**No meta-analytic support for greater male variability in animal personalities**

**What about:**

**Personality tests: men are more variable than women, but are males more variable than females?**

**Abstract**

Human and animal studies alike often report ‘greater male variability’; males have greater representation at the extremes of most traits by virtue of their sex chromosomes. Men have greater variability for personality traits that attract women, yet it is unclear whether animal personalities show similar patterns. We conducted a meta-analysis of animal personality behaviours (*n*=2,248 effect sizes) across five taxonomic groups (*n*=231 species) to look for sex differences in personalities and the influence of sexual selection (degree of sexual size dimorphism, SSD) on trait means and variability. We found some male- and female-biased sex differences in personality means, but differences in trait variance always favoured greater female, not male, variability. Finally, SSD had significant interactions only with mean personality for mammals and invertebrates, which could reflect sex differences in life-histories rather than sexual selection. We suggest future studies take caution when discussing greater male variability for shared traits.

**Introduction**

*These first two paragraphs just won’t work no matter what I try to write – help!*

That males are more variable than females is a trend observed across the animal kingdom, consistent with the ‘greater male variability’ hypothesis. Male-biased variability than females across a range of behavioural, physiological and morphological traits (Pomiankowski & Møller 1995). In humans, male-biased sex differences in variability have been reported for traits like intelligence (Arden & Plomin 2006; Johnson *et al.* 2008), birth weights and several blood parameters (Lehre *et al.* 2009), and creativity (Ju *et al.* 2015; Karwowski *et al.* 2016). Furthermore, there are often more men at the extreme ends of aggressive personalities, leading to their overrepresentation in the prison system (Deary *et al.* 2003). Yet the mechanisms behind such widespread greater male variability remain unclear.

In the realm of human personality, our behavioural quirks can be categorised into one of five factors: Extraversion, Neuroticism, Openness, Conscientiousness and Agreeableness (Costa & McCrae 1992). Like most biological traits, these ‘Big Five’ factors are continuous and there are costs and benefits associated with personalities that fall at their extremes. For example, … Like human personalities, consistent animal behaviours, or personalities, can be grouped into trait types that resemble the human ‘big five’: Activity, Aggression, Boldness, Exploration and Sociality/Sociability (Reale *et al*. 2007). For organisms living in the wild, consistent and predictable (i.e. repeatable) behaviour across time and contexts seems maladaptive, especially when there is such large inter-individual variation in response to the same environmental conditions (Dingemanse *et al.* 2010). However, the consistency of personality within individuals, and its persistence across generations, suggests that low intra-individual variation can be adaptive (Smith & Blumstein 2008a; Dingemanse & Wolf 2010). On the surface, human and animal personality dimensions seem to share behavioural traits, so it’s possible that the mechanisms behind such widespread, sex-biased variation in human personalities might also drive sex differences in animal personalities.

Firstly, the ‘sex chromosome’ hypothesis (James 1973) suggests that greater male variability is the result of sex chromosomal arrangement. For mammals, females have two copies of the X chromosome (homogamety), so the effects of any genes found on the X chromosome become averaged out. The averaging of gene expression can occur either through epigenetic inactivation of some regions during development (Amos-Landgraf *et al.* 2006), or through mosaicism where the ratio of genes on either X chromosome are variably expressed (Lyon 1961). Male mammals, however, only have one copy of the X chromosome (heterogamety). As such, all genes present on the male copy of the X chromosome are fully expressed, leading to more extreme phenotypes and, consequently, higher variances for shared traits (i.e. dosage compensation; see review by Charlesworth 1996). However, not all males are heterogametic – male birds have ZZ sex chromosomal arrangement while female birds have ZW, making females the heterogametic sex (reviewed in Graves & Shetty 2001). If sex chromosomal arrangement is the mechanism driving sex-biased variability, species with heterogametic females should have greater trait variability than males. Indeed, in a meta-analysis covering a broad range of taxonomic groups, Reinhold & Engqvist (2013) found that for species with heterogametic males, males had greater variability in body size than did females, and in species with heterogametic females, females had greater variability in body size than males. Sex chromosomal arrangement and greater male variability can also extend to behaviours, as they too have a genetic underpinning (Dochtermann *et al.* 2014). For example, men have greater representation at the tail ends of intelligence distributions, even when there are no mean differences in intelligence between men and women (Feingold 1992; Halpern & LaMay 2000; Arden & Plomin 2006; Johnson *et al.* 2008).

Second, under sexual selection, the sex with the greatest variation in their mating success (often males) should also have greater variation in traits used to obtain matings (e.g. weapons, ornaments) (Darwin 1871). In order for sexual selection to operate, traits important for male reproduction should have increased variance, compared to the same traits in females or other non-sexually selected traits (Pomiankowski & Møller 1995). For example, a meta-analysis comparing phenotypic trait variances for males and females found that there was a significant male-bias for coefficients of phenotypic variance for reproductive and non-reproductive traits (Wyman & Rowe 2014). However, traits important for reproduction showed more male-biased phenotypic variance than traits unrelated to reproduction (Wyman & Rowe 2014). While it is reasonable to expect male sexual traits to have greater variability than non-sexually selected traits, traits shared by both sexes are also frequently expected to show male-biased variability. For example, male-male competition plays an important role in male reproduction in many animal systems, thus aggression is a personality trait where males might have greater variability than females. In human antisocial personality types (low Agreeableness and high Neuroticism; similar to the Aggression dimension in animals) males show greater variability than females (Budaev 1999, Archer & Mehdikhani 2003). Greater male variability in human aggression is thought to be maintained by negative frequency-dependent selection whereby aggressive tendencies have both costs (risk of physical injuries) and benefits (more mating opportunities via social dominance) for male reproduction (Buss 1996, Budaev 1999).

Non-random mate choice might also contribute to sex differences in personalities. Females might prefer males that consistently behave in a certain way if such behaviour signals that they can provide direct or indirect benefits to her (Jennions & Petrie 1997; Schuett *et al.* 2010; Munson *et al.* 2020). For example, boldness might function as an honest signal of a potential partner’s quality; consistently bolder males might be better able to provision offspring with food compared to shy males, thus providing direct benefits to the female for her choice (Munson *et al*. 2020). Where the choosy sex (usually females) shows a strong preference for partners with a particular trait, that trait will have increased variance because it is under sexual selection (Pomiankowski & Møller 1995). Finally, different sex roles can also drive sex differences in behaviour. Both mating system and parental care can change the strength of sexual selection on one sex (usually males), producing larger variances in male reproductive success (Janicke *et al.* 2016). For example, species with monogamous males that share parental care might have no sex differences in their personalities, but where males mate multiply and do not care for offspring there might be more male-biased personality traits. Regardless of the underlying mechanism, most hypotheses surrounding sex differences in all traits tend to focus on greater male variability, as males tend to benefit more, or incur greater costs, for having greater representation at the tail-ends of trait distributions.

To investigate the prevalence of male-biased sex differences in variability for shared behavioural traits, we chose to compare animal personality traits across five broad taxonomic groups. Personalities, as part of their definition, should be consistent over time (i.e. low intra-individual variation), making personality a useful trait for comparing sex differences in behaviours. Additionally, tests used to measure personality are often similar enough to enable comparison over many studies, allowing us to conduct a systematic review and meta-analysis of sex differences in personalities. Although sex differences in animal personalities have been the focus of a recent meta-analysis (Tarka *et al*. 2018), it’s unclear whether the mechanisms behind greater male variability in human personalities are similarly responsible for sex differences in animal personalities. Therefore, we chose to include SSD as an additional moderator in our meta-analysis because it is a good proxy for the strength of sexual selection (i.e. larger male size reflects a condition-dependent trait, polygynous mating system, and/or different sex roles; Reiss 1986). Additionally, including SSD as a moderator can help disentangle the expected roles of sexual selection and sex chromosomal arrangement on sex-biased variability. First, if males do have greater variability than females in their personality traits, because of sexual selection, then SSD should be a significant moderator of such differences. Second, if sex chromosomal arrangement drives sex differences in behaviour, we should see greater male variability in taxonomic groups where males are the homogametic sex (i.e. mammals) and greater female variability in taxonomic groups where females are the homogametic sex (i.e. birds).

We had three main questions around which we formed our three meta-analytic models: 1) do males and females differ in either their mean trait expression, or variability in personality, 2) are there sex differences in mean or variability for the different personality types, and 3) does the degree of sexual size dimorphism explain sex differences in personality means or variability?

**Results**

*Summary of the Dataset*

Our final dataset comprised 2,248 effect sizes from five broad taxonomic groups: mammals, birds, reptilia (amphibians and reptiles combined), fish and invertebrates. The number of species (n = 10-106), studies (n = 11-61) and effect sizes (n = 95-684) per taxa are shown in Table 1. Boldness was the most well-studied and sociality the least-studied of the five personality types (n=825 and 169 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 meta-analytic intercept models all had high heterogeneity (*I*2Total SMD > 0.70; *I*2Total lnCVR > 0.50), except for reptilia (*I*2Total SMD = 0.45; *I*2Total lnCVR = 0.01) and fish (*I*2Total lnCVR = 0.49). Heterogeneity within mean personality effect sizes mostly came from between-study differences (*I*2StudyID), while phylogenetic relationships and among species differences (*I*2phylo and *I*2species, respectively) explained heterogeneity in variability effect sizes for endotherms and reptilia only (see Supplementary Table S1).

The lack of a sex difference in mean 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 four cases. Females were significantly more sociable in birds (=-0.68, *p*=0.005), less aggressive and less bold in invertebrates (=0.45, *p*=0.05; =0.32, *p*=0.04), and less exploratory in reptiles (= 0.25, *p*=0.02). There were no significant sex differences in mean values for mammals or fish (Table 2, Figures 1 & 2).

There were also significant sex differences in variability in personality in two cases. Female were more variable in their exploratory behaviour in birds (=-0.26, *p*=0.05), and their aggressive behaviour in fish (=-0.13, *p*=0.04). There were no significant sex differences in variability in invertebrates, mammals or reptiles (Table 2, Figures 1 & 2 and Supplementary Table S2). Given that 50 tests for sex differences were conducted, none of the six sex differences were significant after correcting for multiple testing (critical Bonferroni p-value of 0.05/50 = 0.001).

*Sexual Size Dimorphism and Personality Differences between the Sexes*

Sexual size dimorphism (SSD) had a significantly effect on sex differences in mean personality in two of the four taxa test (Supplementary Tables S5 & S6), but its influence depended on the personality trait (Table 3). There was no effect of SSD in fish or birds. In invertebrates, the effect of SSD differed significantly for boldness and activity. When the sexes were the same size there was no difference in either boldness (*β* =0.37, *p*=0.06) or activity (*β* =0.03, *p*=0.90). When males were larger than females, males were less active (*β* =-0.87, *p*=0.19) but bolder (SSD x boldness: *β* =1.29, *p*=0.02). In mammals, the effect of SSD differed significantly for activity, aggression, boldness and exploration. When the sexes were the same size, there was only a sex difference in activity, with males being more active (*β* =0.27, *p*=0.05). When males were larger than females, males were significantly less active (*β* =-2.02, *p*<0.0001), but significantly more aggressive (SSD x aggression: *β* =3.45, *p*<0.0001), significantly bolder (SSD x boldness: *β* =1.96, *p*=0.0001) and significantly more exploratory (SSD x exploration: *β* =1.92, *p=*0.0002) than females.

There were no significant relationships between SSD and sex-differences in personality trait variability in any of the four taxa (see Table 3 and Supplementary Table S4). There was also no sex difference in variability when the sexes were the same size.

*Moved mating system to supplementary materials – I can easily move it back if we want*

*Publication bias*

We found evidence of publication bias for estimates of sex differences in mean or variances for personality traits in only one of 10 tests. The precision of the study predicted the effect size for the sex difference in the mean for invertebrates (p = 0.004; Supplementary Table S12). The test remained significant after correcting for multiple testing (critical Bonferroni p-value of 0.05/10 = 0.005).

**Discussion**

Surprisingly, our meta-analysis failed to find significant sex differences in either trait means or trait variability for personality-like behaviours. When personalities were divided into their respective trait type (the Big Five) we found several significant sex differences in means for some personality types within some taxonomic groups (birds: Sociality; reptilia: Exploration; invertebrates: Aggression, Boldness). However, only exploratory behaviour for birds and aggressive behaviour for fish had significant sex differences in variability, both skewing towards females. Finally, personality and SSD interacted significantly to show strong sex-bias for mean personality, yet only within mammals (male-bias for all personality types except activity, which had a female-bias) and invertebrates (male-biased aggression), and did not moderate sex differences in variability for any personality type or taxonomic group.

The sex chromosome hypothesis suggests that variability for shared traits should reflect sex chromosomal arrangement. While we did observe a tendency for mean variability estimates to skew towards the heterogametic sex (especially mammals and birds), most estimates were nonsignificant. This was surprising considering the number of studies that suggest heterogamety is the mechanism driving greater male variability. For example, greater male than female variability has been reported for traits like personality types in humans (Archer & Mehdikhani 2003; Karwowski *et al.* 2016), cognition and intelligence (Halpern & LaMay 2000; Jones *et al.* 2003; Arden & Plomin 2006; Johnson *et al.* 2008, 2009), and for morphological traits like body size (Reinhold & Engqvist 2013), and brain structure (Arnold 2004; van der Linden *et al.* 2017; DeCasien *et al.* 2020). Importantly, these traits are also considered important for reproduction; in men, extraversion and creative personality types attract more mating opportunities (Buss 1995; Nettle 2006), while body size is important for male-male competition (Darwin 1871). Therefore, we should expect traits related to reproduction to have greater male variability, not just because of sex chromosomal arrangement.

Despite suggestions that most shared traits should have greater male variability, our meta-analysis found no evidence for male-biased variability for any personality trait. We expected aggression to show strong male bias within all taxonomic groups for mean differences and variability, yet only invertebrates had significant sex differences in mean aggression, and only fish had significant sex differences in variability for aggression (female-biased). These results were surprising given that male aggression should correlate with male-male competition and sexual selection (e.g. Schuett *et al*. 2010). However, female aggression is common for establishing social hierarchies and for defending offspring (e.g. Campbell 1999). Although sex roles can vary across the animal kingdom, females generally have the greatest investment in offspring (via anisogamy, Janicke *et al.* 2016), so females who provide parental care should be more aggressive than males in this context. Parental care could also potentially explain why only invertebrate males were significantly more aggressive than females; in many species, females don’t often provide care beyond egg-laying. However, we lacked suitable data to test this idea.

An additional explanation for the absence of sex differences comes from the human personality literature. Creativity and aggressive personality types have greater male variability and are important for female mate choice. Yet other personality traits frequently show no sex differences in either trait means or their variability. Archer and Mehdikhani (2003) compared means and variances for males and females for a range of personality-like traits and found that traits related to sexual selection (directly like physical aggression, or indirectly through a change in reproductive strategy) consistently had significant greater male variability, while traits unrelated to sexual selection had no sex differences in variability (Archer & Mehdikhani 2003). Our chosen sexual selection moderators (SSD and mating system) did not significantly change either mean trait expression or variability; while SSD did interact with personality trait type for mean trait expression in mammals, SSD had no significant effect on variability for any taxonomic group. If sexual selection operates to maintain inter-individual variation in animal personalities (Schuett *et al.* 2010), we would then expect greater variability for those traits in the sex under sexual selection. Since we did not observe significant sex differences in variability, animal personalities could instead reflect behaviours that don’t differentially affect reproductive outcomes.

Personality traits are frequently measured under the context of survival; personality encompasses antipredator responses (‘risky’ behaviour, or Boldness), foraging (Activity/Exploration), and other behaviours related to survival that are important for both sexes. Indeed, we found no sex differences in activity and only one instance of male-biased boldness behaviour for invertebrates. Where we did find sex differences in personality, the traits and their direction tended to reflect life-history differences between the sexes. For example, male reptiles were more explorative than females which might indicate male-biased dispersal (e.g. salamanders *Plethodon cinereus*; Liebgold *et al.* 2011), while female birds were more sociable than males that reflects different social group behaviour (e.g. zebra finches *Taeniopygia guttata*; Schuett & Dall 2009). Additionally, because there are costs (e.g. energetic investment) and benefits (e.g. faster growth, more feeding opportunities) associated with the expression of personality types (Wolf *et al.* 2007), both sexes likely experience trade-offs that lead to similar trait means and variabilities. Previous meta-analyses have found evidence of trade-offs between personality and survival (Smith & Blumstein 2008b; Moiron *et al.* 2020), and we found some evidence that the degree of SSD might impact trade-offs between personality and survival for the sexes differently. For example, strong interactions between SSD and personality trait types for mammals, and between boldness and SSD for invertebrates, show that species with larger males trade-off ‘shy’ behaviour for larger body size. Larger animals are less likely to suffer mortality from predation, but also require more food to fuel their bigger bodies, thus need to feed more often and cannot play it safe. In mammals, as males became larger than females, males became bolder, more explorative and more aggressive than females. Male mammals are quite often the larger sex, due to sexual selection, so as males become larger they likely face more intense male-male competition (Darwin 1871)). However, SSD encompasses several different forms of selection, not just sexual selection (Blanckenhorn 2005). Because SSD did not interact significantly with trait variability, the sex differences in mean personality expression could instead be explained by adaptive, life-history differences.

Finally, phylogenetic variance (*I*2phylo) explained a large proportion of variance for personality variability (*lnCVR*) in both birds and mammals. This would suggest that variability in personality traits are heritable in endotherms and not ectotherms (Supplementary Table S1). Personality behaviours are heritable (Dochtermann *et al.* 2014), so assortative mating or mate choice for similar personality types could potentially allow variation in personality traits to persist, while keeping means and their variances relatively similar for both males and females (Schuett *et al.* 2010; Munson *et al.* 2020).

**Materials and Methods**

*Literature search and data collection*

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 S?. 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 one 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 in 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 LH. 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 and variances (i.e. standard error or standard deviation) for behavioural measures of personality for both males and females. 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 where individuals were neutered, 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). 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. We therefore included all behaviours that are commonly described as being indices of animal personality (see Sih *et al.* 2004; Réale *et al.* 2007).

In total, we identified 211 eligible articles that provided us with suitable data to calculate effect sizes (Fig. 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 file (see Supplementary Material).

*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. 1*

,

*Eq. 2*

*Eq. 3*

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. Hedges’ *d* is corrected by multiplying by the correction factor *J* to obtain *g*. The associated sampling error variance of Hedges’ *g* is:

*Eq. 4*

*Eq. 5*

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 indicted 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). We used this effect size rather than the lnVR to control for any mean-variance relationships and quantify sex difference in variances independent of the mean (Nakagawa *et al.* 2015; see Senior et al. 2020). In our data set there was a positive relationship between the mean and variance in personality measurements (see XXX). We calculated the effect size (lnCVR) and its associated samping variance () as follows:

,

*Eq. 6*

*Eq. 7*

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.

*Data transformations*

Our dataset contained some means, and their variances, that needed 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. 8*

*Eq. 9*

Second, some behavioural measures were presented as proportions, which constrains their distribution, so we converted them to the logit scale log (p/[1-p]) 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.

*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.* 2018). 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 trait (e.g. wing length). We then used the following index of SSD (Lovich & Gibbons 1992):

*Eq. 10*

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. We then used data from located studies to calculate the SSD index. Second, we noted whether the mating system of the species was characterised by ‘multiple mating’ or ‘monogamy’. We used the same search approach described above if the relevant data was not provided in the focal article, using the search string: “species name” AND mating system. The location of data collected for SSD index, mating system and parental care 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 S7-S12, 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) intercept models (only considering 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 reptilia (amphibians and reptiles were 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. In addition, we could only construct well-resolved phylogenies for some taxa.

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; Schuett *et al.* 2010). More importantly, this analysis provides a baseline that allows us 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 observation level identity to account for variation among effect sizes within a study to ensure that the total and residual variance are correctly estimated. Without this term, within-study effects are confounded with sampling variance (see Nakagawa *et al.* 2017). To correct for the non-independence of species due to their shared evolutionary history we also 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 that arising due to 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 factor to provide an estimate of the mean effect size for each trait (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. Interaction terms for personality traits when there were fewer than 10 species with SSD data were dropped from the models. There were too few data on reptilia to run a model that included any interactions (Table 3).

Studies included in our meta-analysis varied greatly in their design and there was 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*

It is possible that 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 small sample sizes (e.g. low precision) drive funnel plot asymmetry if there is a true non-zero effect and a bias towards publishing significant results (REF). Simple inspection of funnel plots is misleading, however, as we need to account for source of variation in effect sizes (i.e. moderator variables and random factors) other than sample size. 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. 11*

Where is the sampling variance of Hedge’s *g* (Eq. 4-5).

**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.

Blanckenhorn, W.U. (2005). Behavioral causes and consequences of sexual size dimorphism. *Ethology*, 111, 977–1016.

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.

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., Kazem, A.J.N., Réale, D. & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.*, 25, 81–89.

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

Dochtermann, N.A., Schwab, T. & Sih, A. (2014). The contribution of additive genetic variation to personality variation: Heritability of personality. *Proc. R. Soc. B Biol. Sci.*, 282.

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.

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.

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.

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.

Jennions, M.D. & Petrie, M. (1997). Variation in mate choice and mating preferences: A review of causes and consequences. *Biol. Rev.*, 72, 283–327.

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.

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.

Liebgold, E.B., Brodie, E.D. & Cabe, P.R. (2011). Female philopatry and male-biased dispersal in a direct-developing salamander, Plethodon cinereus. *Mol. Ecol.*, 20, 249–257.

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.

Moiron, M., Laskowski, K.L. & Niemelä, P.T. (2020). Individual differences in behaviour explain variation in survival: a meta-analysis. *Ecol. Lett.*, 23, 399–408.

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.

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.

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

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. (2013). The variability is in the sex chromosomes. *Evolution (N. Y).*, 67, 3662–3668.

Schuett, W. & Dall, S.R.X. (2009). Sex differences, social context and personality in zebra finches, Taeniopygia guttata. *Anim. Behav.*, 77, 1041–1050.

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

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

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

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

Tarka, M., Guenther, A., Niemelä, P.T., Nakagawa, S. & Noble, D.W.A. (2018). 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.

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. & 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.

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** | **Estimated mean** | **95% CI** | **Prediction intervals** | ***t* score** | ***p*-value** | ***n* effect sizes** |
| *SMD (Hedge’s g)* |  |  |  |  |  |  |
| Mammals | 0.08 | -0.28, 0.45 | -1.10, 1.27 | 0.46 | 0.65 | 684 |
| Birds | -0.13 | -0.36, 0.10 | -1.85, 1.59 | -1.12 | 0.26 | 493 |
| Reptiles/Amphibians | 0.07 | -0.08, 0.22 | -0.52, 0.67 | 0.94 | 0.35 | 95 |
| Fish | -0.04 | -0.35, 0.28 | -1.00, 0.93 | -0.24 | 0.28 | 493 |
| Invertebrates | 0.24 | -0.03, 0.51 | -1.76, 2.24 | 1.73 | 0.08 | 483 |
| *lnCVR* |  |  |  |  |  |  |
| Mammals | 0.07 | -0.20, 0.35 | -0.64, 0.79 | 0.54 | 0.59 | 684 |
| Birds | -0.15 | -0.66, 0.36 | -1.95, 1.65 | -0.59 | 0.56 | 493 |
| Reptiles/Amphibians | 0.05 | -0.04, 0.14 | -0.06, 0.15 | 1.13 | 0.26 | 95 |
| Fish | -0.04 | -0.09, 0.01 | -0.64, 0.56 | -1.44 | 0.15 | 493 |
| Invertebrates | -0.02 | -0.09, 0.06 | -0.72, 0.69 | -0.44 | 0.66 | 483 |

Table 2. MLMR models for each taxa with personality type as a fixed moderator. Values highlighted in bold indicate a significant difference 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*** |
| ***Mammals*** | | | | | | |
| Activity | -0.16 | -0.56, 0.23 | 0.41 | 0.11 | -0.19, 0.41 | 0.48 |
| Aggression | 0.11 | -0.27, 0.48 | 0.58 | 0.11 | -0.20, 0.42 | 0.48 |
| Boldness | 0.15 | -0.20, 0.50 | 0.39 | 0.06 | -0.21, 0.34 | 0.65 |
| Exploration | 0.06 | -0.31, 0.42 | 0.77 | 0.05 | -0.24, 0.34 | 0.73 |
| Sociality | 0.09 | -0.29, 0.47 | 0.64 | 0.06 | -0.24, 0.37 | 0.68 |
| ***Birds*** | | | | | | |
| Activity | -0.13 | -0.42, 0.16 | 0.37 | 0.04 | -0.25, 0.33 | 0.79 |
| Aggression | -0.11 | -0.40, 0.18 | 0.47 | -0.06 | -0.39, 0.27 | 0.72 |
| Boldness | -0.19 | -0.44, 0.07 | 0.15 | -0.01 | -0.23, 0.21 | 0.93 |
| **Exploration** | 0.08 | -0.19, 0.35 | 0.56 | **-0.26** | **-0.52, 0.00** | **0.05** |
| **Sociality** | **-0.68** | **-1.16, 0.21** | **0.005** | 0.13 | -1.59, 1.86 | 0.61 |
| ***Reptiles/Amphibians*** | | | | | | |
| Activity | -0.05 | -0.45, 0.36 | 0.82 | -0.11 | -0.51, 0.29 | 0.60 |
| Aggression | -0.07 | -0.32, 0.19 | 0.60 | 0.33 | -0.05, 0.72 | 0.09 |
| Boldness | 0.08 | -0.15, 0.31 | 0.68 | 0.10 | -0.13, 0.33 | 0.41 |
| **Exploration** | **0.25** | **0.05, 0.45** | **0.02** | -0.10 | -0.35, 0.14 | 0.40 |
| Sociality | -0.05 | -0.60, 0.50 | 0.86 | -0.12 | -0.76, 0.52 | 0.70 |
| ***Fish*** | | | | | | |
| Activity | -0.16 | -0.65, 0.33 | 0.53 | -0.03 | -0.16, 0.09 | 0.63 |
| **Aggression** | -0.05 | -0.52, 0.42 | 0.53 | **-0.13** | **-0.25, -0.01** | **0.04** |
| Boldness | -0.16 | -0.63, 0.32 | 0.52 | -0.02 | -0.63, 0.59 | 0.59 |
| Exploration | -0.05 | -0.54, 0.44 | 0.84 | -0.03 | -0.16, 0.09 | 0.62 |
| Sociality | -0.40 | -0.91, 0.11 | 0.12 | 0.07 | -0.11, 0.24 | 0.47 |
| ***Invertebrates*** | | | | | | |
| Activity | 0.08 | -0.23, 0.39 | 0.62 | -0.06 | -0.17, 0.05 | 0.26 |
| **Aggression** | **0.45** | **0.00, 0.90** | **0.05** | 0.02 | -0.17, 0.21 | 0.83 |
| **Boldness** | **0.32** | **0.01, 0.62** | **0.04** | -0.02 | -0.12, 0.08 | 0.65 |
| Exploration | 0.06 | -0.31, 0.43 | 0.74 | 0.05 | -0.10, 0.20 | 0.51 |
| Sociality | 0.30 | -0.26, 0.86 | 0.29 | 0.23 | -0.08, 0.54 | 0.15 |

Table 3. MLMR models with personality trait type and SSD moderator terms, as well as their interactions, for each taxonomic group. 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.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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** | **0.27** | **-0.01, 0.54** | **0.05** | 0.07 | -0.22, 0.37 | 0.63 | 14 | 12 | 84 |
| Aggression | 0.15 | -0.06, 0.37 | 0.17 | 0.10 | -0.18, 0.39 | 0.48 | 15 | 13 | 85 |
| Boldness | 0.13 | -0.07, 0.33 | 0.20 | 0.10 | -0.18, 0.39 | 0.48 | 26 | 26 | 163 |
| Exploration | 0.03 | -0.15, 0.22 | 0.72 | 0.04 | -0.24, 0.32 | 0.78 | 19 | 16 | 223 |
| **SSD** | **-2.02** | **-2.94, -1.10** | **<0.0001** | 0.30 | -0.34, 0.94 | 0.36 |  |  |  |
| **Aggression x SSD** | **3.45** | **2.05, 4.85** | **<0.0001** | -0.15 | -1.14, 0.83 | 0.76 |  |  |  |
| **Boldness x SSD** | **1.96** | **0.96, 2.96** | **0.0001** | -0.30 | -1.02, 0.41 | 0.41 |  |  |  |
| **Exploration x SSD** | **1.92** | **0.91, 2.92** | **0.0002** | -0.07 | -0.80, 0.65 | 0.84 |  |  |  |
| ***Birds*** | | | | | | |  |  |  |
| Intercept (Boldness) | -0.25 | -0.85, 0.34 | 0.40 | 0.03 | -0.04, 0.11 | 0.38 | 21 | 78 | 235 |
| SSD | -0.24 | -1.45, 0.97 | 0.70 | 0.10 | -0.16, 0.37 | 0.44 |  |  |  |
| ***Fish*** |  |  |  |  |  |  |  |  |  |
| Aggression | -0.09 | -0.66, 0.47 | 0.75 | -0.11 | -0.31, 0.09 | 0.27 | 16 | 13 | 93 |
| Boldness | -0.13 | -0.70, 0.44 | 0.66 | -0.03 | -0.22, 0.16 | 0.74 | 23 | 12 | 172 |
| SSD | 0.16 | -0.85, 1.17 | 0.76 | -0.28 | -1.03, 0.48 | 0.47 |  |  |  |
| Boldness x SSD | -0.38 | -1.45, 0.69 | 0.48 | 0.39 | -0.44, 1.21 | 0.36 |  |  |  |
| ***Invertebrates*** |  |  |  |  |  |  |  |  |  |
| Activity | 0.03 | -0.37, 0.42 | 0.90 | -0.08 | -0.22, 0.06 | 0.28 | 22 | 19 | 204 |
| **Boldness** | **0.37** | **-0.02, 0.76** | **0.06** | -0.02 | -0.15, 0.11 | 0.79 | 27 | 26 | 178 |
| SSD | -0.87 | -2.15, 0.42 | 0.19 | 0.05 | -0.56, 0.66 | 0.87 |  |  |  |
| **Boldness x SSD** | **1.29** | **0.22, 2.36** | **0.02** | 0.07 | -0.62, 0.76 | 0.84 |  |  |  |

Table 4.Primary and secondary keyword searches used for our literature search on 11 December 2018. Keyword searches differed slightly to account for how operator terms are employed by each database. The searches were further refined by using relevant biology field categories in ISI or SCOPUS.

|  |  |  |
| --- | --- | --- |
|  | ***ISI Web of Science*** | ***SCOPUS*** |
| *Primary keyword search terms* | (personalit\* OR “behavioural syndrome\*” OR “behavioral syndrome\*” OR temperament) AND (sex\*) NOT (man OR men OR woman OR women OR human) | personalit\* OR “behavioural syndrome” OR “behavioral syndrome” OR temperament AND sex AND NOT man AND NOT woman AND NOT human |
| *Secondary keyword search terms* | animal\* AND behav\* AND (bold\* OR shy\* OR neoph\* OR aggress\* OR explor\* OR hid\*) AND “sex differences” NOT man NOT woman NOT human | animal\* AND (bold\* OR shy\* OR neoph\* OR aggress\* OR explor\* OR hid\*) AND “sex differences” AND NOT man AND NOT woman AND NOT human |
| *Refined by categories* | zoology, ecology, biology, multidisciplinary sciences, evolutionary biology | agriculture & biology |



**Figure 1**. Phylogenetic relationships and orchard plots for a) mammals and b) birds. Node colours reflect SSD ratios for each species (no data is grey) and heatmap data shows *n* effect sizes for each personality trait type for each species. Orchard plots show all effect sizes and mean effect size for SMD and lnCVR MLMR 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 bigger circles have greater precision.



**Figure 2.** Phylogenetic relationships and orchard plots for c) reptilia, d) fish and e) invertebrates. Node colours reflect SSD ratios for each species (no data is grey) and heatmap data shows *n* effect sizes for each personality trait type for each species. Orchard plots show all effect sizes and mean effect size for SMD and lnCVR MLMR 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 bigger circles have greater precision.



**Figure 3.** *PRISMA* diagram showing the process of finding, screening and including/excluding studies for this meta-analysis. Searches were further refined to agriculture & biology (Scopus), zoology, ecology, biology, multidisciplinary sciences, evolutionary biology (WoS) categories.