**Sex differences or similarities? Males are not the more variable sex when it comes to personality**

Word count: 6,902

Main text (5,000 max.): 3,937

Methods (3,000 max.): 2,965

**Abstract (160 words)**

The notion that men are more variable than women is embedded in scientific thinking. For personality traits, this greater variability is partly attributed to biology, underpinned by claims of more morphological variation among male than female animals. There is, however, little information for wild animals about sex differences in personality-like behaviours. In a meta-analysis we quantified sex differences in means and variances for five traits: boldness, aggression, activity, sociality and exploration (2167 effects, 203 studies, 226 species, five taxa). We also tested if sexual size dimorphism, a proxy for sexual selection, explains sex differences in personality. In 25 taxon-trait specific tests there were only two significant sex differences in mean personality, and none for variance. Including size dimorphism explained sex differences in mean personality in two of nine taxon-trait specific tests (aggression and activity in mammals), and in one test for variance (aggression in fish). We find little evidence for widespread sex differences in variability in personality in wild animals.

**Introduction**

Numerous studies have quantified average differences between men and women in traits ranging from height and physiology to, more controversially, behaviours, including intelligence and personality. Far fewer studies have investigated sex differences in humans in variability among individuals (but see Lehre *et al.* 2009). Recently, however, sex differences in behavioural variability, the underlying causes and the possible consequences have been widely discussion in the social sciences (Stuart-Williams & Halsey 2020). For example, sex differences in variability in academic performance (Machin & Pekkarinen 2008) or in the classroom (Baye & Monseur 2016; O’Dea *et al.* 2018; Gray *et al.* 2019) have been linked to reported sex differences in variability in intelligence (Arden & Plomin 2006; Johnson *et al.* 2008), creativity (Ju *et al.* 2015; Karwowski *et al.* 2016), aggressiveness (Deary *et al.* 2003), personality (Borkenau *et al.* 2013) and, ultimately, brain structure itself (Wierenga *et al.* 2020). The general conclusion of these studies is that boys or men are more variable in their behaviour than girls or women (the ‘greater male variability hypothesis’), a finding which is sometimes used to partly explain lower female representation in STEM disciplines (Pinker 2002).

Greater variability among men than women in behaviour, personality and cognition is readily attributed to social and cultural factors that differ between the sexes, but also to biological factors (Feingold 1992; Miller & Halpern 2014). Some commentators have, however, argued that the role of biology is underplayed (Stuart-Williams & Halsey 2020). A key line of reasoning is that there is a trend across animals for males to be more variable than females. Although true that greater male variability in animals has been reported for some traits, the strength and generality of this claim is unclear. Specifically, the strongest evidence is for greater variability in male than female morphology, especially for traits under sexual selection, including sexual ornaments, courtship, weaponry and body size (Price *et al.* 1993; Pomiankowski & Moller 1995; Archer & Mehdikhani 2003; Reinhold & Engqvist 2013; Wyman & Rowe 2014). To date, the evidence is weak or absent for greater male variability for behaviours exhibited by both sexes (Tarka *et al.* 2018; Zajitschek *et al.* 2020).

Human personality is often quantified by classifying individuals of both sexes based on their scores for five components of behaviour (the ‘Big Five’): extraversion, neuroticism, openness, conscientiousness and agreeableness (Costa & McCrae 1992). Individuals tend to have repeatable personality scores over time. In the last 15 years, comparable evidence has emerged for consistent behavioural differences in animals that are akin to variation in human personality. These animal behaviours ate often grouped into five personality axes that loosely resemble those in humans: activity, aggression, boldness, exploration and sociability (Réale *et al.* 2007). It is therefore now possible to test for greater male variability in personality traits in animals.

The four main, non-mutually-exclusive explanations invoked by biologists to explain the maintenance of variation in animal personality despite natural selection tending to eliminate less fit variants (Smith & Blumstein 2008; Dingemanse & Wolf 2010) are: (a) sexual selection; (b) negative frequency-dependent selection; (c) life-history trade-offs; and (d) developmental or genetic constraints. Crucially, all four explanations might partially account for why males are more variable than females in personality traits.

First, sexual selection is usually stronger on males than females (Fromhage & Jennions 2016; Janicke & Morrow 2018). It favour individuals with the most extreme expression of traits that increases the likelihood of obtaining mates (e.g. weapons, ornaments and coercive behaviours) (Darwin 1871), or fertilizing eggs (Parker 1970), because success depends upon relative competitive advantage. Competition generates strong directional selection on these traits. Sexual selection also favour condition-dependent trait expression because individuals in better condition can incur greater costs (Rowe and Houle 1996). Even minor differences in resource acquisition among individuals due to chance events or small genetic differences in, say, foraging efficiency therefore translate into differential expression of sexual traits (Rowe and Houle 1996), increasing phenotypic variance among individuals. Stronger sexual selection on males is therefore predicted to result in sexual traits with greater variance among males than the equivalent traits in females, or than naturally selected traits in either sex (Pomiankowski & Møller 1995). There is some support for this prediction. A meta-analysis of phenotypic variance of traits reported significantly higher variation among males than females for reproductive and non-reproductive traits, but a greater difference for reproductive traits (Wyman & Rowe 2014).

Second, negative frequency-dependent selection can maintain variation in traits. This is particularly relevant for behavioural traits that often affect the intensity of competition among similar individuals (Wolf & McNamara 2012). For example, there are two foraging morphs in *Drosophila melanogaster* larvae: ‘rover’ (actively explores) and ‘sitter’ (sedentary feeders). Both morphs have lower fitness when common, as within-morph competition increases with population density at food sources (Fitzpatrick *et al.* 2007). Alternative mating tactics that are under negative frequency-dependent selection are usually associated with a suite of morphological traits that differ from those for the dominant mate acquisition tactic. For example, sneakers are smaller and duller than males using the dominant mating tactic. Sexual selection, which promotes the evolution of alternative male mating tactics, therefore tends to generate higher variation among males than females.

Third, individuals with greater life expectancy value their future reproduction more highly. This can result in a life history trade-off such that these individuals are, for example, less bold or exploratory (Wolf *et al.* 2007). Sexual selection often reduces male life expectancy to below that of females due to injuries during fighting (e.g. Piper *et al.* 2008), the energetic costs of sexual advertising (e.g. Hunt *et al.* 2004), and ornaments attracting predators (e.g. male guppy colouration: Gordon *et al.* 2011). Lower life expectancy, in conjunction with condition-dependent trait expression, could led to greater phenotypic variation among males than females.

Four, greater male variability might arise from genetic and developmental pathways associated with sex determination mechanisms (James 1973). In some taxa females have two copies of each gene, while males have one. In mammals, for example, XX/XY sex determination means that females have two X chromosomes and males a single X chromosome. In females the phenotypic effects of genes on the X chromosome are averaged across their expression on both chromosomes, either through epigenetic inactivation of different regions of maternally and paternally inherited X chromosomes (Amos-Landgraf *et al.* 2006), or by the expression of a single X chromosome per cell (Lyon 1961). In contrast, males only have a maternally inherited X chromosome, so that all genes on it are expressed. All else being equal, this should create more extreme phenotypes in males (review: Charlesworth 1996), hence greater variation among males than females. It should be noted, however, that females are the heterogametic sex in some taxa, including birds, butterflies, and some fish and reptiles (Beukeboom & Perrin 2014). If this mechanism drives sex differences in variability then female heterogametic taxa should exhibit greater trait variability in females (see Reinhold & Engqvist 2013).

Based on the above explanations, sexual selection is likely to amplify variation in sexually selected behaviours and, as a by-product, any associated behaviours that contribute to an individual’s personality. For example, sexual selection on fighting behaviour implies that aggressiveness is an aspect of personality where males will be more variable than females. Indeed, in humans, variation in two components of personality (agreeableness and neuroticism) where low and high scores, respectively, are linked to aggressive behaviour, show greater variability among men than women (Budaev 1999, Archer & Mehdikhani 2003). This line of reasoning should apply to all species, such that male variability in personality increases when sexual selection on males is more intense. This prediction is directly relevant to biological explanation for sex differences in variation in human behaviour, but has yet to be tested.

Here we conduct a large-scale meta-analysis of animal personality studies to test the robustness of broad claims that males show greater variability than females. More specifically, we test three key questions: 1) Do males show greater variability than females in five personality factors? 2) Are sex differences in variability similar for the five components of personality? And 3) Does sexual selection (estimated using the proxy of sexual size dimorphism, SSD) explain differences in the magnitude of sex differences in personality (mean and variance)? In many taxa, the strength of sexual selection on males is correlated with the magnitude of SSD (Reiss 1986; e.g. primates: Cassini 2020; fish: Horne *et al.* 2020). We therefore included SSD as a moderator in our meta-analyses. Additionally, our data contained taxa where males are heterogametic (mammals), homogametic (birds), or a mixture of the two (insects, fish and reptiles/frogs). As such, we partially tested for an effect of sex chromosomes on sex differences in variation in personality by testing for a moderating effect of taxa.

**Results**

*Dataset Summary*

Our final dataset comprised 2,167 effect sizes from five broad taxonomic groups: mammals, birds, fish, invertebrates, and reptiles/amphibians (combined). The number of species (n = 10-106), studies (n = 11-61) and effect sizes (n = 95-674) per taxa are shown in Table 1. Boldness was the most well-studied, and sociality the least-studied, of the five personality types (n=814 and 166 effect sizes, respectively) (Table 2).

*Sex Differences*

Combining all five personality traits, males and females did not differ significantly in their mean personality in any of the five taxa, nor was there a significant sex difference in variability (Table 1 and Supplementary Table S1). The effect size estimates from our basic meta-analytic intercept models all had high heterogeneity (*I*2Total SMD > 0.70; *I*2Total lnCVR > 0.60), while there was moderate-low heterogeneity for reptiles/amphibians (*I*2Total SMD = 0.45; *I*2Total lnCVR = 0.01) and fish (*I*2Total lnCVR = 0.49). Heterogeneity in the sex difference in mean personality mostly came from between-study differences (*I*2StudyID), while phylogenetic relationships and among species differences (*I*2phylo and *I*2species, respectively) explained heterogeneity in the variability of effect sizes for mammals, birds and reptiles/amphibians only (see Supplementary Table S1).

The lack of a sex difference in mean and variability in personality could arise if the direction of any difference in sex-specific values varied across the five personality traits. When the traits were analysed separately we found significant sex differences in mean values in only two cases. Females were significantly more sociable in birds (=-0.68, 95% CIs: -1.16, -0.21; *p*=0.005), but less exploratory in reptiles/amphibians (= 0.25, 95% CIs: 0.05, 0.45; *p*=0.02). While the magnitude of the estimated effect sizes suggest that sex-differences might exist for personality traits in other taxa (e.g., invertebrates and fish) none of the other 23 tests for sex differences in mean personality were statistically significant (Table 2, Figures 1 & 2).

There was only a single significant sex difference in variability when traits were analysed separately. Females were more variable in their aggressive behaviour in fish (=-0.13, 95% CIs: -0.25, -0.01; *p*=0.04). Again, while the magnitude of the estimated effect sizes in other groups were equivalent or even larger for the difference in variance between males and females, none of the 24 other tests were statistically significant (Table 2, Figures 1 & 2 and Supplementary Table S2).

*Sexual Size Dimorphism and Personality Differences between the Sexes*

Sexual size dimorphism (SSD) explained the extent of the sex differences in mean personality in only one of the four taxa tested, but its influence depended on the personality trait (Table 3 and Supplementary Tables S5-8). There was no effect of SSD in fish, invertebrates or birds. In mammals, the SSD of a species predicted sex differences in activity and aggression, but not of boldness or exploration. When the sexes were the same size (SSD=0), there were no differences in aggression between males and females (*β* = -0.09, 95% CIs: -1.29, 1.10; *p*=0.88). However, as sexual size dimorphism became more male-biased (SSD>0) males were significantly more aggressive than females (*β* =-2.16, 95% CIs: -3.99, -0.32; *p*=0.02). And when the sexes were the same size (SSD=0) males were no more active than females (*β* = 0.44, 95% CIs: -1.74, 2.62; *p*=0.69), but as sexual size dimorphism increased (SSD>0) females were significantly more active than males (*β* =-2.16, 95% CIs: -3.99, -0.32; *p*=0.02).

There were no significant relationships between SSD and sex differences in personality trait variability in any of the four taxa in which we could carry out this test (see Table 3 and Supplementary Tables S5-8). When the sexes were the same size, female fish were more variable in their aggressive behaviour than males, but this effect was weak (*β* = -0.12, 95% CIs: -0.23, 0.00; *p*=0.05). There were no sex differences in variability when the sexes were the same size for any other taxonomic group, nor any effect of SSD on the sex difference in variability.

*Publication bias*

Overall, we found little evidence that publication bias affected estimates of sex differences in the mean or variance in personality traits. Out of 10 tests, publication bias was only evident for mean personality in invertebrates (*p* = 0.005; Supplementary Table S14). After accounting for this, males were, on average, bolder and more active than females (Table S14).

**Discussion**

Our results provided little evidence for widespread male-female differences in personality-linked behaviours in animals in any of the five taxa that we examined. This was the case for both sex differences in the mean values of personality traits, and for the level of variation among individuals. A finer-scale analysis did, however, reveal some sex differences when separately investigating each of the five personality types. Although there were male-female differences for certain personality traits in three taxonomic groups, there was no consistent pattern. For example, females were significantly more social in birds but less exploratory in reptiles/amphibians. There were no detectable sex differences for mammals or fish. Crucially, there was no evidence for greater male than female variability in any taxa for any of the five personality factors. Indeed, the only instance in which there was a significant sex difference revealed greater variability among females (aggression in fish). Finally, the magnitude of SSD, which is a proxy for the strength of sexual selection on males, only explained sex differences in mean personality for activity and aggression in mammals. In species where males and females were the same size, males were more active and less aggressive than females, but in species where males are larger than females, males are significantly less active and more aggressive than females. Crucially, the magnitude of SSD did not explain sex differences in variability for any of the five personality types in the four taxa with sufficient data to test for its effect.

Starting with Darwin (1874, p224) it has been repeatedly stated that males vary more in their appearance (i.e. phenotypes) than do females, partly because of the effects of sexual selection. On closer inspection this empirical claim of greater male phenotypic variability has limited empirical support. One general finding from a small-scale, cross-species study is that sexually selected traits in male animals show more variation than naturally selected traits (Pomiankowski & Møller 1995). Given that sexual selection is usually stronger on males this implies that they will exhibit greater phenotypic variation than females when pooled across all traits. However, another cross-species study showed no significant male-female difference in variation for traits broadly associated with reproduction that are expressed in both sexes, including some traits that might be under direct sexual selection. Furthermore, traits not linked to reproduction had only marginally greater variation among males than females (Wyman & Rowe 2014). In another cross-species study, variation in body size was significantly greater in males than females in taxa where males are the heterogametic sex, but the pattern was reversed in taxa where males are the homogametic sex (Reinhold & Engqvist 2013).

These three cross-species studies mainly focussed on morphological traits, but studies of other types of traits have produced similar findings. For example, a meta-analysis of behavioural, physiological and life history traits (e.g. time to maturity) that mediate the link between current and future reproductive effort (i.e. ‘pace-of-life’ traits) reported no significant sex difference in the level of variation among individuals (Tarka *et al.* 2018). There were also no significant male-female differences in variation when the data was partitioned by breeding system, mating system, study environment or trait type (which included the category ‘behaviour’). Recently, another meta-analysis has investigated a vast dataset on sex differences in probably the most heavily studied model laboratory vertebrate species, the house mouse *Mus musculus* (Zajitschek *et al.* 2020). The main finding is clear: across all examined traits there is no sex difference in trait variability. For specific traits types there is, however, a clear bias towards either females (e.g. immunological traits, eye morphology) or males being more variable (e.g. morphological traits).

Given the available empirical data it seems that the ‘greater male variability hypothesis’ is, at best, only weakly supported for morphological and physiological traits in animals. In contrast, in humans and chimpanzees, greater male variability has been shown for a range of morphological and physiological traits including brain structure (Arnold 2004; van der Linden *et al.* 2017; DeCasien *et al.* 2020) and, perhaps more notably, for behavioural traits like personality (Archer & Mehdikhani 2003; Borkenau *et al.* 2013; Karwowski *et al.* 2016), cognitive ability (Halpern & LaMay 2000; Jones *et al.* 2003; Arden & Plomin 2006; Johnson *et al.* 2008, 2009; Roalf *et al.* 2014) and academic achievement (Lehre *et al.* 2009; Baye & Monseur 2016; O’Dea *et al.* 2018). Our current findings are therefore intriguing, because we show using a larger database of 226 species that personality-like behavioural traits are, in general, not more variable in males than females. If anything, the trend is towards greater variation among females (supporting Tarka *et al.* 2018), which has been previously hypothesised to arise, at least in mammals, due to the effect of females being at different stages of their oestrus cycle (Beery & Zucker 2011). Our findings for animals raises doubts about the extent to which biological factors and evolutionary arguments explain why men have greater trait variation than women for behavioural traits.

One widespread biological explanations for greater variation among men than women is attributed to sex chromosomes, and the fact that men are XY and women XX. Men therefore only express genes from a single X chromosome, while women, on average, express genes on both. The net effect is greater variance in gene expression among men, which should tend to increase the level of phenotypic variation (Reinhold & Engqvist 2013a). However, we found no difference in the sex difference in variability between mammals and birds even though males are heterogametic in mammals and homogametic in birds. One explanation is that genes on sex chromosomes do not affect, or only weakly affect, behaviours associated with personality in animals. In mammals, genes responsible for body size, brain gender and brain characteristics are X-linked and frequently escape X chromosome inactivation processes that silence gene expression, leading to variability (Carruth *et al.* 2002; Skuse 2006; Snell & Turner 2018).

Another major biological explanation for greater variation among men than women in behavioural traits is sexual selection. This is especially relevant for personality as certain traits elevate the likelihood of reproducing. For example, more extravert and creative men tend to have more mating opportunities in some societies (Buss 1995; Nettle 2006; Karwowski *et al.* 2016). Similarly, there is evidence that consistent levels of aggression (i.e. consistent, average levels of aggression) elevate success during male-male competition (e.g. killifish: McGhee & Travis 2010; social lizards: McEvoy *et al.* 2013). These behavioural traits are therefore sexually selected. Sexual selection is expected to result in condition-dependent expression of traits which should increase phenotypic variation (Rowe and Houle 1996). Indeed, there is evidence in humans that some traits under sexual selection show a greater sex difference in variability than other traits (Archer & Mehdikhani 2003). It is therefore relevant that we found no moderating effect of sexual size dimorphism, which is a standard proxy for the level of sexual selection on males (e.g. Rohner *et al.* 2016), on the sex difference in variation in personality in any of the animal taxa that we examined, including mammals. One explanation for our finding is that natural selection on females might lead to comparable directional selection. For example, females generally invest more than males in parental care (Janicke *et al.* 2016), and, in this context, there might therefore be equally strong selection on females as males to be aggressive. Similarly, in group living animals, females often establish social hierarchies where dominance is maintained through aggressive interactions (e.g. Kappeler 2017). Another possible explanation for not finding greater male variability in personality in animals is that many of the measures of personality involve behaviours that are likely to be under similar natural selection in both sexes. For example, many animal personality traits affect survival, notably boldness, which is linked to anti-predator responses, and activity or exploration, which is linked to foraging.

Of course, a lack of evidence for greater male variability in animals for personality traits does not preclude biological factors contributing towards greater male variation in a range of behavioural and allied traits in humans (Snell & Turner 2018). Given that phylogeny (*I*2phylo) explained a large proportion of variance in sex-specific differences in personality variability in mammals it would be valuable to conduct a more focussed meta-analysis looking at sex-specific variability in behaviour in primates. Nonetheless, our findings for animals, alongside the weak evidence for greater male variability in other traits (Reinhold & Engqvist 2013; Tarka *et al.* 2018; Zajitschek *et al.* 2020), suggests that accepting evolutionary explanations for greater behavioural variability in men than women is premature. Greater attention needs to be paid to the possible role of social factors that might select for a wider range of developmental pathways in boys than girls yielding greater behavioural variability in men than women (Gray *et al.* 2019).

**Materials and Methods**

*Literature search and data collection*

This meta-analysis was pre-registered with the Open Science Foundation (OSF; study details available at: <https://osf.io/bwjyt/>). We conducted a systematic search of the literature for empirical studies of animals that quantify personality traits, which were categorised into five types: ‘boldness’, ‘aggression’, ‘activity’, ‘sociality’ and ‘exploration’ (Sih *et al.* 2004; Réale *et al.* 2007). We ran primary and secondary keyword searches using the search strings shown in Supplementary Table S25. In the primary searches we looked for behavioural measures commonly used by those studying animal personality to quantify each of the five personality types, such as ‘hiding time’ as an index of boldness or ‘distance moved’ as an index of exploration (see Sih *et al.* 2004; Réale *et al.* 2007). Our secondary searches were designed to locate studies of ‘personality-like’ behaviours that were alternate measures of activity, aggression, boldness, exploration or sociality. We then conducted additional forward citation searches to include all articles that had cited any of five influential reviews of animal personality: (Dall *et al.* 2004; Sih *et al.* 2004; Réale *et al.* 2007; Schuett *et al.* 2010; Dingemanse & Wolf 2010). In all cases we searched for studies that collected data on both males and females so that the sexes could be compared for data collected by the same researchers from the same population using equivalent methods. In total, we collected 9,698 records from the *ISI Web of Science* and *SCOPUS* databases on 11 December 2018. The PRISMA diagram summarising our search protocol and the number of articles located using each search method is shown in Fig. 3.

After removing duplicates, 3,739 articles were available for title and abstract screening by LMH. Of these, 942 were excluded at the title/abstract screening stage because they were: a) duplicates (*n* = 130), b) non-animal studies (*n* = 256), c) reviews or non-empirical studies (*n* = 516), d) not in English (*n* = 8), or e) inaccessible (*n* = 32). We then carried out a full-text screening of the remaining 2,797 articles. We included articles in the final dataset only if they provided raw means, variances (i.e. standard error or standard deviation) and sample sizes for behavioural measures of personality for both sexes. Raw summary statistics are required to calculate variance-based effect sizes, which also provide greater opportunities to control for sources of non-independence (Nakagawa *et al.* 2015; Noble *et al.* 2017). Consequently, studies that only report principle components or factor loadings were excluded (see Tarka *et al.* 2018). We excluded articles during full-text screening using the following criteria: 1) it was not a personality study (i.e. reported a behaviour that could not be assigned to any of the ‘Big Five’ categories, e.g. ‘social node position’), 2) the study was on domesticated or agricultural animals, 3) the study did not report separate data for males and females (i.e. the study was on hermaphrodites, clones or gynandromorphs; or data from both sexes were combined; or only one sex was measured; or data was only reported for one sex), or 4) data were unsuitable/missing (i.e. raw data was missing, missing sample sizes, sampling error type was not reported, or mean of one sex was zero leading to effect size calculation issues) (see Figure 3 and Supplementary Material). Recently, there has been a push for stricter definitions of personalities in animal behaviour studies (e.g. Dingemanse & Wright 2020). By definition, personality traits are repeatable (Sih *et al*. 2004), but very few of the available studies reported repeatability, or cited previous work that has documented repeatability of the behavioural measure used to quantify personality. As such, many earlier studies of personality-like animal behaviours do not necessarily meet these criteria. To ensure sufficient sample sizes, we therefore included behaviours that are commonly described as indices of animal personality, and where the authors interpreted behaviours as ‘personalities’ (see Sih *et al.* 2004; Réale *et al.* 2007).

In total, we identified 210 eligible articles with suitable data to calculate effect sizes (Figure 3). We extracted means, measures of variance (standard deviation or standard error) and sample sizes from the text, tables, figures or supplementary data files for both sexes for all relevant behavioural measures of personality traits that were reported by the authors. We used the R package *metaDigitise* (v1.0.0, Pick *et al.* 2019) to extract summary statistics from figures. The location of the data in the original article is provided in our raw data files (see Supplementary Material).

From our final 210 eligible studies, a further n=7 invertebrate studies were removed from analysis because our model sensitivity checks found that effect sizes calculated from scores were significantly different from the rest of the invertebrate dataset (but not for any other taxonomic group, see Supplementary Table S15). As such, we decided to remove these effect sizes, which reduced our final dataset to n=2,167 effect sizes, n=203 studies and n=226 species.

*Data transformations*

Our dataset contained some means, and associated variances, that had to be transformed to meet distribution assumptions about normality before we could calculate *g* or lnCVR. First, any latency data (e.g. time to resume behaviour) that was right-skewed was log-transformed using the following calculations to obtain means and standard deviations, respectively:

,

*Eq. 1*

*Eq. 2*

Second, some behavioural measures were presented as proportions, which constrains their distribution, so we converted them to the logit scale: mean= log (p/[1-p]); SD= to meet normality assumptions before calculating the relevant effect size. Information on which effect sizes are based on transformed values are provided in our raw data and code supplied in the Supplementary Materials.

*Effect sizes and sampling variances*

To quantify sex differences in means and variances for personality traits we used unbiased standardised effect sizes. First, to quantify the difference between the sexes in the mean value of personality traits we calculated Hedges’ *g* (sample size adjusted standardised mean difference; Hedges & Olkin 1985) as follows:

,

*Eq. 3*

,

*Eq. 4*

*Eq. 5*

Where is the mean of the behavioural measure, is the pooled standard deviation and N and SD are the sample size and standard deviation for males (M) and females (F), respectively. The associated sampling error variance of Hedges’ *g* is:

*Eq. 6*

We used Hedges *g* instead of log response ratios as some behavioural measures were based on ranks or scores. As such, response ratios could not be calculated because score and rank data are not bound to zero (Houle *et al.* 2011; Nakagawa *et al.* 2015). The relevant direction of the effect size varies depending on the focal behavioural measure of personality. For example, boldness is often measured as either ‘latency to flee’ or ‘time to resume a behaviour’ following a simulated predator approach. Here a bolder individual is therefore indicated by a larger or a smaller value respectively. We examined all measurement protocols and, where necessary, reversed the sign of the male-female difference to ensure that the direction of the effect size had a consistent interpretation for each personality trait. Specifically, a positive value of *g* indicates that males are more social, aggressive, exploratory, active or bold.

Second, to quantify a sex difference in the variance in personality traits we used the log coefficient of variation (lnCVR) (Nakagawa *et al.* 2015). In our data set there was a strong, positive relationship between the mean and variance in personality measurements (males: *r* =0.90; females: *r* =0.91). As such, using lnCVR controlled for mean-variance relationships and allowed us to quantify sex differences in variances independent of the mean (Nakagawa *et al.* 2015; Senior *et al.* 2020). We calculated lnCVR and its associated samping variance () as:

,

*Eq. 7*

*Eq. 8*

where CV*M* and CV*F* are *SDM /M* and *SDF /F* , respectively; and . and are the correlation between the logged means and standard deviations of males and females respectively. A positive value of lnCVR indicates that males are variable than females.

*Moderator variables*

We extracted information on factors that differed among studies where we had an *a priori* expectation that they might moderate the magnitude and/or direction of the effect size (Tarka *et al.* 2018a). Specifically, we recorded the taxa (‘invertebrates’, ‘fish’, ‘amphibians’, ‘reptiles’, ‘birds’, ‘mammals’), the age of individuals (‘juvenile’ or ‘adult’), whether the study population was from the lab (captive breed) or the wild, whether the behaviours were measured in the lab or field, and whether the data was collected in an experiment or during natural behaviour of the subject (‘experimental’ or ‘observational’). Most importantly, we generated two moderator variables to quantify the strength of sexual selection. First, we quantified the degree of sexual size dimorphism (SSD), which is often strongly correlated with indicators of the strength of sexual selection (e.g. primates: Cassini 2020; fish: Horne *et al.* 2020). We calculated SSD as the ratio of male to female mean body length, mass or the size of another focal, dimorphic trait (e.g. wing length) using the following index of SSD (Lovich & Gibbons 1992):

*Eq. 9*

Using the SSD index rather than the sex difference in raw means allows for: 1) a single continuous moderator that is zero when the sexes are the same size, and positive when males are larger than females; and 2) comparison of SSD across a wide range of absolute size measures. Larger values of the SSD index are interpreted as species in which there is increasingly stronger sexual selection on males. Where body size measures for males and females were not reported in the research article, we searched *Web of Science*, *Scopus* and *Google Scholar* using the search terms: “species name” AND male AND female AND body size OR length. For birds, we first searched *CRC Handbook of Avian Body Masses* (Dunning Jr 2007) and the online reference database *Birds of the World* (birdsoftheworld.org; accessed via an ANU library subscription in 2019) for body size measures. We then used data from located studies to calculate the SSD index.

Second, we also quantified mating system (“monogamous” or “multiple mating”). Where included studies did not report mating system, we searched *Web of Science*, *Scopus* and *Google Scholar* using the search terms: “species name” AND “mating system”. The location of data collected for SSD index and mating system are provided in the Supplementary Material.

Based on the number of available studies for different levels of the prospective moderators (age, population source, test location, experimental/observational), or the level of subjectivity required to categorise species (mating system), we decided upon completion of data collection that the only moderators we would formally analyse using a null hypothesis framework for their influence on the effect sizes were the SSD index and taxa. The relationships between the other moderators and the effect sizes are presented in Supplementary Tables S9-S13, but these should be treated as strictly exploratory analyses.

*Meta-analyses*

We modelled the effect sizes Hedge’s *g* and lnCVR using multi-level meta-analytic (MLMA) models (intercept only models that consider random effects) and then ran multi-level meta-regression (MLMR) models (including fixed effect moderators) in R (version 3.5.1, R Core Team 2018) using the package *metafor* (version 2.4.0, Viechtbauer 2010). We ran separate models for each of our five taxonomic groups, namely: birds, mammals, fish, invertebrates and reptiles/amphibians (combined due to low sample sizes). Although we were interested in whether sex differences varied across these taxonomic groups, the available sample sizes generally precluded running models with taxa as a moderator and would have resulted in overly complicated interpretation. Furthermore, the diversity of taxa made it challenging to create a full phylogeny that included all taxa to account for evolutionary relationships and non-independence (Noble et al. 2017). Focusing on broad taxonomic groups separately allowed us to construct phylogenies for each group. Even then, phylogenies were better resolved for some groups than others (e.g., mammal and bird ones were better than those for insects – see below).

We first estimated the overall evidence for a sex difference in the mean and variance across all personality traits using MLMA models (Table 1). This analysis is based on the simple premise that we predict males will show larger trait values for personality traits (with the possible exception of sociality) (Dingemanse & Wolf 2010; Wolf & Weissing 2010), and greater variance for all five traits in accordance with the ‘greater male variability’ hypothesis (Feingold 1992). More importantly, this analysis provides a baseline to look for sources of heterogeneity in the data. To account for the non-independence of data we included species identity and study identity as random effects, as there were multiple effect sizes from the same species or study. We also included an observation level random effect to estimate a residual / within study variance. Without this term, within-study effects are assumed to solely result from sampling variance (see Nakagawa *et al.* 2017). To correct for the non-independence of species due to their shared evolutionary history we included phylogeny as a random effect. Phylogenetic correlation matrices were derived for each taxa either from existing phylogenetic tree databases (e.g. *BirdTree.org* for birds; Jetz *et al.* 2012, 2014) or using *TimeTree.org* (Hinchliff *et al.* 2015). For the bird phylogeny, we used the Ericson tree backbone (Ericson *et al.* 2006) to generate 1,000 trees and then took a sample of 100 trees. We used *ape* and *phytools* packages in R to generate an average tree from these 100 trees, which we then used in our bird taxa models. TimeTree phylogenies were derived by exporting a list of the relevant species from *TimeTree.org*. If a species in our dataset had no phylogenetic representation we substituted the next closest available species (e.g. same genera or family). In all cases, we resolved synonymous taxa across our dataset so that species were correctly categorised and pruned our trees where needed. In the two initial MLMA models for *g* and lnCVR we derived heterogeneity estimates (*I*2; Higgins & Thompson 2002; Nakagawa & Santos 2012). We partitioned heterogeneity arising among species (*I*2species), studies (*I*2study), and due to phylogenetic heritability (*I*2phylo; Hadfield & Nakagawa 2010; Nakagawa & Santos 2012). The total heterogeneity (*I*2Total), is the proportion of the total variance in effect size estimates excluding total sampling variance (see Supplementary Material for calculations).

Next, we fit separate MLMR models for each taxonomic group that included key moderator variables. Our first set of models included personality type as a moderator to provide an estimate of the mean effect size for each of the five personality factors (i.e. ‘activity’, ‘aggression’, ‘boldness’, ‘sociability’, ‘exploration’) (Table 2). We expected that the magnitude of sex-specific differences in mean values and variance would depend on the type of personality trait, because trait types are likely to be correlated with sex roles (e.g. parental care might affect sociability; Schuett *et al.* 2010), life-histories (e.g. sex-biased dispersal is likely to affect exploration and activity; Cote *et al.* 2010), and sexual selection (e.g. the level of male-male competition or female mate choice might affect male levels of aggression and female levels of exploration; Munson *et al.* 2020). We then tested whether the degree of sexual selection, as measured by sexual size-dimorphism (SSD index), moderated effect sizes. We predicted that species with a greater male-bias in SSD would show stronger sex differences in the mean and variance. However, we also expected the strength of its moderating effect to differ among the personality traits. As such, we fit an MLMR model that included personality type, SSD and their interaction. We only fit interaction terms in these models when there were 10 or more species for each personality type. There were too few data, and low heterogeneity, for reptiles/amphibians to run a model that included interactions (Table 3).

The studies included in our meta-analysis varied greatly in their design and there were a number of additional sources of non-independence within studies (Noble *et al.* 2017). First, multiple personality traits were quantified on the same sets of individuals (e.g. ‘boldness’ and ‘aggression’). Additionally, some studies measured the same individuals multiple times for the same trait (i.e. repeatability), or for the same trait type using a different test (e.g. several measures of boldness). We conducted a series of sensitivity analyses to ascertain the impact that these sources of non-independence had on our final results. We created correlation matrices among effect sizes that shared the same sets of individuals in the sample used to derive effect sizes. Given that we did not know the exact correlation among traits we created three different dependency matrices (**D**; i.e., correlation matrices) that assumed r = 0.3, r = 0.5 and r = 0.8. We refit our models replacing the identity matrix (**I**) that was assumed when estimating the residual error variance with our **D** matrices.

For all models we present meta-analytic mean estimates and 95% confidence intervals (Tables 1-3).

*Publication bias*

Published studies might disproportionately report certain findings (e.g. greater male variability). To look for publication bias, we first checked for funnel plot asymmetry for both *g* and lnCVR. Studies with large error (e.g. low precision) drive funnel plot asymmetry if there is a true non-zero effect and a bias towards publishing significant results (REF). Visual inspection of funnel plots is misleading, however, as we need to account for additional sources of variation in effect sizes (i.e. moderator variables and random factors) beyond effect size precision. We therefore included precision as a moderator term in our MLMR models to test where it explained some of the variation in the reported effect sizes (Van Aert *et al.* 2019). If precision has a significant influence, this is suggestive of publication bias. Precision was calculated as:

*Eq. 10*

Where is the sampling variance of effect sizes’ Hedge’s *g* (Eq. 3-6) or lnCVR (Eq. 7-8).

**Data availability**

xxx

**References**

Van Aert, R.C.M., Wicherts, J.M. & Van Assen, M.A.L.M. (2019). *Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis*. *PLoS One*.

Amos-Landgraf, J.M., Cottle, A., Plenge, R.M., Friez, M., Schwartz, C.E., Longshore, J., *et al.* (2006). X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am. J. Hum. Genet.*, 79, 493–499.

Archer, J. & Mehdikhani, M. (2003). Variability among Males in Sexually Selected Attributes. *Rev. Gen. Psychol.*, 7, 219–236.

Arden, R. & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Pers. Individ. Dif.*, 41, 39–48.

Arnold, A.P. (2004). Sex chromosomes and brain gender. *Nat. Rev. Neurosci.*, 5, 701–708.

Baye, A. & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessments Educ.*, 4, 1–16.

Beery, A.K. & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.*, 35, 565–572.

Beukeboom, L.W. & Perrin, N. (2014). *The evolution of sex determination*. Oxford University Press, New York.

Borkenau, P., McCrae, R.R. & Terracciano, A. (2013). Do men vary more than women in personality? A study in 51 cultures. *J. Res. Pers.*, 47, 135–144.

Buss, D.M. (1995). Psychological Sex Differences: Origins Through Sexual Selection. *Am. Psychol.*, 50, 164–168.

Campbell, A. (1999). Staying alive: Evolution, culture, and women’s intrasexual aggression. *Behav. Brain Sci.*, 22, 203–252.

Carruth, L.L., Reisert, I. & Arnold, A.P. (2002). Sex chromosome genes directly affect brain sexual differentiation. *Nat. Neurosci.*, 5, 933–934.

Cassini, M.H. (2020). Sexual size dimorphism and sexual selection in primates. *Mamm. Rev.*, 50, 231–239.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.*, 6, 149–162.

Costa, P.T. & McCrae, R.R. (1992). Four ways five factors are basic. *Pers. Individ. Dif.*, 13, 653–665.

Cote, J., Clobert, J., Brodin, T., Fogarty, S. & Sih, A. (2010). Personality-dependent dispersal: Characterization, ontogeny and consequences for spatially structured populations. *Philos. Trans. R. Soc. B Biol. Sci.*, 365, 4065–4076.

Dall, S.R.X., Houston, A.I. & McNamara, J.M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecol. Lett.*, 7, 734–739.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. Murray, London.

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

DeCasien, A.R., Sherwood, C.C., Schapiro, S.J. & Higham, J.P. (2020). Greater variability in chimpanzee ( *Pan troglodytes* ) brain structure among males. *Proc. R. Soc. B Biol. Sci.*, 287, 20192858.

Dingemanse, N.J. & Wolf, M. (2010). Recent models for adaptive personality differences: A review. *Philos. Trans. R. Soc. B Biol. Sci.*, 365, 3947–3958.

Dingemanse, N.J. & Wright, J. (2020). Criteria for acceptable studies of animal personality and behavioural syndromes. *Ethology*, 126, 865–869.

Dunning Jr, J.B. (2007). *CRC Handbook of Avian Body Masses*. 2nd Editio. CRC Press, Boca Raton.

Ericson, P.G.P., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, M., *et al.* (2006). Diversification of Neoaves: Integration of molecular sequence data and fossils. *Biol. Lett.*, 2, 543–547.

Feingold, A. (1992). Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy. *Rev. Educ. Res.*, 62, 61–84.

Fitzpatrick, M.J., Feder, E., Rowe, L. & Sokolowski, M.B. (2007). Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature*, 447, 210–212.

Fromhage, L. & Jennions, M.D. (2016). Coevolution of parental investment and sexually selected traits drives sex-role divergence. *Nat. Commun.*, 7.

Gordon, S.P., Lopez-Sepulcre, A. & Reznick, D.N. (2011). Predation-associated differences in sex linkage of wild guppy coloration. *Evolution (N. Y).*, 66, 912–918.

Graves, J.A.M. & Shetty, S. (2001). Sex from W to Z: Evolution of vertebrate sex chromosomes and sex determining genes. *J. Exp. Zool.*, 290, 449–462.

Gray, H., Lyth, A., McKenna, C., Stothard, S., Tymms, P. & Copping, L. (2019). Sex differences in variability across nations in reading, mathematics and science: a meta-analytic extension of Baye and Monseur (2016). *Large-Scale Assessments Educ.*, 7.

Gross, M.R. (1996). Alternative reproductive strategies and tactics: diversity within sexes. *TREE*, 11, 92–98.

Hadfield, J.D. & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.*, 23, 494–508.

Halpern, D.F. & LaMay, M.L. (2000). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. *Educ. Psychol. Rev.*, 12, 229.

Hedges, L. V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. *Biometrics*. Academic Press, New York.

Higgins, J.P.T. & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, 21, 1539–1558.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., *et al.* (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci.*, 112, 12764–12769.

Horne, C.R., Hirst, A.G. & Atkinson, D. (2020). Selection for increased male size predicts variation in sexual size dimorphism among fish species. *Proc. R. Soc. B Biol. Sci.*, 287.

Houle, D., Pélabon, C., Wagner, G. & Hansen, T.F. (2011). Measurement and meaning in biology. *Q. Rev. Biol.*, 86, 3–34.

Hunt, J., Brooks, R., Jennions, M.D., Smith, M.J., Bentsen, C.L. & Bussière, L.F. (2004). High-quality male field crickets invest heavily in sexual display but die young. *Nature*, 432, 1024–1027.

James, J.W. (1973). 353. Note: Covariances Between Relatives due to Sex-Linked Genes. *Biometrics*, 29, 584.

Janicke, T., Häderer, I.K., Lajeunesse, M.J. & Anthes, N. (2016). Evolutionary Biology: Darwinian sex roles confirmed across the animal kingdom. *Sci. Adv.*, 2, 1–11.

Janicke, T. & Morrow, E.H. (2018). Operational sex ratio predicts the opportunity and direction of sexual selection across animals. *Ecol. Lett.*, 21, 384–391.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012). The global diversity of birds in space and time. *Nature*, 491, 444–448.

Jetz, W., Thomas, G.H., Joy, J.B., Redding, D.W., Hartmann, K. & Mooers, A.O. (2014). Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Curr. Biol.*, 24, 919–930.

Johnson, W., Deary, I.J. & Carothers, A. (2008). Sex Differences in Variability in General Intelligence: A New Look at the Old Question. *Perspect. Psychol. Sci.*, 3, 518–531.

Johnson, W., Deary, I.J. & Carothers, A. (2009). A Role for the X Chromosome in Sex Differences in Variability in General Intelligence? *Perspect. Psychol. Sci.*, 4, 598–611.

Jones, C.M., Braithwaite, V.A. & Healy, S.D. (2003). The evolution of sex differences in spatial ability. *Behav. Neurosci.*, 117, 403–411.

Ju, C., Duan, Y. & You, X. (2015). Retesting the greater male variability hypothesis in mainland China: A cross-regional study. *Pers. Individ. Dif.*, 72, 85–89.

Kappeler, P.M. (2017). Sex roles and adult sex ratios: insights from mammalian biology and consequences for primate behaviour. *Phil. Trans. R. Soc. B*, 372, 20160321.

Karwowski, M., Jankowska, D.M., Gajda, A., Marczak, M., Groyecka, A. & Sorokowski, P. (2016). Greater Male Variability in Creativity Outside the WEIRD World. *Creat. Res. J.*, 28, 467–470.

Lehre, A.C., Lehre, K.P., Laake, P. & Danbolt, N.C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Dev. Psychobiol.*, 51, 198–206.

van der Linden, D., Dunkel, C.S. & Madison, G. (2017). Sex differences in brain size and general intelligence (g). *Intelligence*, 63, 78–88.

Lovich, J.E. & Gibbons, J.W. (1992). A review of techniques for quantifying sexual size dimorphism. *Growth Dev Aging*, 56, 269–281.

Lyon, M.F. (1961). Gene Action in the X-chromosom (Mus musculus L.). *Nature*, 190, 372–373.

Machin, S. & Pekkarinen, T. (2008). Assessment: Global sex differences in test score variability. *Science (80-. ).*, 322, 1331–1332.

McEvoy, J., While, G.M., Sinn, D.L. & Wapstra, E. (2013). The role of size and aggression in intrasexual male competition in a social lizard species, Egernia whitii. *Behav. Ecol. Sociobiol.*, 67, 79–90.

McGhee, K.E. & Travis, J. (2010). Repeatable behavioural type and stable dominance rank in the bluefin killifish. *Anim Behav*, 79, 497–507.

Miller, D.I. & Halpern, D.F. (2014). The new science of cognitive sex differences. *Trends Cogn. Sci.*, 18, 37–45.

Munson, A.A., Jones, C., Schraft, H. & Sih, A. (2020). You ’ re Just My Type : Mate Choice and Behavioral Types. *Trends Ecol. Evol.*, 1–11.

Nakagawa, S., Noble, D.W.A., Senior, A.M. & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biol.*, 15, 1–14.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., *et al.* (2015). Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods Ecol. Evol.*, 6, 143–152.

Nakagawa, S. & Santos, E.S.A. (2012). Methodological issues and advances in biological meta-analysis. *Evol. Ecol.*, 26, 1253–1274.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *Am. Psychol.*, 61, 622–631.

Noble, D.W.A., Lagisz, M., O’dea, R.E. & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Mol. Ecol.*, 2410–2425.

O’Dea, R.E., Lagisz, M., Jennions, M.D. & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nat. Commun.*, 9.

Parker, G.. (1970). Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.*, 45, 535–567.

Pick, J.L., Nakagawa, S. & Noble, D.W.A. (2019). Reproducible, flexible and high-throughput data extraction from primary literature: The metaDigitise r package. *Methods Ecol. Evol.*, 10, 426–431.

Piper, W.H., Walcott, C., Mager, J.N. & Spilker, F.J. (2008). Fatal battles in common loons: a preliminary analysis. *Anim. Behav.*, 75, 1109–1115.

Pomiankowski, A. & Moller, A.P. (1995). A resolution of the lek paradox. *Proc. R. Soc. B Biol. Sci.*, 260, 21–29.

PRICE, T., SCHLUTER, D. & HECKMAN, N.E. (1993). Sexual selection when the female directly benefits. *Biol. J. Linn. Soc.*, 48, 187–211.

R Core Team. (2016). R: A language and environment for statistical computing.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. (2007). Integrating animal temperament within ecology and evolution. *Biol. Rev.*, 82, 291–318.

Reinhold, K. & Engqvist, L. (2013a). The variability is in the sex chromosomes. *Evolution (N. Y).*, 67, 3662–3668.

Reinhold, K. & Engqvist, L. (2013b). The variability is in the sex chromosomes. *Evolution (N. Y).*, 67, 3662–3668.

Reiss, M.J. (1986). Sexual dimorphism in body size: Are larger species more dimorphic? *J. Theor. Biol.*, 121, 163–172.

Roalf, D.R., Gur, R.E., Ruparel, K., Calkins, M.E., Satterthwaite, T.D., Bilker, W.B., *et al.* (2014). Within-individual variability in neurocognitive performance: Age- and sex-related differences in children and youths from ages 8 to 21. *Neuropsychology*, 28, 506–518.

Rohner, P.T., Blanckenhorn, W.U. & Puniamoorthy, N. (2016). Sexual selection on male size drives the evolution of male-biased sexual size dimorphism via the prolongation of male development. *Evolution (N. Y).*, 70, 1189–1199.

Rowe and Houle. (1996). The lek paradox and the capture of genetic variance. *Proc. Biol. Sci.*, 263, 1415–1421.

Schuett, W., Tregenza, T. & Dall, S.R.X. (2010). Sexual selection and animal personality. *Biol. Rev.*, 85, 217–246.

Senior, A.M., Viechtbauer, W. & Nakagawa, S. (2020). Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio. *Res. Synth. Methods*, 11, 553–567.

Sih, A., Bell, A.M. & Johnson, J.C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *TREE*, 19, 372–378.

Skuse, D.H. (2006). Sexual dimorphism in cognition and behaviour: The role of X-linked genes. *Eur. J. Endocrinol. Suppl.*, 155, 99–106.

Smith, B.R. & Blumstein, D.T. (2008). Fitness consequences of personality: A meta-analysis. *Behav. Ecol.*, 19, 448–455.

Snell, D.M. & Turner, J.M.A. (2018). Sex Chromosome Effects on Male–Female Differences in Mammals. *Curr. Biol.*, 28, R1313–R1324.

Tarka, M., Guenther, A., Niemelä, P.T., Nakagawa, S. & Noble, D.W.A. (2018a). Sex differences in life history, behavior, and physiology along a slow-fast continuum: a meta-analysis. *Behav. Ecol. Sociobiol.*, 72, 132.

Tarka, M., Guenther, A., Niemelä, P.T., Nakagawa, S. & Noble, D.W.A. (2018b). Sex differences in life history, behavior, and physiology along a slow-fast continuum: a meta-analysis. *Behav. Ecol. Sociobiol.*, 72, 132.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.*, 36, 1–48.

Wierenga, L.M., Doucet, G., Dima, D., Agartz, I., Aghajani, M., Akudjedu, T., *et al.* (2020). Greater male than female variability in regional brain structure across the lifespan. *bioRxiv*, 2020.02.17.952010.

Wolf, M., Van Doorn, G.S., Leimar, O. & Weissing, F.J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, 447, 581–584.

Wolf, M. & McNamara, J.M. (2012). On the evolution of personalities via frequency-dependent selection. *Am. Nat.*, 179, 679–692.

Wolf, M. & Weissing, F.J. (2010). An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. B Biol. Sci.*, 365, 3959–3968.

Wyman, M.J. & Rowe, L. (2014). Male Bias in Distributions of Additive Genetic, Residual, and Phenotypic Variances of Shared Traits. *Am. Nat.*, 184, 326–337.

Zajitschek, S., Zajitschek, F., Bonduriansky, R., Brooks, R.C., Cornwell, W., Falster, D.S., *et al.* (2020). Sex and Power: sexual dimorphism in trait variability and its eco-evolutionary and statistical implications. *BioRxiv*, 1–20.

Table 1. Multi-level meta-analytic models for each taxonomic group for a sex difference in the mean (SMD) and variability (lnCVR) in personality traits. Positive estimates indicate that the mean or variability in personality is greater for males than females.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameters** | **Overall mean** | **95% CI** | **Prediction interval** | ***t* score** | ***p*-value** | ***n* effect sizes** | ***n studies*** | ***n species*** |
| *SMD (Hedge’s g - Mean)* |  |  |  |  |  |  |  |  |
| Mammals | 0.08 | -0.28, 0.45 | -1.10, 1.27 | 0.44 | 0.66 | 674 | 61 | 45 |
| Birds | -0.14 | -0.36, 0.09 | -1.82, 1.57 | -1.17 | 0.24 | 483 | 50 | 106 |
| Reptiles / Amphibians | 0.07 | -0.08, 0.22 | -0.52, 0.67 | 0.94 | 0.35 | 95 | 11 | 10 |
| Fish | -0.04 | -0.35, 0.28 | -1.00, 0.93 | -0.24 | 0.28 | 493 | 44 | 22 |
| Invertebrates | 0.30 | -0.02, 0.62 | -1.89, 2.48 | 1.82 | 0.07 | 422 | 37 | 36 |
|  |  |  |  |  |  |  |  |  |
| *lnCVR (Variance)* |  |  |  |  |  |  |  |  |
| Mammals | 0.07 | -0.20, 0.34 | -0.64, 0.79 | 0.51 | 0.61 | 674 | 61 | 45 |
| Birds | -0.14 | -0.65, 0.37 | -1.94, 1.64 | -0.56 | 0.58 | 483 | 50 | 106 |
| Reptiles / Amphibians | 0.05 | -0.04, 0.14 | -0.06, 0.15 | 1.13 | 0.26 | 95 | 11 | 10 |
| Fish | -0.04 | -0.09, 0.01 | -0.64, 0.56 | -1.44 | 0.15 | 493 | 44 | 22 |
| Invertebrates | 0.00 | -0.13, 0.12 | -0.77, 0.76 | -0.04 | 0.97 | 422 | 37 | 36 |

Table 2. MLMR models for each taxa with personality type as a fixed moderator. Values highlighted in bold indicate a significant difference (< 0.05) between males and females. Positive estimates indicate that the mean and variability are greater for males than females. These models are graphically represented in Figures 1 & 2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameters** | **SMD**  **mean** | **SMD 95%CI** | **SMD**  ***p-value*** | **lnCVR**  **mean** | **lnCVR**  **95%CI** | **lnCVR**  ***p-value*** | **n effect sizes** | **n studies** | **n species** |
| ***Mammals*** | | | | | | |  |  |  |
| Activity | -0.17 | -0.56, 0.23 | 0.40 | 0.10 | -0.20, 0.41 | 0.50 | 84 | 14 | 12 |
| Aggression | 0.10 | -0.27, 0.48 | 0.59 | 0.11 | -0.20, 0.42 | 0.49 | 87 | 16 | 14 |
| Boldness | 0.15 | -0.20, 0.50 | 0.39 | 0.06 | -0.22, 0.34 | 0.67 | 193 | 27 | 27 |
| Exploration | 0.05 | -0.31, 0.41 | 0.79 | 0.04 | -0.25, 0.34 | 0.78 | 213 | 19 | 16 |
| Sociality | 0.09 | -0.29, 0.47 | 0.64 | 0.06 | -0.25, 0.37 | 0.70 | 97 | 12 | 10 |
|  | | | | | | |  |  |  |
| ***Birds*** | | | | | | |  |  |  |
| Activity | -0.14 | -0.43, 0.15 | 0.35 | 0.05 | -0.24, 0.34 | 0.72 | 63 | 14 | 9 |
| Aggression | -0.14 | -0.43, 0.14 | 0.33 | -0.07 | -0.39, 0.25 | 0.68 | 50 | 11 | 10 |
| Boldness | -0.19 | -0.44, 0.06 | 0.13 | -0.005 | -0.23, 0.22 | 0.97 | 261 | 24 | 96 |
| Exploration | 0.09 | -0.18, 0.36 | 0.51 | -0.25 | -0.50, 0.01 | 0.06 | 78 | 16 | 9 |
| **Sociality** | **-0.68** | **-1.16, 0.21** | **0.005** | 0.14 | -0.38, 0.66 | 0.60 | 31 | 3 | 2 |
|  | | | | | | |  |  |  |
| ***Reptiles / Amphibians*** | | | | | | |  |  |  |
| Activity | -0.05 | -0.45, 0.36 | 0.82 | -0.11 | -0.51, 0.29 | 0.60 | 5 | 3 | 3 |
| Aggression | -0.07 | -0.32, 0.19 | 0.60 | 0.33 | -0.05, 0.72 | 0.09 | 30 | 2 | 2 |
| Boldness | 0.08 | -0.15, 0.31 | 0.68 | 0.10 | -0.13, 0.33 | 0.41 | 25 | 4 | 3 |
| **Exploration** | **0.25** | **0.05, 0.45** | **0.02** | -0.10 | -0.35, 0.14 | 0.40 | 32 | 5 | 4 |
| Sociality | -0.05 | -0.60, 0.50 | 0.86 | -0.12 | -0.76, 0.52 | 0.70 | 3 | 2 | 2 |
|  | | | | | | |  |  |  |
| ***Fish*** | | | | | | |  |  |  |
| Activity | -0.16 | -0.65, 0.33 | 0.53 | -0.03 | -0.16, 0.09 | 0.63 | 92 | 9 | 5 |
| **Aggression** | -0.05 | -0.52, 0.42 | 0.53 | **-0.13** | **-0.25, -0.01** | **0.04** | **95** | **17** | **14** |
| Boldness | -0.16 | -0.63, 0.32 | 0.52 | -0.02 | -0.63, 0.59 | 0.59 | 174 | 24 | 13 |
| Exploration | -0.05 | -0.54, 0.44 | 0.84 | -0.03 | -0.16, 0.09 | 0.62 | 103 | 10 | 7 |
| Sociality | -0.40 | -0.91, 0.11 | 0.12 | 0.07 | -0.11, 0.24 | 0.47 | 29 | 7 | 6 |
|  | | | | | | |  |  |  |
| ***Invertebrates*** | | | | | | |  |  |  |
| Activity | 0.33 | -0.04, 0.70 | 0.08 | -0.06 | -0.19, 0.08 | 0.41 | 166 | 18 | 17 |
| Aggression | 0.35 | -0.36, 1.06 | 0.33 | 0.17 | -0.10, 0.43 | 0.22 | 35 | 5 | 6 |
| Boldness | 0.31 | -0.05, 0.67 | 0.09 | -0.04 | -0.16, 0.08 | 0.56 | 161 | 22 | 22 |
| Exploration | 0.00 | -0.44, 0.45 | 0.98 | 0.07 | -0.12, 0.26 | 0.47 | 54 | 7 | 6 |
| Sociality | 0.39 | -0.38, 1.16 | 0.32 | 0.27 | -0.14, 0.68 | 0.20 | 6 | 1 | 1 |

Table 3. Subset analyses with personality trait type and SSD as moderator terms for four of the five taxonomic groups. Estimates highlighted in bold indicate significant differences between males and females. Positive estimates indicate that the mean and variability are greater for males than females. Separate models were run for each trait type with SSD as a moderator, therefore estimates show the personality trait mean when males and females are the same size, and when males are larger than females. Only interactions with 10 or more species for each personality type were estimated.

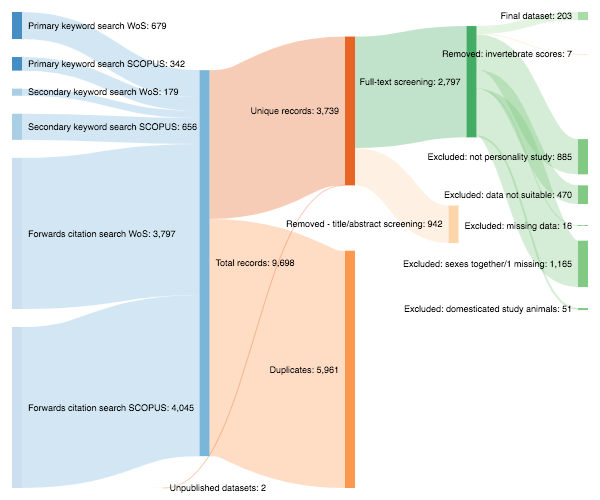
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameters** | **SMD**  **mean** | **SMD 95%CI** | **SMD**  ***p-value*** | **lnCVR**  **mean** | **lnCVR**  **95%CI** | **lnCVR**  ***p-value*** | **n studies** | **n**  **species** | **n effect sizes** |
| ***Mammals*** | | | | | | |  |  |  |
| Activity | 0.44 | -1.74, 2.62 | 0.69 | 0.05 | -0.15, 0.25 | 0.60 | 14 | 12 | 84 |
| **SSD** | **-2.16** | **-3.99, -0.32** | **0.02** | 0.13 | -0.56, 0.81 | 0.72 |  |  |  |
| Aggression | -0.09 | -1.29, 1.10 | 0.88 | 0.09 | -0.21, 0.39 | 0.56 | 15 | 13 | 85 |
| **SSD** | **1.36** | **-0.01, 2.73** | **0.05** | -0.05 | -1.43, 1.33 | 0.94 |  |  |  |
| Boldness | 0.09 | -0.09, 0.27 | 0.34 | 0.07 | -0.03, 0.16 | 0.16 | 26 | 26 | 163 |
| SSD | -0.16 | -0.50, 0.17 | 0.34 | 0.08 | -0.09, 0.25 | 0.35 |  |  |  |
| Exploration | 0.00 | -0.18, 0.18 | 0.99 | -0.06 | -0.36, 0.24 | 0.69 | 19 | 16 | 213 |
| SSD | -0.05 | -0.60, 0.50 | 0.85 | 0.13 | -0.37, 0.64 | 0.61 |  |  |  |
|  | | | | | | |  |  |  |
| ***Birds*** | | | | | | |  |  |  |
| Boldness | -0.75 | -0.87, 0.33 | 0.38 | 0.03 | -0.04, 0.11 | 0.37 | 21 | 78 | 233 |
| SSD | -0.23 | -1.45, 0.98 | 0.70 | 0.11 | -0.16, 0.37 | 0.44 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| ***Fish*** |  |  |  |  |  |  |  |  |  |
| **Aggression** | -0.16 | -0.96, 0.63 | 0.68 | **-0.12** | **-0.23, 0.00** | **0.05** | 16 | 13 | 93 |
| SSD | 0.27 | -0.84, 1.37 | 0.63 | -0.13 | -0.81, 0.55 | 0.70 |  |  |  |
| Boldness | 0.06 | -0.23, 0.34 | 0.70 | -0.04 | -0.33, 0.25 | 0.78 | 23 | 12 | 172 |
| SSD | -0.32 | -0.93, 0.29 | 0.30 | 0.10 | -0.31, 0.52 | 0.63 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| ***Invertebrates*** |  |  |  |  |  |  |  |  |  |
| Activity | 0.31 | -0.42, 1.04 | 0.41 | -0.04 | -0.25, 0.17 | 0.72 | 18 | 16 | 165 |
| SSD | -0.66 | -2.26, 0.93 | 0.41 | 0.27 | -0.53, 1.07 | 0.50 |  |  |  |
| Boldness | 0.18 | -0.03, 0.38 | 0.10 | -0.04 | -0.16, 0.09 | 0.56 | 22 | 22 | 161 |
| SSD | 0.28 | -0.54, 1.10 | 0.50 | 0.02 | -0.47, 0.51 | 0.95 |  |  |  |



**Figure 1**. Phylogenetic relationships and orchard plots for a) mammals and b) birds. Node colours reflect SSD ratios for each species (male-biased are blue, female-biased are red, monomorphic are white and no data are grey) and heatmap data shows *n* effect sizes for each personality trait type for each species. Orchard plots show total effect sizes and mean effect size for SMD and lnCVR meta-regression models with personality trait as a moderator term. Thin bars are 95% CIs and thick bars are prediction intervals. Circle size reflects effect size precision where larger circles have greater precision.

****

**Figure 2.** Phylogenetic relationships and orchard plots for c) reptiles and amphibians, d) fish and e) invertebrates. Node colours reflect SSD ratios for each species and heatmap data shows *n* effect sizes for each personality trait type for each species. Orchard plots show total effect sizes and mean effect size for SMD and lnCVR meta-regression models with personality trait as a moderator term. Thin bars are 95% CIs and thick bars are prediction intervals. Circle size reflects effect size precision where larger circles have greater precision. Code and data used to generate Figures 1 & 2 are available in the accompanying Supplementary Material.



**Figure 3.** *PRISMA* diagram showing the process of finding, screening and including/excluding studies for this meta-analysis. See Supplementary Material accompanying this article for detailed keyword search terms.