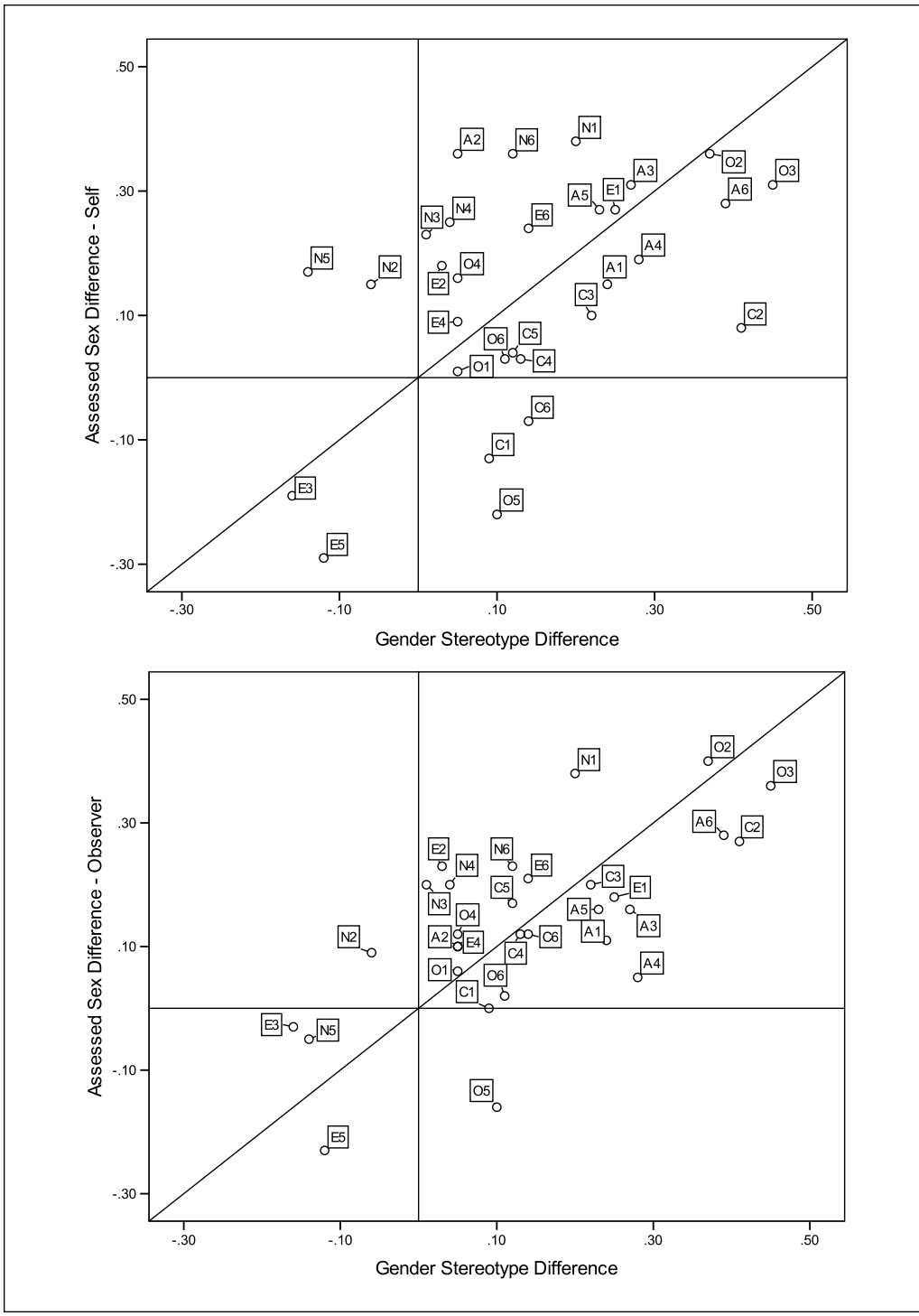

Figure 1. Gender stereotype differences plotted against assessed sex differences in self-ratings (top) and observer-ratings (bottom) for the 30 facets across all age groups.
Note. Positive values indicate women score higher than men. See Table 2 for full facet labels.

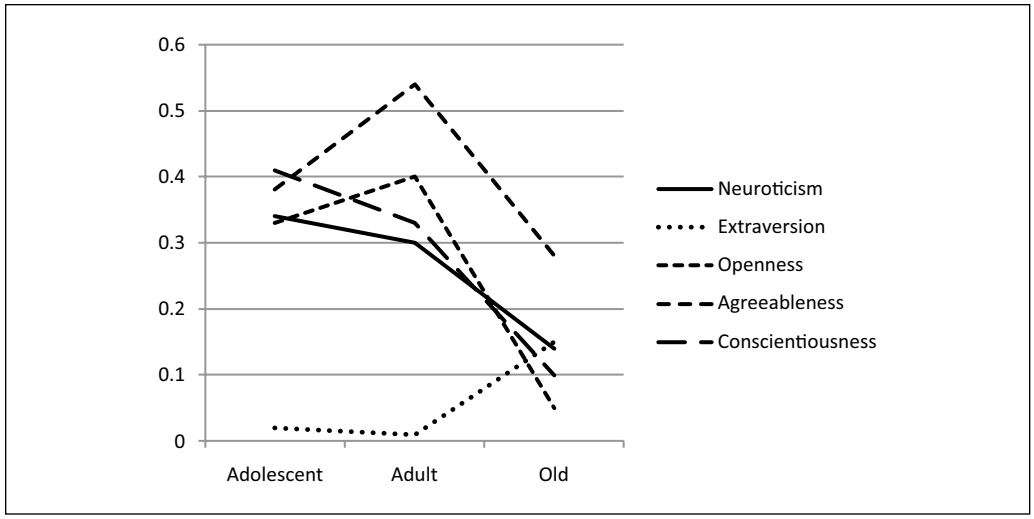

Figure 2. Age trends in the size of gender stereotype differences.

Table 3. Correlations Across Facets of Age-Specific GSDs With Age-Specific Sex Differences.

	GSDs		
Assessed sex differences	Adolescent	Adult	Old
Self-reports			
College age[a]	.36*	.45*	.46**
Adult[a]	.40*	.49**	.50**
Observer-ratings			
Early adolescence (12-14)[b]	.70***†	.60***	.41*†
Middle adolescence (15-17)[b]	.63***	.54**	.45*
College age (18-21)[c]	.48**	.41*	.43*
Adult (40-98)[c]	.47**‡	.57***	.67***‡

Note. GSD = gender stereotype differences. $N = 30$. Correlations with the same superscript (†/‡) are significantly different, $p < .05$, one-tailed.
[a]From Costa, Terracciano, & McCrae (2001).
[b]From De Bolle et al. (2013).
[c]From McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project (2005).
*$p < .05$. **$p < .01$. ***$p < .001$.

National Differences

To explore cross-national concordance in the direction, size, and age trends of GSDs, we first examined the degree of similarity in the profiles of perceived gender differences across the 30 facet ratings within each of the three age groups. To this end, we calculated Cronbach's alpha treating the 90 facet ratings (30 within each age group) as cases and the 26 nations as items. The resulting alpha score was .94, suggesting a fair amount of cross-national similarity. Since Cronbach's alpha may have been skewed by the fairly high number of items (i.e., 26), we also examined the relative consistency of GSDs within each nation with the overall pattern. To this end, we computed item–total correlations, that is, correlations between the means of the age-specific facet ratings for each nation and the means averaged across all other nations (Table 1).

With the exception of India (n.s.), all correlations reached statistical significance (range = .31-.87, median = .61, all *p*s < .01). In combination, these results indicated considerable cross-national consistency in the direction of GSDs across age groups and facets.

We also computed an index capturing the degree to which GSD differences within each nation map onto cross-national patterns in the magnitude and direction of assessed sex differences (Costa et al., 2001; McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005). Within a given nation, we collapsed GSD scores across age groups and summed them into factor scores based on the scoring procedure described in McCrae, Terracciano, and 78 Members of the Personal Profiles of Cultures Project (2005). Specifically, factor scores for N and A were computed by summing each of the six corresponding facets. Scores for E, O, and C, in contrast, were composed of facets that were previously found to show consistent sex differences across nations (McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005). Feminine Extraversion/Introversion was calculated by summing facet scores for E1, E2, and E6, subtracting scores for E3 and E5, and dividing the result by five. Feminine Openness/Closedness was calculated by summing scores for O2, O3, and O4, subtracting the score for O5, and dividing the result by 4. Feminine Conscientiousness/Unconscientiousness, finally, was calculated by summing scores for C2 and C3, subtracting the score for C1, and dividing the result by 3 (see McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005).

As seen in Table 1 (middle columns), 96% of the resulting scores were positive, indicating high consistency across nations. The magnitude and direction of effects was also similar across the five domains within a given nation; all intercorrelations among the five middle columns of Table 1 were positive (median $r = .45$) and the alpha (treating domains as items and nations as cases) was .82. Thus, it was reasonable to compute a total stereotype differentiation score (i.e., the mean across the five domain composites; Table 1, right column), which captures the average size of GSDs across domains within a given nation.

We then used this stereotype differentiation score to examine the degree to which cross-national patterns in the size of GSDs mapped onto cross-national patterns in assessed sex differences as well as nation-level indicators of socioeconomic development, gender equality, value systems, and the strength of social norms. To account for deviations from normality in stereotype differentiation scores, we computed rank-order correlations (Spearman's ρ).

Cross-national associations between stereotype differentiation scores and the size of observer-rated sex differences were significant across the 20 overlapping nations for adolescent targets in the APPOC (De Bolle et al., under review, $\rho = .69$, $p < .01$) and marginally significant across the 25 overlapping nations for college-aged and adult targets in the PPOC (McCrae, Terracciano, & 78 Members of the Personal Profiles of Cultures Project, 2005; $\rho = .39$, $p = .05$). For self-reports, access to appropriate comparison data was limited. The article by Costa and colleagues (2001) only reported sex differentiation scores for four of the factors (no data on Feminine Conscientiousness/Unconscientiousness) and only 14 nations overlapped with the present sample. Although the correlation between differentiation scores was positive, it did not reach statistical significance ($\rho = .40$, $p = .15$). Taken together, these analyses imply that respondents from nations with larger assessed sex differences in personality also showed larger GSDs, particularly in observer-ratings.

Consistent with prior research comparing the size of assessed sex differences in Asian versus Western cultural contexts, stereotype differentiation scores were significantly smaller in Asian nations (Hong Kong, P.R. China, India, Iran, Japan, Malaysia, and South Korea; $M = .16$, $SD = .11$) than in Western nations (the United States, Australia, New Zealand, and all European countries; $M = .27$, $SD = .08$), $t(20) = 2.73$, $p < .05$.[7] In contrast, although the general direction of effects pointed toward larger stereotype differentiation scores in more developed and egalitarian nations, none of the associations with socioeconomic development (HDI), gender equality (GDI

and GEM), or cultural values (as defined by Hofstede, 2001; Inglehart & Baker, 2000; Schwartz, 2005) reached statistical significance (for all associations $|\rho| < .35$, $p > .1$). However, although indicators of tightness–looseness (Gelfand et al., 2011) were only available for 14 of the nations, we found a noteworthy trend suggesting that GSDs were less pronounced in tighter nations ($\rho = -.50$, $p = .07$).

Discussion

The present study adds to our understanding of gender stereotypes in five-factor personality traits in several ways. In most of the 26 nations examined, women were perceived as higher than men in O, A, and C as well as facets of E and N, and in general, GSDs mapped well onto assessed sex differences. This finding is consistent with earlier work reporting considerable descriptive accuracy for gender stereotypes about abilities and behavioral tendencies (Swim, 1994) as well as the broader literature that suggests that there may be a "kernel of truth" in other forms of stereotypes including those about age (Chan et al., 2012) and race (for reviews, see Ryan, 2002, and Jussim, 2012).

From a theoretical point of view, our findings suggest that—to some extent—GSDs reflect valid social judgments about the size and direction of sex differences in personality (Brunswik, 1952; Funder, 1995), although some important caveats apply. First, it should be noted that—like sex differences—GSDs of personality were small in size and did not exceed one third of a standard deviation. Second, we examined *consensual stereotypes* aggregated over multiple respondents; *personal stereotypes* held by individual respondents are likely to be much less accurate (see Chan et al., 2012). Third, we define accuracy as statistical agreement of perceived and assessed differences between men and women. Thus, our findings do not speak to the accuracy of beliefs about the underlying *causes* of sex differences (e.g., biological vs. sociocultural). Finally, our findings do not question the reality of cognitive biases in the perception of sex differences (Campbell, 1967; Stangor & McMillan, 1992) but merely suggest that accurate judgments can emerge in spite of such biases.

We also found some notable discrepancies between assessed and perceived sex differences. In part, such discrepancies may reflect differences in item wording between the personality questionnaires used to obtain self- and observer-ratings (e.g., NEO-PI-R, BFI) and the adjective-based NCS that was used to assess stereotypes in the present study (see Allik, Mottus, & Realo, 2010). At a more substantive level, if gender stereotypes are considered as reflections of actual sex differences, they may be more pronounced for readily observable characteristics such as differences in orderliness and organization (i.e., C) than for internal experiences of negative emotionality (i.e., N).

Gender stereotypes were also sensitive to target age. Most notably, gender effects were substantially smaller for older adult targets than for adults and adolescents, an effect that appears to match the pattern seen in assessed sex differences for A (Chapman et al., 2007). Raters may have been influenced by aging stereotypes suggesting that older adults lose some of their differentiated gender identities with the end of child-rearing roles in women and the end of work-related roles in men (Gailey, 1987; Gutmann, 1985). With regard to direction, however, the accuracy of gender stereotypes did not differ by target age, and accuracy was higher when stereotypes of a given age group were compared with assessed sex differences in that group. Thus, GSDs reflect some aspects of age trends in sex differences.

With regard to cross-national patterns, several notable findings emerged. First, like assessed sex differences, GSDs showed remarkable cross-national consistency not only in direction but also in developmental patterns across age groups. The latter finding suggests that GSDs are more closely aligned with universal maturational changes in male and female personality than with gender role expectations that may vary considerably across cultures in terms of content and

timing. However, although directions and age effects in GSDs were quite consistent, the average size of GSDs (summed across factors) systematically varied across nations: They were more pronounced among nations with stronger sex differences in assessed personality. Our findings also provide a new perspective on somewhat counterintuitive findings in the prior literature indicating that assessed sex differences are stronger in highly developed, egalitarian, and Western contexts (Costa et al., 2001; Schmitt et al., 2008). Substantive explanations of such effects have focused on differences in the size and expression of genetically based sex differences (Lippa, 2010) in different human groups, whereas artifactual explanations have emphasized the role of attribution (i.e., stronger gender norms in some nations may lead raters to discount sex-stereotypic behavior; Costa et al., 2001) and frame of reference effects (i.e., raters in gender-segregated societies may assess personality relative to the target's own gender, whereas raters in egalitarian societies assess personality relative to both men and women; Guimond et al., 2007). In the present study, we found that—like assessed sex differences—GSDs were somewhat more pronounced in Western versus Asian nations. However, in contrast to assessed sex differences, the magnitude of GSDs was not significantly associated with national indicators of gender equality and human development. Instead, we found a trend toward smaller gender stereotypes in tighter nations (Gelfand et al., 2011). Conceivably, nations with stronger social norms and lower tolerance for deviant behavior may curtail the behavioral range for both men and women, and thus limit opportunities to express and observe innate sex differences in personality. Alternatively, tighter nations may enforce sex-role differentiated behaviors and thus encourage both attribution and frame of reference effects. Although our correlational findings, of course, cannot directly speak to causal mechanism, these considerations suggest that our findings are more consistent with artifactual (as compared with substantive) interpretations of cross-national differences in GSDs. However, given that tightness–looseness data were only available for a small subset of the nations and effects remained at the trend level, these findings need to be corroborated by future research.

Several additional limitations need to be considered when interpreting our results. First, respondents rated either men or women. As such, gender differences were examined at the group, not the individual level, and raters did not serve as their own controls, which likely increased random error and may have led to an underestimation of effect sizes. On the positive side, our design also avoided contrast effects that are a concern when the same respondent rates both males and females. In any case, it is reassuring to note that the effect sizes in the present study are comparable with those reported in prior research on gender stereotypes of personality (Williams et al., 1999).

Estimates of accuracy need to be interpreted with caution as well. First, although assessed sex differences were drawn from the best available comparison samples, the samples differed in national composition, age groupings, and assessment approaches from those in the present study and such discrepancies may have led to an underestimation of accuracy. Second, even though assessed sex differences are based on personality ratings for specific individuals (i.e., self or familiar other), they may still be affected—to some extent—by implicit theories about male and female personality held by the rater. This might have inflated estimates of accuracy, particularly for observer-ratings. Moreover, questionnaire-based personality measures such as the NEO-PI-R may show cross-cultural variations in response styles and item relevance that would also reduce the validity of our accuracy estimates.

A final concern is our reliance of samples of convenience that largely drew on student populations. Conceivably, students may be more likely than the general population to support egalitarian beliefs, thus leading to an underestimation of stereotypes. Ideally, future research would therefore recruit samples representative of the population in each nation and collect assessed sex differences and gender stereotypes for identical age groups from the same raters on parallel instruments designed to minimize cross-national variations in measurement characteristics.

Authors' Note

Löckenhoff, McCrae, Costa, and Terracciano were U.S. government employees at the time of data collection, Chan was a U.S. government employee at the time of manuscript preparation.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: McCrae and Costa receive royalties from the NEO Inventories.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging. Realo and Allik were supported by grants from the Estonian Ministry of Education and Science (SF0180029s08 and IUT2-13). Hřebíčková and Graf were supported by the Czech Science Foundation (13-25656S) and by the Institute of Psychology, Academy of Sciences of the Czech Republic (RVO: 68081740).

Notes

1. In the remainder of this article, we will use the term "sex differences" to refer to assessed differences between men's and women's personalities and the term "gender stereotype differences" (GSDs) to refer to respondents' perceptions of such differences.
2. Note that we distinguish the Hong Kong Special Administrative Region from the rest of the People's Republic of China because it differs significantly in language, political administration, and cultural background.
3. Previous analyses comparing student and adult samples in this subset of cultures suggested that patterns of responses are comparable across raters of different ages (Chan et al., 2012).
4. Each participant rated targets from three age groups in counterbalanced order, raising concerns about possible age contrast effects. To some extent, computing difference scores within age groups protects against such effects. Also, supplemental analyses indicated that the pattern of results was comparable if only the first of each participant's responses was analyzed. Another concern is the appropriateness of computing difference scores across the whole sample, given that sample sizes differed by country. Supplemental analyses found a similar pattern of results when a combination of the 26 culture means instead of the total sample mean was used.
5. Because data were collected by multiple investigators, data on the specific age distribution are not available.
6. The *p*-values for these and other facet-level correlations should be interpreted cautiously because NEO facets are not fully independent. Note, however, that non-parametric analyses (Spearman's ρ) yielded the same pattern of findings.
7. Consistent with our prior work (Löckenhoff et al., 2009), we used the United Nations geographical regions to categorize nations into Asian versus Western categories.

References

Allik, J., Mottus, R., & Realo, A. (2010). Does national character reflect mean personality traits when both are measured by the same instrument? *Journal of Research in Personality*, *44*, 62-69. doi:10.1016/j.jrp.2009.10.008

Allik, J., Realo, A., Mõttus, R., Pullmann, H., Trifonova, A., & McCrae, R. R., & 56 Members of the Russian Character and Personality Survey. (2009). Personality traits of Russians from the observer's perspective. *European Journal of Personality*, *23*, 567-588. doi:10.1002/per.721

Baron-Cohen, S. (2003). *The essential difference: The truth about the male and female brain*. New York, NY: Basic Books.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155-162.

Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, *75*, 729-750. doi:10.1037//0022-3514.75.3.729

Best, D. L., & Thomas, J. J. (2004). Cultural diversity and cross-cultural perspectives. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The psychology of gender* (pp. 296-327). New York, NY: Guilford.

Branje, S. J. T., Van Lieshout, C. F. M., & Gerris, J. R. M. (2007). Big five personality development in adolescence and adulthood. *European Journal of Personality*, *21*, 45-62. doi:10.1002/per.596

Brunswik, E. (1952). *The conceptual framework of psychology*. Oxford, UK: University of Chicago Press.

Buss, D. M. (1997). Evolutionary foundations of personality. In R. Hogan (Ed.), *Handbook of personality psychology* (pp. 317-344). London, England: Academic Press.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist*, *22*, 817-829. doi:10.1037/h0025079

Chan, W., McCrae, R., De Fruyt, F., Jussim, L., Löckenhoff, C. E., De Bolle, M., . . . Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology*, *103*, 1050-1066.

Chapman, B. P., Duberstein, P. R., Soerensen, S., & Lyness, J. M. (2007). Gender differences in Five Factor Model personality traits in an elderly cohort. *Personality and Individual Differences*, *43*, 1594-1603. doi:10.1016/j.paid.2007.04.028

Costa, P. T., & McCrae, R. R. (1992). *Professional manual: Revised NEO Five-Factor Inventory*. Lutz, FL: Psychological Assessment Resources.

Costa, P. T., McCrae, R. R., & Martin, T. A. (2008). Incipient adult personality: The NEO-PI-3 in middle-school-aged children. *British Journal of Developmental Psychology*, *26*, 71-89. doi:10.1348/026151007x196273

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. doi:10.1037/0022-3514.81.2.322

De Bolle, M., De Fruyt, P., McCrae, R. R., Löckenhoff, C. E., Costa, P. T. Jr., Aguilar-Vafaie, M. E., . . . Terracciano, A. (2013). *Sex differences in adolescent personality: A cross-cultural study*. Manuscript submitted for publication.

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., Costa, P. T., & Collaborators Adolescent, P. (2009). Assessing the Universal Structure of Personality in Early Adolescence The NEO-PI-R and NEO-PI-3 in 24 Cultures. *Assessment*, *16*, 301-311. doi:10.1177/1073191109333760

De Fruyt, F., Van Leeuwen, K., De Bolle, M., & De Clercq, B. (2008). Sex differences in school performance as a function of conscientiousness, imagination and the mediating role of problem behaviour. *European Journal of Personality*, *22*, 167-184. doi:10.1001/per.675

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*, 1171-1188. doi:10.1177/0146167200262001

Eagly, A. H. (1987). *Sex differences in social behavior: A social role interpretation*. Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men: An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly*, *15*, 203-216. doi:10.1111/j.1471-6402.1991.tb00792.x

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*, 429-456. doi:10.1037/0033-2909.116.3.429

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652-670. doi:10.1037/0033-295x.102.4.652

Gailey, C. W. (1987). Evolutionary perspectives on gender hierarchy. In B. B. Hess & M. M. Ferree (Eds.), *Analyzing gender: A handbook of social science research* (pp. 32-67). Thousand Oaks, CA: Sage.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. A., Lun, J., Lim, B. C., . . . Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*, 1100-1104. doi:10.1126/science.1197754

Goldberg, L. R. (1990). An alternative description of personality—the big-five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.

Goodwin, R. D., & Gotlib, I. H. (2004). Gender differences in depression: The role of personality factors. *Psychiatry Research*, *126*, 135-142. doi:10.1016/j.psychres.2003.12.024

Guimond, S., Branscombe, N. R., Brunot, S., Buunk, A. P., Chatard, A., Desert, M., . . . Yzerbyt, V. (2007). Culture, gender, and the self: Variations and impact of social comparison processes. *Journal of Personality and Social Psychology*, *92*, 1118-1134. doi:10.1037/0022-3514.92.6.1118

Gutmann, D. L. (1985). The parental imperative revisited: Towards a developmental psychology of adulthood and later life. *Contributions to Human Development*, *14*, 31-60.

Hines, M. (2011). Gender development and the human brain. *Annual Review of Neuroscience*, *34*, 69-88. doi:10.1146/annurev-neuro-061010-113654

Hofstede, G. (1980). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, *65*, 19-51.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York, NY: Oxford University Press.

Kite, M. E., Deaux, K., & Miele, M. (1991). Stereotypes of young and old: Does age outweigh gender. *Psychology and Aging*, *6*, 19-27. doi:10.1037//0882-7974.6.1.19

Klimstra, T. A., Hale, W. W., III, Raaijmakers, A. W., Branje, S. J. T., & Meeus, W. H. J. (2009). Maturation of Personality in Adolescence. *Journal of Personality and Social Psychology*, *96*, 898-912. doi:10.1037/a0014746

Lehmann, R., Denissen, J. J. A., Allemand, M., & Penke, L. (2013). Age and gender differences in motivational manifestations of the Big Five from age 16 to 60. *Developmental Psychology*, *49*, 365-383. doi:10.1037/a0028277

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. doi:10.1007/s10508-008-9380-7

Löckenhoff, C. E., De Fruyt, F., Terracciano, A., McCrae, R. R., De Bolle, M., Costa, P. T., . . . Yik, M. (2009). Perceptions of aging across 26 cultures and their culture-level associates. *Psychology and Aging*, *24*, 941-954. doi:10.1037/a0016901

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *Journal of Social Psychology*, *137*, 369-373.

Maccoby, E. E., & Jacklin, C. N. (1974). Myth, reality and shades of gray: What we know and don't know about sex differences. *Psychology Today*, *8*(7), 109-112.

Marcus, D. K., & Lehman, S. J. (2002). Are there sex differences in interpersonal perception at zero acquaintance? A social relations analysis. *Journal of Research in Personality*, *36*, 190-207. doi:10.1006/jrpe.2001.2346

McCrae, R. R., & Costa, P. T. Jr. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 159-181). New York, NY: Guilford.

McCrae, R. R., Costa, P. T. Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*, 261-270. doi:10.1207/s15327752jpa8403_05

McCrae, R. R., Costa, P. T. Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology*, *83*, 1456-1468. doi:10.1037//0022-3514.83.6.1456

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175-215.

McCrae, R. R., Terracciano, A., & 78 Members of the Personal Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*, 547-561. doi:10.1037/0022-3514.88.3.547

Miettunen, J., Veijola, J., Lauronen, E., Kantojarvi, L., & Joukamaa, M. (2007). Sex differences in Cloninger's temperament dimensions— a meta-analysis. *Comprehensive Psychiatry*, *48*, 161-169. doi:10.1016/j.comppsych.2006.10.007

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401-421.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313-345. doi:10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*, 1-25. doi:10.1037/0033-2909.132.1.1

Ruble, D. N., & Martin, C. L. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology*, Vol. 3 (pp. 933-1016). New York, NY: John Wiley.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3, Social, emotional, and personality development* (6th ed., pp. 858-932). Hoboken, NJ: John Wiley.

Ryan, C. S. (2002). Stereotype accuracy. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 13, pp. 75-109). Hove, UK: Psychology Press/Taylor & Francis.

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. doi:10.1037/a0014651

Schwartz, S. H. (2005). *Schwartz Value Survey*. Jerusalem: Israel Social Sciences Data Center.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*, 330-348. doi:10.1037/a0021717

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. Austin: University of Texas Press.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*, 42-61. doi:10.1037/0033-2909.111.1.42

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21-36. doi:10.1037/0022-3514.66.1.21

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, L., Ahn, C., Ahn, H. N., . . . R. R. McCrae. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science*, *310*, 96-100. doi:10.1126/science.1117199

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. (2005). Hierarchical linear modeling analyses of the NEO-PI-R Scales in the Baltimore longitudinal study of aging. *Psychology and Aging*, *20*, 493-506.

Vecchione, M., Alessandri, G., Barbaranelli, C., & Caprara, G. (2012). Gender differences in the Big Five personality development: A longitudinal investigation from late adolescence to emerging adulthood. *Personality and Individual Differences*, *53*, 740-746. doi:10.1016/j.paid.2012.05.033

Wiggins, J. S., & Broughton, R. (1985). The interpersonal circle: A structural model for the intergration of personality research. In R. Hogan & H. Jones (Eds.), *1985* (pp. 1-47). Greenwich, CT: JAI press.

Williams, J. E., Satterwhite, R. C., & Best, D. L. (1999). Pancultural gender stereotypes revisited: The Five Factor Model. *Sex Roles*, *40*, 513-525. doi:10.1023/a:1018831928829