



Sex differences in personality scores on six scales: Many significant, but mostly small, differences

Adrian Furnham¹ · Luke Treglown²

Accepted: 24 March 2021
© The Author(s) 2021

Abstract

This study examined sex differences in domain *and* facet scores from six personality tests in various large adult samples. The aim was to document differences in large adult groups which might contribute new data to this highly contentious area. We reported on sex differences on the Myers-Briggs Type Indicator (MBTI); the Five Factor NEO-PI-R; the Hogan Personality Indicator (HPI); the Motives and Values Preferences Indicator (MVPI); the Hogan Development Survey (HDS) and the High Potential Trait Indicator (HPTI). Using multivariate ANOVAs we found that whilst there were many significant differences on these scores, which replicated other studies, the Cohen's *d* statistic showed very few (3 out of 130) differences $>.50$. Results from each test were compared and contrasted, particularly where they are measuring the same trait construct. Implications and limitations for researchers interested in assessment and selection are discussed.

Keywords Sex · Personality · Traits · Effect size · Bright/Dark side

Introduction

It has been observed that researchers have to be courageous to investigate or write about sex differences (Furnham, 2017). Even the terminology is a sensitive issue: the terms “man” and “woman” are typically used in reference to gender, whereas the terms “male” and “female” are used in reference to sex. In this paper we shall examine sex differences and refer to male and female.

What is most surprising in this complex research area is comparing the radically different conclusions of researchers and reviewers on exactly the same topic. Early intelligence researchers put in considerable effort to ensure tests showed *minimal* sex differences (Mackintosh, 1986; Mackintosh, 2011), though personality researchers seemed less concerned with evidence of sex differences. However, over the course of the last 20 years there have been a great number of studies concerned with gender differences in personality world-wide (Del Giudice, 2009; Del Giudice et al., 2012; Schmitt, 2015; Schmitt et al., 2008; Weisberg et al., 2011). Some have

concentrated on particular group differences using clinical populations, different age groups, and different cultures, or whether scores change much over time (Furnham & Cheng, 2019). Few have been interested in devising valid tests that minimise sex differences, but rather trying to establish the size, and more importantly the cause, of the differences they find currently exist.

There have also been meta-analysis in the area, some done many years ago. Thus, using now less known and used tests of an earlier era, Feingold (1994) concluded that males were more assertive, and had higher self-esteem than females, who scored higher than males in extraversion, anxiety, trust, and, especially, tender-mindedness. There were no sex differences in social anxiety, impulsiveness, activity, locus of control, and orderliness.

Some argue that even if they are small, actual, verifiable (not artefactual) sex differences, they should not be explored or explained because of the divisive personal and social effect that it can have on both sexes. Others believe there are important explicable reasons for sex differences which warrant scientific description and explanation (Buss, 1995; Eagly, 1995; Furnham, 2017).

Another curiosity in this highly disputed area is the apparent contradiction between *popular and scientific writers* (Gray, 1992; Pease & Pease, 2002). There are many popular books that portray a simple evolutionary perspective that describes, and even rejoices in, sex differences in almost all

✉ Adrian Furnham
adrian@adrianfurnham.com

¹ Norwegian Business School (BI), Nydalveien, Oslo, Norway

² Thomas International, London, UK

human behaviour, but particularly communication, relationships and work. These are contrasted with the measured and cautious academic books and papers that note how complex some of these seemingly simple questions are, and how all the answers require numerous qualifiers (Halpern et al., 2007).

Inevitably there are two strongly competing, opposite forces: those who stress the *biology* of difference and those who stress the *sociology* of similarity. The former often suggest that these differences are immutable, though it is accepted that all innate traits can be changed with experience. Whilst nearly everyone acknowledges that we are biopsychosocial beings there are those who see us more as BIOpsychosocial as opposed to biopsychosocial. This all concerns explaining how and why observed differences occur (Furnham, 2017).

At the heart of the issue is the quality and quantity of sex differences, their cause and consequence. Though the focus has always been on differences, the trend has been to talk of similarities which is what a great deal of the literature suggests. Indeed, it has been argued that the word difference is too easily confused with deficiency. The same is true of the words sex and gender: the former applying to biological distinctions and the latter sociological categories (Furnham, 2017).

As a result, there is a sophisticated and subtle nature-nurture debate that persists across many interrelated disciplines whose practitioners' study human behaviour (Furnham & Kanazawa, 2020). Many academics who view gender as the product of socialization and cultural factors are split into opposing camps based on whether there are large or insignificant differences between women and men/ males and females (discussed by Buss & Schmitt, 2011).

There is an extensive and growing literature on how evolution creates systematic variation in personality (Nettle, 2006; Penke et al., 2007). Scholars attempt to explain how culture, biology, and evolution interact to collectively shape personality (Fischer, 2018). Evolutionary psychology posits various sources of sex differences, such as sexual selection (intersexual selection and intrasexual competition) and the theory of obligatory parental investment (Archer, 1996, 2009; Buss, 1995; Geary, 2010). Moreover evolutionary psychologists attempt to describe and explain how evolutionary processes shape sex differences in personality and the specific reasons as to why we might expect, or not expect, to see sex differences specific personality traits (Del Giudice et al., 2012; Lippa, 2010; Schmitt et al., 2008). There is also a salient literature on the proposed cultural origins of gender, more particularly the purported sociocultural factors that shape gender symmetry (Hyde, 2007) or differences (Eagly & Wood, 1999).

Behaviour geneticists decompose total variance in personality and other individual traits into three components: heritability (genes), shared environment (everything that happens within the family that makes siblings from one family similar

to each other but different from those from other families), and unshared environment (everything that happens within and outside the family that makes siblings from one family different from each other) (Plomin et al., 2012). Behaviour geneticists contend that the rough rule of thumb when it comes to the determinants of adult personality and other traits is 50–0–50, that is, roughly 50% of the variance in personality, behaviour, and other traits is heritable (influenced by genes), roughly 0% by the shared environment (what happens within the family and is experienced similarly by all siblings), and roughly 50% by the non-shared environment (what happens inside and outside the family not shared by siblings). Tooby and Cosmides' (2005) talk of the standard social science model (SSSM) of the brain. Adherents of the SSSM argue that the brain is a general-purpose device that is almost entirely shaped by culture and that individual differences are explained by social environment and learning (Vrabel & Zeigler-Hill, 2017). Tooby and Cosmides (2005) purport that their integrated model is superior to the SSSM because it integrates both culture and evolutionary biology. However, among evolutionarily-minded scholars, some believe that this distinction represents a false dichotomy (Richardson, 2007; Wallace, 2010). Suffice it to say this remains a highly contentious academic area of research.

Reviewers of this topic can be described as Maximizers vs Minimizers. Maximizers want to find and explain the (many large) differences between the sexes while the minimisers want to emphasize how few real and meaningful differences there are (Furnham, 2017). Part of this debate can be seen in the interpretation of Cohen's *d*, which is an indicator of difference. Whilst there are conventions about how to label the difference as: none, trivial, small, medium, large and very large, this is also contested. Most researchers quote Cohen who suggested that $d = 0.2$ be considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size. This means that if two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant, although these cut off points have been disputed. However there is an interesting literature which suggests *d* differs according to a number of factors (sub-discipline, sample size) and that in some areas of research a *d* of .25 to .35 could be considered medium (Hemphill, 2003; Greenwald et al., 2015; Schäfer & Schwarz, 2019).

To what extent do these results matter? In the applied context, it might be useful to contrast the approach taken with IQ tests, where they were originally designed to eliminate, as much as possible, sex differences. This does not seem to have been done by personality test creators. Most researchers do not worry about sex differences on IQ tests, because in essence there are none, but perhaps we should worry about personality tests in selection contexts. This could clearly have an impact in selection. It is particularly interesting when different personality tests

essentially measure the same trait but yield large differences.

This Study

This study is on sex differences in personality. In this study we report data on six questionnaires, four of which are well known and have overlapping dimensions (like Conscientiousness). We were fortunate to have large data sets on each of these, though we are aware that there are other important and well-used personality tests some of which measure other dimensions (e.g. HEXACO). We also aware that one test we report on, namely the MBTI, has been heavily criticized academically, though still very frequently used in applied and consulting settings and thus we examine it along with the others (Barbuto Jr, 1997; Furnham, 2018). Also, we also examine the MVPI (see below) which strictly speaking measures motives and values rather than traits, but yields some interesting and important results.

We believe this study has various unique features. *First*, while it replicates many other studies, it does so in overpowered samples often comprising many thousands of adults. *Second*, it examines the differences in six different well-known tests, whereas previous studies nearly all examined only one test. This allows the possibility of looking at differences between tests that measure the same construct (e.g. NEO Neuroticism, HPI Adjustment; HPTI Adjustment). *Third*, for two of the well know measures (NEO; HPI) we were able to examine differences at both domain *and* facet level. *Fourth*, for two questionnaires there were two large samples so that it was possible to examine replications. In all studies the respondents were first language English-speaking adults.

Usually test manuals provide information on group differences such as ethnicity and gender. Sometimes this data is very out-of-date and restricted to one continent. Surprisingly, the N is also often modest. Moreover, it seems to be the case that test publishers are eager to show as few group differences as possible as this may influence potential buyers of the test (Furnham, 2018). For each of the tests used in this study the manuals were consulted to examine the data on sex differences. Each provide good evidence of the internal and test-retest reliability of the test scores. This led to the development of the hypotheses, though the major concern was in the size of the differences.

Based on an observation of various previous studies in personality and evolutionary psychology and the test manuals, the following hypotheses were derived (Del Giudice, 2009; Del Giudice et al., 2012; Furnham, 2008; Schmitt et al., 2017):

1. MBTI: Males would score significantly higher than females on Thinking vs Feeling (H1) and Judging vs Perceiving (H2) (Furnham, 1996, 2018).
2. NEO-PI-R: Males would score significantly lower than females on Neuroticism (H3) and Agreeableness (H4) (Costa Jr. et al., 2001).
3. HPI: Males would score significantly higher than females on Adjustment (H5) and Ambition (H6) but lower on Interpersonal Sensitivity (H7) (Hogan et al., 2007).
4. MVPI: Males would score significantly higher than females on Commerce (H8), Power (H9) and Science, but lower than females on Aesthetics (H10) and Altruism (H11). (Hogan et al., 2007).
5. HDS: Males would score significantly higher than females on Bold (H12) and Mischievous (H13) but lower on Cautious (H14), Dutiful (H15) and Excitable (H16). (Hogan et al., 2007).
6. HPTI: Males would score higher on Adjustment (H17), Risk Approach (H18) and Competitiveness (H19) (MacRae & Furnham, 2020).

Method

Participants

There were seven different samples, most over 1000 participants, used in this study. The focus was on sex differences and these are shown in each table. People ranged in age from 24 to 69 years with the majority being in their late thirties. For each questionnaire there was no significant sex difference in age between males and females. In most of the samples (over 50%) were graduates and once again it was established that there was no difference in the education level between males and females. Most were working adults in supervisory and management positions from a very wide range of organisations. We did not have data on the participants socio-economic status or their work history. Because the participants were nearly all at middle manager levels in their organisations there was a bias towards more males often being twice as many as females (see study limitations). Participants self-identified as either male or female: there was very little missing data for this question.

Instruments

1. *The MBTI Myers-Briggs Type Indicator-Form G* (MBTI: Myers & McCauley, 1985). This is a Jungian-based inventory that is composed of 94 forced-choice items that yield scores on each of the eight factors as well as the famous four dimensions: Introversiion-Extraversion, Sensation-Intuition, Thinking-Feeling and Judging-

Perceiving. Respondents are classified into one of 16 personality types based on the largest score obtained for each bipolar scale. The test provides linear scores on each dimension which are usually discussed in terms of types based on cut-off scores. The Myers-Briggs Type Indicator has been the focus of extensive research and substantial evidence has accumulated suggesting the inventory has satisfactory concurrent and predictive validity and reliability.

2. *The NEO Personality Inventory Revised (NEO-PI-R)* (Costa & McCrae, 1992) This questionnaire is a 240-item measure designed to assess the Five Factor Model (FFM) domains (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness) as well as six primary traits/facets for every domain. The test takes approximately 35 min to complete. Research has provided evidence for the validity and the reliability of this instrument (McCrae et al., 2011). In the current study only the five domains and not the traits were taken into consideration.
3. *Hogan Personality Inventory (HPI)* (Hogan et al., 2007). The HPI consists of 206 items that are used to produce seven personality traits and six criterion scores. Participants respond to each question on a five-point Likert scale. The scales' internal consistency and test-retest reliabilities are well established, with both the manual and independent research citing internal consistency alphas of over .71 and test-retest reliability between .74 and .86 (Hogan et al., 2007).
4. *Hogan Development Survey (HDS)* (Hogan et al., 2007). Similar to the HPI, the HDS is a contextualised measure as it seeks to identify dysfunctional behaviours that impair work performance. The HDS taxonomy is closely related to classical personality disorders (PD) described by the DSM-IV-TR (American Psychiatric Association, 2015). The HDS adopts a dimensional model, opposed to categorical. The HDS consists of 154 items that are completed via participants stating either their agreement or disagreement. The manual reporting internal reliabilities ranging between .50 and .79 (average alpha = .67) and test-retest reliabilities between .58 and .87 (average alpha = .75).
5. *The Motives, Values, Preferences Inventory (MVPI)* (Hogan et al., 2007) measures 10 Motives/Preferences. Each scale is composed of five themes: a) Lifestyles, which concern the manner in which a person would like to live; b) Beliefs, which involve 'shoulds', ideals and ultimate life goals; c) Occupational Preferences, which include the work an individual would like to do, what constitutes a good job, and preferred work materials; d) Aversions, which reflect attitudes and behaviours that are either disliked or distressing; and e) preferred Associates,

which include the kind of persons desired as co-workers and friends. MVPI scores are quite stable over time, with test-retest reliabilities ranging between .64 and .88 ($M = .79$). More than 100 validation studies have been conducted on the MVPI with results indicating that the inventory is effective in predicting job performance and outcome variables such as turnover.

6. *High Potential Trait Indicator (HPTI)* (MacRae & Furnham, 2014). The HPTI is a measure of personality traits, specifically within a workplace context. It is comprised of six factors, Adjustment, Curiosity, Ambiguity Acceptance, Conscientiousness, Courage, and Competitiveness. The inventory is 78 items in length. Each trait is converted into a percentile rank based off the normal distribution of the sample. Various papers have been published using this measure (Furnham & Treglown, 2018)

Procedure

Participants were tested by three well established British-based psychological consultancies over a period from 10 to 16 years, where participants attended assessment centres and their data was logged. The same participants tended to complete the MBTI and NEO-PI-R where the data were obtained from one consultancy, the three Hogan Instruments (HPI, HDS, MVPI) where data were obtained from the other consultancies, and the HPTI the third consultancy. This was done in either in assessment centres or online as a part of recruitment or development process, and all participants were given full feedback on their test performance. They came from a wide range of organisations in the private and public sector. Participants agreed to take part in research and anonymised data was used in the analysis with their permission. Data sets were given to the authors for analysis with all tests scored which means we could not calculate alphas, though we have no reason to believe there were any problems with them (Hogan et al., 2007). Ethics permission was requested and received (CEHP: 2017; 514).

Results

Data was first screened for random responding, missing data, and other errors. In each analysis we started with MANOVAs and each was significant, followed by one way ANOVAs. Bonferroni corrections ($p < .01$) were made which meant a number of analyses (12 in all) ceased to be significant. In the interpretation we only focused on results where $p < .001$ though our primary focus was on the Cohen's d score/ We assumed that under $d < .20$ was a small difference and $d < .50$ a medium sized difference.

1. MBTI

Table 1 shows that males scored higher than females on all dimensions, particularly Thinking-Feeling where the *d* was in the $.20 > d < .50$ range. Males scored higher in Sensing and higher on Judging which is consistent with the literature. This confirms H1 and H2.

2. NEO-PI-R

Table 2 shows that all big five factors showed significant sex differences. Females scored higher on four of the five traits, particularly Openness and Neuroticism, but lower on Conscientiousness. All but three of the facets revealed significant differences. With few exceptions the facets within each domain showed consistent differences. Exceptions were Assertiveness and Excitement Seeking in Extraversion where males scored higher than females. Of the 30 *d* scores, 17 were $< .20$, 16 were $.20 > d < .50$ and only one $> .50$ (Feelings in the Openness factor). This confirms H3 and H4.

3. HPI

Table 3 shows all seven domain factors were significant. Males scored significantly higher on Adjustment, Ambition, Sociability and Inquisitive, but lower on Interpersonal Sensitivity, Prudence, and Learning Approach. Once again, the facets within each domain score tended to be consistent both in direction and significance. Of the 50 analyses 32 showed *d* scores $< .20$, 17 were $.20 > d < .50$ and one $> .50$ (Curiosity in the Inquisitive factor). This confirms H5, H6 and H7.

4. MVPI

Table 4 shows the results from the two different samples. The results were reasonable consistent. In both samples males scored significantly higher than females on Recognition,

Power, Commerce and Science but lower on Hedonism, Altruism, Affiliation, and Aesthetics. Combining the two in all 8 of the *d*'s were $< .20$, 10 were $.20 > d < .50$ and two $d > .50$ (Commerce in Sample 1 and Science in Sample 2). This confirms H8 to H11.

5. HDS

Table 5 also shows data from two different samples, which were one again were reasonably consistent. In both samples females were significantly higher on Excitable (Borderline), Cautious and Dutiful, while males scored higher on Sceptical, Reserved, Bold, and Mischievous. Combining the two samples on the 22 differences 16 were $d < .20$ and 6 were $.20 > d < .50$. This confirms H12 to H16

6. HPTI

Table 6 shows the results of gender difference tests for the six HPTI traits. Significant differences were noted for all six traits, with males scoring higher on Conscientiousness, Adjustment, Risk Approach, Ambiguity Acceptance, and Competitiveness, whereas female participants scored higher on Curiosity. Effect sizes revealed that only Risk Approach ($d = .32$) and Ambiguity Acceptance ($d = .33$) had small effect sizes, whilst the rest can be regarded as negligible ($d < .20$).

Discussion

The results of this study can be interpreted in various different ways. A sex-difference maximiser would note that a cursory glance at the six tables shows that the vast majority of ANOVAs (over 80%) shows significant sex differences, many at the $p < .001$ illustrating the fundamental point that there are many and important sex differences in personality, using a variety of measures and assessed at both the domain and facet level. On the other hand, the minimiser might take comfort in the effect size

Table 1 Data from the Myer-Briggs Type Indicator: MBTI

	Male (<i>N</i> = 7535)	<i>SD</i>	Female (<i>N</i> =1945)	<i>SD</i>	<i>F</i>	<i>d</i>
Extraversion-Introversion	22.70	14.95	23.36	14.54	3.09	0.04
Sensing-Intuition	22.42	14.68	21.32	14.28	8.76**	0.08
Thinking-Feeling	28.44	15.78	23.46	15.17	156.37***	0.32
Judging-Perceiving	23.64	14.60	22.32	14.30	12.86***	0.09

High Scores indicate the dimension on the right, low on the left

* $p < .05$, ** $p < .01$ *** $p < .001$

Table 2 Data from the NEO-PI

	Male (<i>N</i> =7168)	<i>SD</i>	Female (<i>N</i> =2094)	<i>SD</i>	<i>F</i>	<i>d</i>
Neuroticism	62.38	19.50	69.37	20.40	203.87***	-0.35
Anxiety	11.65	5.11	13.42	5.36	188.57***	-0.34
Angry Hostility	9.78	4.56	10.55	4.62	45.01***	-0.17
Depression	9.08	4.64	10.19	4.99	88.74***	-0.23
Self-Consciousness	11.44	4.23	12.29	4.42	62.92***	-0.20
Impulsiveness	14.31	4.38	15.65	4.40	151.60***	-0.31
Vulnerability	6.17	3.46	7.38	3.60	194.21***	-0.34
Extraversion	127.42	18.80	130.60	17.84	47.52***	0.17
Warmth	23.36	4.04	25.21	3.63	354.99***	0.48
Gregariousness	19.90	4.74	20.74	4.60	51.54***	-0.18
Assertiveness	21.12	4.43	20.38	4.63	44.86***	0.16
Activity	21.69	4.15	22.13	3.99	18.47***	-0.11
Excitement Seeking	19.10	4.47	18.29	4.62	52.64***	0.18
Positive Emotion	22.29	4.66	23.90	4.26	199.68***	-0.36
Openness	119.18	18.59	126.91	17.61	286.73***	-0.43
Fantasy	16.59	4.85	17.96	4.76	130.24***	-0.29
Aesthetics	17.08	6.00	19.22	5.51	213.95***	-0.37
Feelings	21.44	4.23	23.57	3.79	428.49***	-0.53
Actions	19.68	4.21	21.24	3.85	230.41***	-0.39
Ideas	20.75	5.27	20.80	5.21	0.13	-0.01
Values	23.65	3.49	24.11	3.12	30.10***	-0.14
Agreeableness	118.93	15.82	123.90	14.84	164.09***	-0.32
Trust	22.05	4.13	22.40	4.07	11.43***	-0.09
Straightforwardness	18.64	4.52	19.47	4.25	56.19***	-0.19
Altruism	23.61	3.49	24.91	3.27	231.10***	-0.38
Compliance	17.97	4.03	18.59	3.97	38.80***	-0.15
Modesty	17.28	4.65	18.36	4.40	88.56***	-0.24
Tender Mindedness	19.39	3.50	20.12	3.24	72.38***	-0.22
Conscientiousness	134.79	17.40	132.65	17.19	24.73***	0.12
Competence	24.47	3.23	23.97	3.25	38.76***	0.15
Order	18.96	4.48	19.12	4.59	2.23	-0.04
Dutifulness	25.12	3.52	24.60	3.52	35.68***	0.15
Achievement Striving	23.39	4.04	23.17	3.93	4.83*	0.06
Self-Discipline	24.00	4.06	23.87	4.14	1.83	0.03
Deliberation	18.87	4.37	17.93	4.47	73.92***	0.21

* $p < .05$, ** $p < .01$ *** $p < .001$

data (Cohen's d) and note that there are very few large or even medium effect sizes, though this depends on how size is categorised.

Nearly all the hypotheses based on the previous literature were confirmed. Overall, the MBTI showed relatively small differences except in the Thinking-Feeling variable which has been the topic of much debate. It has been suggested (and refuted) that this factor is essentially measuring Neuroticism and hence the higher score for females which is consistent with the previous literature (Furnham, 2018),

The results from the NEO-PI-R confirmed some previous studies which showed males higher only on Conscientiousness but lower on Extraversion, Agreeableness, Openness and Neuroticism. The biggest domain differences were for three traits where females scored higher than males. The most unusual finding was the big difference on Openness (which was also shown in the HPTI trait of curiosity) where there is a limited literature and few speculations on sex differences. The smallest and fewest differences were on Conscientiousness and its facets. The facet analysis

Table 3 Data from the Hogan Personality Inventory (HPI)

	Male (<i>N</i> =7296)	<i>SD</i>	Female (<i>N</i> =3341)	<i>SD</i>	<i>F</i>	<i>d</i>
Adjustment	29.88	5.37	28.15	6.18	217.46***	0.30
Empathy	3.83	1.23	3.81	1.28	0.58	0.02
Not Anxious	3.04	1.16	2.62	1.37	276.01***	0.33
No Guilt	4.79	1.39	4.68	1.45	12.65***	-0.08
Calmness	3.43	0.72	3.02	0.91	615.78***	0.50
Even Tempered	4.22	1.09	4.01	1.22	73.93***	0.18
No Complaints	4.66	0.70	4.33	1.01	379.98***	0.38
Trusting	2.00	0.94	2.04	0.93	5.07*	0.04
Good Attachment	3.93	1.34	3.64	1.49	96.65***	0.20
Ambition	26.24	3.21	25.18	4.06	212.78***	-0.29
Competitive	4.78	0.56	4.66	0.71	77.76***	0.19
Self-Confident	2.78	0.51	2.62	0.68	183.36***	0.27
Accomplished	5.82	0.59	5.72	0.71	53.04***	0.15
Leadership	4.88	1.55	4.61	1.68	67.38***	0.17
Identity	2.70	0.70	2.59	0.83	49.66***	0.14
No Social Anxiety	5.29	1.20	4.98	1.44	128.93***	0.23
Sociability	16.18	4.28	15.37	4.43	79.80***	0.19
Likes Parties	2.96	1.21	3.00	1.14	2.11	-0.03
Likes Crowds	2.80	1.33	2.73	1.36	6.62**	0.05
Experience Seeking	5.12	1.15	4.91	1.28	71.50***	0.17
Exhibitionist	2.68	1.50	2.44	1.51	60.04***	0.16
Entertaining	2.62	1.32	2.30	1.40	128.94***	0.24
Interpersonal Sensitivity	19.24	2.19	19.86	1.93	199.88***	0.30
Easy to Live With	4.65	0.66	4.66	0.63	0.86	0.02
Sensitive	3.51	0.72	3.73	0.56	243.93***	-0.34
Caring	3.71	0.55	3.81	0.47	79.70***	-0.20
Likes People	5.50	0.88	5.60	0.75	33.67***	-0.12
No Hostility	1.89	0.79	2.07	0.84	117.45***	-0.22
Prudence	20.93	4.36	21.18	4.09	7.62**	0.06
Moralistic	3.07	1.18	3.11	1.20	3.48	-0.03
Mastery	3.21	0.88	3.34	0.81	54.98***	-0.15
Virtuous	3.79	1.02	3.83	1.01	3.50	-0.04
Not Autonomous	2.03	1.07	2.10	1.05	10.50***	-0.07
Not Spontaneous	2.67	0.95	2.58	0.95	20.36***	0.09
Impulse Control	2.45	1.48	2.47	1.45	0.35	-0.01
Avoids Trouble	3.72	1.22	3.75	1.14	1.05	0.03
Inquisitive	16.63	4.26	14.76	4.77	413.96***	0.41
Science Ability	3.41	1.36	2.73	1.51	536.64***	0.47
Curiosity	2.60	0.72	2.16	0.89	720.95***	0.54
Thrill Seeking	2.81	1.57	2.19	1.55	364.92***	0.40
Intellectual Games	1.99	0.95	2.06	0.96	11.25***	-0.07
Generates Ideas	3.92	1.16	3.53	1.36	233.20***	0.15
Culture	1.90	1.25	2.08	1.32	46.22***	-0.14
Learning Approach	9.21	3.03	9.31	3.01	2.67	0.03
Education	2.28	0.92	2.43	0.85	64.54***	-0.17
Math Ability	1.96	1.13	1.44	1.19	463.64***	0.45
Good Memory	3.16	1.06	3.07	1.10	15.97***	0.08
Reading	1.82	1.32	2.38	1.33	409.35***	-0.42

Table 3 (continued)

	Male (<i>N</i> = 729- 6)	<i>SD</i>	Female (<i>N</i> = 334- 1)	<i>SD</i>	<i>F</i>	<i>d</i>
Self-Focus	2.21	1.28	2.33	1.24	19.70***	-0.10
Impression Management	2.25	1.27	2.26	1.27	0.17	-0.01
Appearance	2.55	1.48	2.53	1.45	0.37	0.01

* $p < .05$, ** $p < .01$ *** $p < .001$

gave some indication of variability within domain but few where the differences went in the opposite direction. Two exceptions were the facets of assertiveness and excitement seeking in Extraversion where, as in many other studies males scored higher than females. Interestingly the highest *d* was for the Openness facet Feelings ($d = .53$) which reflects the finding in the MBTI. (Furnham, 1996).

The results of the HPI confirm previous studies with the biggest domain *d*'s being for Adjustment, Ambition and Curiosity with males scoring higher and Interpersonal Sensitivity with females scoring higher. Again, most of the facets scores went in the same direction though they did occasionally differ greatly in size: compare empathy and calmness in Adjustment.

The results of the replicated MVPI study showed two things: where there were significant differences the results went in the same direction, and that the biggest differences lay in male's interest in power, business and science, values associated with entrepreneurship and work success (Furnham, 2018). Further, as in previous studies females scored higher in Altruism and Aesthetics.

The findings from the HDS show similar outcome in the two studies. When grouping the eleven traits into the recommended tri-partite system the results are clear: females tend to have scores on those traits moving away from (Cautious but not Reserved) and toward others (Dutiful not Diligent) while males score higher on traits in the moving against others category (especially Mischievous).

Table 4 Data from the Motives and Values Preference Inventory (MVPI)

	Male (<i>N</i> = 874)	<i>SD</i>	Female (<i>N</i> = 584)	<i>SD</i>	<i>F</i>	<i>d</i>
Recognition	40.29	7.35	40.06	7.20	0.34	0.03
Power	47.16	7.00	43.96	7.44	69.34***	-0.44
Hedonistic	40.89	6.67	42.63	6.60	24.17***	0.26
Altruistic	45.23	7.22	48.15	6.11	64.71***	0.44
Affiliation	49.29	5.46	50.19	5.32	9.79**	0.17
Tradition	42.40	5.74	42.63	5.45	0.60	0.01
Security	36.65	7.72	37.09	7.60	1.13	0.06
Commerce/Business	44.79	6.65	41.55	6.78	81.77***	-0.48
Aesthetics/Culture	32.28	7.62	33.92	7.85	15.88***	-0.21
Science/Rationality	40.72	8.00	36.41	7.80	103.52***	0.55
Sample 2	Male (<i>N</i> = 2604)	<i>SD</i>	Female (<i>N</i> = 1427)	<i>SD</i>	<i>F</i>	<i>d</i>
Recognition	41.37	7.45	40.26	6.84	21.90***	0.16
Power	48.24	6.38	46.22	6.64	90.54***	-0.31
Hedonistic	40.32	6.58	41.90	6.57	53.05***	0.24
Altruistic	46.40	7.01	48.14	6.48	59.43***	0.26
Affiliation	49.91	5.27	50.69	4.80	21.17***	0.15
Tradition	42.58	5.94	42.50	5.72	0.21	0.01
Security	38.01	7.31	38.65	7.12	7.05**	0.09
Commerce/Business	45.27	6.13	41.93	6.31	268.10***	-0.54
Aesthetics/Culture	33.44	7.66	35.44	8.16	60.29***	-0.25
Science/Rationality	41.55	7.91	38.34	7.44	158.78***	0.42

* $p < .05$, ** $p < .01$ *** $p < .001$

Table 5 Data from the Hogan Dark Side Inventory (HDS)

Sample 1	Male (N= 3715)	SD	Female (N=1588)	SD	F	d
Excitable	2.87	2.57	3.09	2.77	8.22**	-0.08
Skeptical	4.96	2.49	4.75	2.37	8.35**	0.09
Cautious	2.99	2.48	3.83	2.76	118.91***	-0.32
Reserved	4.23	2.17	3.77	1.99	53.54***	0.22
Leisurely	4.89	2.36	4.82	2.30	0.90	0.03
Bold	7.52	2.73	7.17	2.68	18.93***	0.13
Mischievous	6.93	2.46	6.35	2.39	62.64***	0.24
Colourful	7.69	2.95	7.37	2.87	12.70***	0.11
Imaginative	5.59	2.38	5.39	2.50	7.90**	-0.08
Diligent	9.05	2.43	9.15	2.41	1.94	0.04
Dutiful	7.26	2.20	7.51	2.22	17.33***	-0.11
Sample 2	Male (N= 2828)	SD	Female (N=2115)	SD	F	d
Excitable	2.82	2.57	3.08	2.74	11.59***	-0.10
Skeptical	4.45	2.42	4.39	2.34	0.85	0.03
Cautious	3.07	2.53	3.90	2.79	118.74***	-0.31
Reserved	4.36	2.21	3.64	1.85	146.10***	0.35
Leisurely	5.08	2.32	5.05	2.20	0.21	0.01
Bold	7.32	2.77	6.93	2.65	24.70***	0.14
Mischievous	7.00	2.55	6.35	2.39	82.45***	0.26
Colourful	7.81	3.00	7.67	2.93	2.75	0.05
Imaginative	5.51	2.37	5.41	2.39	2.28	-0.04
Diligent	9.16	2.59	9.20	2.53	0.44	0.02
Dutiful	7.28	2.21	7.63	2.27	29.06***	-0.16

* $p < .05$, ** $p < .01$ *** $p < .001$

The final scale showed two of the six HPTI scales with relatively large differences: males score higher in Risk Approach and Ambiguity Acceptance which has been shown many times before. Although there was a sex difference on Competitiveness, the size of this was modest.

One interesting comparison could be between the scores of different tests which essentially (claimed to) measure the same construct. Thus, the sex difference d for Neuroticism in the NEO-PI-R was .35, Adjustment in the HPI was .30 and Adjustment in the HPTI was .14. Similarly, Conscientiousness in the NEO-PI-R was .12 and in the

HPTI was .11, and Prudence .06. Equally the sex difference d in Agreeableness in the NEO-PI-R was .32 and Interpersonal Sensitivity in the HPI was .30. Therefore, the results seem to suggest similar sex differences on scales of different length and question measuring the same phenomenon. There were however exceptions: females were more Extraverted and Open on the NEO=PI-R, but less Sociable and Curious on the HPI.

One interesting issue concerns revisiting each question and facet to determine whether there was any inherent sex bias in the question wording and whether if these were removed the overall d would decline. This is not an issue of attempted to deny or reduce differences that exist but rather trying to reduce artefacts arising from question selection. Certainly, with changes in society, particularly with reference to sex and gender differences, questionnaire wording could cause both offense and differences in interpretation unless they are constantly updated.

Another issue to arise from this study is the great variability in the facet score items and labels that are essentially measuring the same dimension. Compare for instance the six Openness facets of the NEO-PI-R with six facets of the HPI. Given these labels it is expected that these two measures are relatively weakly correlated and measuring rather different factors.

Finally accepting that there are some real, biologically based, stable sex differences, as opposed to socialised gender differences, in personality traits the question arises as to why they occur. Results such as these cannot inform the nature-nurture debate, with (most) evolutionary psychologists offering a cohesive (and for some convincing) argument as to why there are replicable, consistent and cross-cultural findings. Minimizers who reject the “biology as destiny” approach attempt to explain all these differences in terms of primary and secondary socialisation (Buss, 1995). However, in a big review study Schmitt et al. (2017) concluded: “*Social role theory appears inadequate for explaining some of the observed cultural variations in men’s and women’s personalities. Evolutionary theories regarding ecologically-evoked gender differences are described that may prove more useful in explaining global variation in human personality*” (p45).

Table 6 ANOVAs and effect sizes of gender differences in the six HPTI traits

	Male (N= 4023)	SD	Female (N=5572)	SD	F	d
Conscientiousness	76.13	8.33	75.22	7.76	30.3***	.11
Adjustment	71.35	10.35	69.90	10.02	47.4***	.14
Curiosity	69.31	8.17	70.06	8.07	19.9***	-.09
Risk Approach	70.45	8.54	67.65	8.73	244.1***	.32
Ambiguity Accep	55.46	9.33	52.41	9.22	252.2***	.33
Competitiveness	55.95	10.68	55.04	11.09	16.4***	.08

*** $p < .001$

This study, like all, others has limitations. All participants were British adults taking part in a compulsory assessment centre. Though they might have been tempted by impression management there is no reason to suspect that there were sex differences in this behaviour. The reason why males outnumber females tended to reflect the profile of middle managers in those organisations which reflected all sectors, public and private. The sample was thus biased in terms of age, education and class and the question remains whether a more representative sample of people from a wider age range and social class background would have shown more or fewer sex differences. Furthermore, nearly all participants were from Europe and the effects of culture were thus not explored. It could be that sex differences are smaller in more Western, individualistic, democratic, egalitarian, and higher gender-parity cultural contexts than those from more traditional, developing countries.

It has been argued that personality changes over time and it may be that sex differences and similarities in personality are different for young, middle-aged and older participants (Roberts et al., 2001). Finally there is always the possibility that there are sex differences in self-report behaviours and biases, such that females exhibit more humility and males more hubris and that therefore some observed differences are more due to other factors and artefacts than actual personality differences.

Funding Open access funding provided by Norwegian Business School.

Declarations Committee approved (CEHP/2017/514).

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- American Psychological Association (2015). Definitions related to sexual orientation and gender diversity in APA documents. <https://www.apa.org/pi/lgbt/resources/sexuality-definitions.pdf>
- Archer, J. (1996). Sex differences in social behavior: Are the social role and evolutionary explanations compatible? *American Psychologist*, 51(9), 909–917. <https://doi.org/10.1037/0003-066X.51.9.909>.
- Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences*, 32(3–4), 249–266. <https://doi.org/10.1017/S0140525X09990951>.
- Barbuto Jr., J. E. (1997). A critique of the Myers-Briggs type Indicator and its operationalization of Carl Jung's psychological types. *Psychological Reports*, 80(2), 611–625. <https://doi.org/10.2466/pr0.1997.80.2.611>.
- Buss, D. M. (1995). Psychological sex differences: Origins through sexual selection. *American Psychologist*, 50(3), 164–168. <https://doi.org/10.1037/0003-066X.50.3.164>.
- Buss, D. M., & Schmitt, D. P. (2011). Evolutionary psychology and feminism. *Sex Roles*, 64(9–10), 768–787. <https://doi.org/10.1007/s11199-011-9987-3>.
- Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, 6(4), 343–359. <https://doi.org/10.1521/pedi.1992.6.4.343>.
- Costa Jr., 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. <https://doi.org/10.1037/0022-3514.81.2.322>.
- Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology*, 7(2), 264–279. <https://doi.org/10.1177/147470490900700209>.
- Del Giudice, M., Booth, T., & Irwing, I. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS One*, 7(1), e29265. <https://doi.org/10.1371/journal.pone.0029265>.
- Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, 50(3), 145–158. <https://doi.org/10.1037/0003-066X.50.3.145>.
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, 54(6), 408–423. <https://doi.org/10.1037/0003-066X.54.6.408>.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116, 429–456. <https://doi.org/10.1037/0033-2909.116.3.429>.
- Fischer, R. (2018). *Culture and psychology*. Cambridge University Press. <https://doi.org/10.1017/9781316091944>.
- Furnham, A. (1996). The big five versus the big four: The relationship between the Myers-Briggs type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, 21(2), 303–307. [https://doi.org/10.1016/0191-8869\(96\)00033-5](https://doi.org/10.1016/0191-8869(96)00033-5).
- Furnham, A. (2008). *Personality and intelligence at work*. Routledge. <https://doi.org/10.4324/9780203938911>.
- Furnham, A. (2017). Biological sex and cognitive development. In K. Nadal (Ed.), *The SAGE encyclopedia of psychology and gender* (pp. 150–152). SAGE publications, Inc. <https://doi.org/10.4135/9781483384269.n55>.
- Furnham, A. (2018). The great divide: Academic versus practitioner criteria for psychometric test choice. *Journal of Personality Assessment*, 100(5), 498–506. <https://doi.org/10.1080/00223891.2018.1488134>.
- Furnham, A., & Cheng, H. (2019). The change and stability of NEO scores over six-years: A British study and a short review. *Personality and Individual Differences*, 144(1), 105–110. <https://doi.org/10.1016/j.paid.2019.02.038>.
- Furnham, A., & Kanazawa, S. (2020). The evolution of personality. In L. Workman, W. Reader, & J. H. Barkow (Eds.), *The Cambridge handbook of evolutionary perspectives on human behavior* (pp. 462–470). Cambridge University Press. <https://doi.org/10.1017/9781108131797.038>.
- Furnham, A., & Treglown, L. (2018). High potential personality and intelligence. *Personality and Individual Differences*, 128(1), 81–87. <https://doi.org/10.1016/j.paid.2018.02.025>.

- Geary, D. C. (2010). *Male, Female: The Evolution of Human Sex Differences* (2nd ed.). American Psychological Association. <https://doi.org/10.1037/12072-000>.
- Gray, J. (1992). *Men are from Mars, women are from Venus*. Harper Collins.
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology, 108*(4), 553–561. <https://doi.org/10.1037/pspa0000016>.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., et al. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*(1), 1–51. <https://doi.org/10.1111/j.1529-1006.2007.00032.X>.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*(1), 78–79. <https://doi.org/10.1037/0003-066X.58.1.78>.
- Hogan, R., Hogan, J., & Warrenfeltz, R. (2007). *The Hogan guide: Interpretation and use of Hogan Inventories*. Hogan Press.
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science, 16*(5), 259–263. <https://doi.org/10.1111/j.1467-8721.2007.00516.X>.
- Lippa, R. A. (2010). Gender differences in personality and interests: When, where, and why? *Social and Personality Psychology Compass, 4*(11), 1098–1110. <https://doi.org/10.1111/j.1751-9004.2010.00320.x>.
- Mackintosh, N. (1986). The biology of intelligence? *British Journal of Psychology, 77*(1), 1–18. <https://doi.org/10.1111/j.2044-8295.1986.tb01977.x>.
- Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford University Press ISBN 978-0-19-958559-5.
- MacRae, I., & Furnham, A. (2014). *High potential: How to spot, manage and develop talented people at work*. Bloomsbury.
- MacRae, I., & Furnham, A. (2020). A psychometric analysis of the high potential trait inventory. *Psychology, 11*(8), 1125–1140. <https://doi.org/10.4236/psych.2020.118074>.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28–50. <https://doi.org/10.1177/1088868310366253>.
- Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs type Indicator*. Consulting Psychologists Press.
- Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist, 61*(6), 622–631. <https://doi.org/10.1037/0003-066x.61.6.622>.
- Pease, A., & Pease, B. (2002). *Why men lie and women cry*. Hatchett.
- Penke, L., Denissen, J. J., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality Psychology, 21*(5), 549–587. <https://doi.org/10.1002/per.629>.
- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2012). *Behavior Genetics* (6th ed.). Worth.
- Richardson, R. C. (2007). *Evolutionary psychology as maladapted psychology*. MIT Press. <https://doi.org/10.7551/mitpress/7464.001.0001>.
- Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology, 81*(4), 670–683. <https://doi.org/10.1037/0022-3514.81.4.670>.
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813. <https://doi.org/10.3389/fpsyg.2019.00813>.
- Schmitt, D. P. (2015). The evolution of culturally-variable sex differences: Men and women are not always different, but when they are... it appears not to result from patriarchy or sex role socialization. In V. A. Weekes-Shackelford & T. K. Shackelford (Eds.), *The evolution of sexuality* (pp. 221–256). Springer. https://doi.org/10.1007/978-3-319-09384-0_11.
- Schmitt, D. P., Long, A. E., McPhearson, A., O'Brien, K., Remmert, B., & Shah, S. (2017). Personality and gender differences in global perspective. *International Journal of Psychology, 52*, 45–56. <https://doi.org/10.1002/ijop.12265>.
- 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*(1), 168–182. <https://doi.org/10.1037/0022-3514.94.1.168>.
- Tooby, J., & Cosmides, L. (2005). Conceptual Foundations of Evolutionary Psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 5–67). Hoboken: John Wiley & Sons, Inc.
- Vrabel, J., & Zeigler-Hill, V. (2017). Standard social science model (SSSM) of personality. In: V. Zeigler-Hill, T.K. Shackelford (eds.), *Encyclopedia of personality and individual differences* (pp.1-3). https://doi.org/10.1007/978-3-319-28099-8_1188-1
- Wallace, B. (2010). *Getting Darwin wrong: Why evolutionary psychology won't work*. Imprint Academic.
- Weisberg, Y. J., De Young, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in Psychology, 2*, 178. <https://doi.org/10.3389/fpsyg.2011.00178>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.