A meta-analysis of sex differences in animal personality: no evidence for the greater male variability hypothesis

Lauren M. Harrison*, Daniel W. A. Noble† and Michael D. Jennions†

Division of Ecology and Evolution, Research School of Biology, The Australian National University, 46 Sullivans Creek Road, Canberra, ACT, 2600, Australia

ABSTRACT

The notion that men are more variable than women has become embedded into scientific thinking. For mental traits like personality, greater male variability has been partly attributed to biology, underpinned by claims that there is generally greater variation among males than females in non-human animals due to stronger sexual selection on males. However, evidence for greater male variability is limited to morphological traits, and there is little information regarding sex differences in personality-like behaviours for non-human animals. Here, we meta-analysed sex differences in means and variances for over 2100 effects (204 studies) from 220 species (covering five broad taxonomic groups) across five personality traits: boldness, aggression, activity, sociality and exploration. We also tested if sexual size dimorphism, a proxy for sex-specific sexual selection, explains variation in the magnitude of sex differences in personality. We found no significant differences in personality between the sexes. In addition, sexual size dimorphism did not explain variation in the magnitude of the observed sex differences in the mean or variance in personality for any taxonomic group. In sum, we find no evidence for widespread sex differences in variability in non-human animal personality.

Key words: sexual selection, personality, behaviour, sex differences, variability, shared traits, meta-analysis, sexual size dimorphism, heterogamety, greater male variability hypothesis

CONTENTS

I. Introduction ........................................................................... 680
(1) ‘Greater male variability’ in humans and other animals ........................... 680
(2) Personality behaviours and sex differences ........................................... 681
(3) Evolutionary explanations ..................................................................... 681
   (a) Condition dependence, life-history trade-offs and sexual selection ........ 681
   (b) Negative frequency dependence and sexual selection .......................... 682
   (c) Genetic and developmental pathways .............................................. 682
(4) Testing the ‘greater male variability’ hypothesis using non-human animal personalities ......................... 682
(5) Predictions ...................................................................................... 683
II. Materials and methods .............................................................. 683
(1) General approach .............................................................................. 683
(2) Literature search and data collection ..................................................... 683
(3) Data transformations .......................................................................... 684
(4) Effect sizes and sampling variances ...................................................... 685
(5) Moderator variables ........................................................................... 685
(6) Meta-analyses .................................................................................. 686
(7) Publication bias ................................................................................ 688

* Address for correspondence (Tel: +61 422 366 788; E-mail: lauren.harrison@anu.edu.au)
† Authors contributed equally to this study.
I. 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 human sex differences in variability among individuals (but see Lehre et al., 2009). The causes and consequences of sex differences in variability have, however, been widely debated in the social sciences (e.g. Stewart-Williams & Halsey, 2021). For example, sex differences in variability in academic performance (Machin & Pekkarinen, 2008) or the classroom (Baye & Monseur, 2016; O’Dea et al., 2018; Gray et al., 2019) have been attributed to sex differences in variability in intelligence (Arden & Plomin, 2006; Johnson, Deary & Carothers, 2008), psychological state (Archer, 2019), creativity (Ju, Duan & You, 2015), aggressiveness (Deary et al., 2003), personality (Borkenau, McCrae & Terracciano, 2013b) and ultimately, brain structure (Ritchie et al., 2018; Wierenga et al., 2020).

The general conclusion is that males are more behaviourally variable than females (‘greater male variability’ hypothesis). This conclusion has led to much controversy as it is sometimes used to explain male-biases in science, technology, engineering and mathematics (STEM) for university enrolment, subsequent employment and recognition of extremely high-level performance (Benbow & Stanley, 1980, 1983; Lubinski & Benbow, 1992; Pinker, 2005; Halpern et al., 2007).

(1) ‘Greater male variability’ in humans and other animals

Greater attention has been paid to describing average differences between the sexes than to differences in trait variability (e.g. average differences in brain size; see Shields, 1975, 1982). It was not until Darwin (1871) suggested that males were more likely than females to express variation in their physical (i.e. phenotypic) traits that the ‘greater male variability’ hypothesis became the focus of scientific attention. Darwin (1871, p. 272, 321–330) argued that the tendency of males to vary more than females indicated that sex-specific selection on males was an important evolutionary process. As evolutionary theory emphasizes the importance of variability as the raw material for selection to operate, it became legitimate to study sex differences in behavioural variation (Shields, 1982; Hyde, 1990). Following Darwin, the prominent sexologist Havelock Ellis (1894) was one of the first formally to describe biological sex differences in variability for both physical and mental traits in men and women. Ellis (1894) concluded that the differences between men and women for variability in brain size indicated that the mental abilities of men were superior to those of women, due to a greater concentration of men at the higher extremes of distributions for both traits. Although the ‘greater male variability’ hypothesis has faced criticism and pushback since its inception (most notably from women scientists (Thompson, 1903; Hollingworth, 1914, 1918; Hyde, 1981; Shields, 1982)), it has remained a controversial topic of unresolved debate for well over a century (see Fausto-Sterling, 1985). Some have argued that the extent to which the ‘greater male variability’ hypothesis has been accepted is related more to societal changes than empirical evidence. For example, Shields (1982) stated that the hypothesis was more readily accepted by the scientific community at a time when women began to enrol in universities in increasing numbers. Even now, as fewer scientists invoke the ‘greater male variability’ hypothesis to explain a male-bias in STEM, the hypothesis, and its possible biological underpinnings (i.e. due to natural and sexual selection – referred to using the shorthand of “biology” or “evolution”) continue to be discussed and tested (e.g. Geary, 2018, 2021).

Greater variability among men than women in behaviour, personality and cognition is widely attributed to socio-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 evolved sex differences in behaviour is still underplayed (Archer, 2019; Stewart-Williams & Halsey, 2021). A key line of reasoning invokes a trend across non-human animals for greater male than female variability (e.g. Geary, 2010). Although greater male variability has been reported for some traits in non-human animals (e.g. reversal learning performance in mountain chickadees; Branch et al., 2020), the robustness of this claim is unclear. Specifically, the strongest evidence is for greater variability in male than female morphology, especially for sexually selected traits, including ornaments, weaponry and body size (Pomiankowski & Moller, 1995; Reinhold & Engqvist, 2013; Wyman & Rowe, 2014). To date, the evidence appears weak or absent for greater male
variability for behaviours that are exhibited by both sexes (Tarka et al., 2018; Zajitschek et al., 2020).

(2) Personality behaviours and sex differences

Human personality is often quantified by scoring five components of behaviour (the ‘Big Five’: extraversion, neuroticism, openness, conscientiousness and agreeableness; Costa & McCrae, 1992), which are repeatable among individuals over time. Over the last 15 years, comparable evidence has emerged for repeatable and heritable behavioural variation among non-human animals that is akin to human personality. These animal behaviours are often grouped into five personality axes that loosely resemble those in humans: activity, aggression, boldness, exploration and sociality/sociability (Réale et al., 2007). We can therefore now test claims for greater male variability in personality traits in non-human animals (e.g. Archer, 1996; Nettle, 2006; Borkenau et al., 2013b).

(3) Evolutionary explanations

There are three non-mutually exclusive explanations frequently invoked by biologists to explain the maintenance of variation in non-human animal personality despite natural selection tending to eliminate less-fit variants (Smith & Blumstein, 2008; Dingemanse & Wolf, 2010): (a) condition-dependent trait expression; (b) negative frequency-dependent selection and (c) developmental or genetic constraints. Crucially, when stronger sexual selection on males than females is taken into account, sexual selection might partially account for why males are more variable than females for some personality traits.

Sexual selection is usually stronger on males than females (Fromhage & Jennions, 2016; Janicke & Morrow, 2018). It favours individuals with the most extreme expression of traits that increase the likelihood of obtaining mates (e.g. weapons, ornaments, courtship and coercive behaviours; Darwin, 1871), or fertilising eggs when females mate multiply (Parker, 1970). There is often far greater among-individual variation in mating success in males than females (Janicke et al., 2016). This is partly attributable to the relative abundance of males and females that are available to mate [i.e. the operational sex ratio (OSR); Emlen & Oring, 1977]. As the OSR becomes more male-biased, the variation in male mating success increases, and hence the opportunity for sexual selection (Wade, 1979; Arnold & Wade, 1984). However, a greater opportunity for sexual selection in males than females does not, by itself, mean that there is stronger sexual selection on males: stochastic factors can create a relationship between the OSR and the opportunity for sexual selection (Klug et al., 2010; Jennions, Kokko & Klug, 2012). Ultimately, stronger sexual selection on males primarily arises due to males spending relatively more of their lives in the mating pool trying to acquire a mate (Kokko, Klug & Jennions, 2012; Fromhage & Jennions, 2016). This increases the costs that males can incur to shorten this interval by investing more heavily in traits that provide an advantage when competing for mates. This sexual competition for mates (and fertilisations) tends to generate strong directional selection on males which, in turn, can sometimes select for condition-dependent dependent trait expression, and alter the life-history trade-offs that males and females experience. Each of these consequences can increase variability in male mating behaviours.

(a) Condition dependence, life-history trade-offs and sexual selection

Sexual selection generally favours condition-dependent trait expression if individuals in better condition can afford to incur greater costs (Rowe & Houle, 1996). Even minor differences in resource acquisition among individuals due to chance or small genetic differences in, say, foraging efficiency can therefore translate into differential expression of sexual traits (Rowe & Houle, 1996), increasing phenotypic variance among individuals. Condition dependence can therefore maintain variation in sexually selected traits, even when directional female choice should otherwise erode genetic variation in male trait expression (i.e. the ‘lek paradox’; Borgia, 1979). Sexually selected traits are likely to show higher condition dependence than naturally selected traits because a small relative advantage can translate into large fitness gains due to the zero-sum nature of competition for mates. In sum, stronger sexual selection on males is predicted to yield sexual traits with greater variance among males than equivalent traits in females, or than naturally selected traits in either sex (Pomiankowski & Moller, 1995; Wyman & Rowe, 2014). It is also worth noting that pleiotropic effects of sexually selected traits might then also increase variance in correlated traits (e.g. Han & Dingemanse, 2015). For example, sexual selection on males to elevate aggressiveness when fighting with rivals might also affect aggressiveness in other contexts, such as when foraging in a group.

Life-history trade-offs have also been proposed to generate variation in personality among individuals (Wolf et al., 2007). Sexual selection on condition-dependent male sexual traits can lead to sex-specific life-history trade-offs that generate greater phenotypic variation among males than females. For example, sexual selection often reduces male life expectancy to below that of females due to the differential expression of sex and stress hormones (e.g. in humans; Regan & Partridge, 2013), fighting-associated injuries (e.g. fatal territorial contests in male common loons Gavia immer; Piper et al., 2008), the high energetic costs of sexual advertising (e.g. the advertisement calls of male field crickets Teleogryllus commodus; Hunt et al., 2004), and ornaments attracting predators (e.g. guppy Poecilia reticulata; male colouration; Gordon, Lopez-Sepulcre & Reznick, 2011). Sexual selection could therefore increase variation in the magnitude of the trade-off between current reproduction and survival, which will elevate age-dependent variation in allocation of resources to different life-history traits, including those under both sexual and natural selection.

Biological Reviews 97 (2022) 679–707 © 2021 Cambridge Philosophical Society
Negative frequency-dependent selection is a form of balancing selection that can maintain genetic variation in traits. It is often associated with the evolution of alternative morphs within a species due to a rarer phenotype having a fitness advantage (e.g. predators learning more readily to detect male P. reticulata guppies with a common colour morph; Olendorf et al., 2006). Negative frequency-dependent selection is particularly relevant for behavioural traits that elevate 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 limited food sources (Fitzpatrick et al., 2007).

Sexual selection often results in the evolution of alternative mating tactics. This is usually because individuals with a poor start in life, or those that are in worse condition due to their recent experience or to the effects of ageing, gain more from a mating tactic that circumvents female choice or direct physical contests (Taborsky & Brockmann, 2010). When early development affects the adult phenotype, this can result in development being canalised into a few discrete pathways such that alternative mating tactics are associated with a suite of morphological traits that differ from those for the dominant mate-acquisition tactic. For example, sneaker males are smaller and more female-like than males using the dominant mating tactic. Sexual selection, because it more often promotes the evolution of alternative mating tactics in males than females, therefore tends to generate higher variation among males than females in both morphological traits and allied behaviours.

Greater male variability might arise from ‘constraints’ imposed by sex-determination mechanisms (James, 1973). In mammals, for example, XX/XY chromosomal sex determination means that females have two X chromosomes and males only one. For females, the phenotypic effects of genes on the X chromosome are therefore averaged across their expression on both chromosomes, often via epigenetic inactivation of one chromosome (Amos-Landgraf et al., 2006). By contrast, males only express genes on the single, maternally inherited X chromosome. All else being equal, this should create more extreme phenotypes in males (Charlesworth, 1996), hence greater variation among males than females. It should be noted, however, that X-inactivation is not always random in mammals. Skewed X-inactivation, the imbalanced expression of paternal and maternal X chromosome genes, is fairly common (Shvetsova et al., 2019) and can generate greater than expected genetic variation in females (Gribnau & Barakat, 2017). Additionally, females are the heterogametic sex in some taxa, including birds, butterflies, and some fish and reptiles (Reukeboom & Perrin, 2014). If the mechanism of sex determination drives sex differences in phenotypic variability, then taxa with heterogametic females should exhibit greater variability in females than males (Reinhold & Engevist, 2013). Similarly, it is possible that when sex determination depends on an environmental threshold, such as temperature or host size in parasitoids, the sexes might differ in their phenotypic variability depending on whether there is greater variation in the environmental cue above or below the threshold.

(4) Testing the ‘greater male variability’ hypothesis using non-human animal personalities

Based on the above explanations, sexual selection and the mechanism of sex determination are likely to amplify both average differences in trait expression and variation in sexually selected behaviours and, as a by-product, any associated behaviours that affect personality (reviewed in Schuett, Tregenza & Dall, 2010). For example, aggressive personality, as a result of sexual selection on fighting behaviour, is expected to lead to both greater average trait expression and more among-individual variance in males compared to females (but see Stockley & Bro-Jørgensen, 2011). Indeed, variation in two components of human personality (agreeableness and neuroticism) that are linked to aggressiveness, show greater variability among men (Budaev, 1999; Archer & Mehlilkhani, 2003). This line of reasoning should apply to all taxa: male variability in personality should be higher in species when sexual selection on males is more intense. This broad-scale prediction about all non-human animals is directly relevant to proposed explanations for sex differences in variation in human behaviour that invoke evolved biological differences between the sexes due to sexual selection (e.g. Archer, 2019; Stewart-Williams & Halsey, 2021), but it remains poorly tested.

Here we conduct a large-scale meta-analysis of non-human animal personality studies (extending an earlier meta-analysis by Tarka et al., 2018) to test the robustness of claims that males differ behaviourally from females for these types of traits. More specifically, we answer three questions: (i) do males show greater mean trait expression or variability than females for each of the five personality components? (ii) Do sex differences in means or variability differ among the five components of personality? And (iii) does sexual selection [estimated using a proxy of sexual size dimorphism (SSD)] explain differences in the magnitude of sex differences in the mean and variance in personality? In many taxa, the strength of sexual selection on males is correlated with the magnitude of SSD (Fairbairn, Blanckenhorn & Székely, 2007). We therefore included SSD as a moderator in our meta-analyses. Additionally, our data set contains taxa where males are heterogametic (mammals), homogametic (birds), or a mixture of both (insects, fish and reptiles/amphibians). As such, we also partially tested if sex determination by sex chromosomal arrangement affects sex differences in variation in personality by testing for a moderating effect of taxa.
(5) Predictions

We expected that males would show larger mean 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). 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 sociality), life histories (e.g. sex-biased dispersal is likely to affect exploration and activity), 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). When testing whether the degree of sex-specific sexual selection (as measured by an index of SSD) 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 due to variation in the magnitude of the difference in sex-specific selection on the trait.

II. MATERIALS AND METHODS

(1) General approach

We conducted a systematic review and meta-analysis to test the generality of the ‘greater male variability’ hypothesis across the animal kingdom (excluding humans). Comparisons of shared behavioural traits often provide conflicting evidence for greater male variability (e.g. Reinhold & Engevist, 2013; Tarka et al., 2018; Zajitschek et al., 2020); we therefore chose to focus on animal personality as a way to compare shared behaviours that are broadly equivalent, and measured in similar ways, across many different non-human animal species. For this synthesis, we extracted the raw means and error for personality and personality-like behaviours for both males and females from the primary literature (see Fig. 1). Recently, there has been a push for stricter definitions of personalities in non-human animal behaviour studies (Dingemans & Wright, 2020). By definition, personality traits are repeatable (Sih, Bell & Johnson, 2004), but very few of the available studies reported repeatability, nor 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 adequate sample sizes, however, we retained behaviours that are commonly described as indices of animal personality, or where the authors explicitly interpreted these behaviours as ‘personalities’ (Sih et al., 2004; Réale et al., 2007). Our inclusion of studies that did not explicitly confirm the repeatability of personality and personality-like behaviours could introduce potential problems with our interpretation, but we included them to allow us to test the ‘greater male variability’ hypothesis for behaviours exhibited by both sexes across a wider array of animal taxa. It is also worth noting that there is no obvious reason for studies that do not report repeatability to be biased with respect to the magnitude or direction of any sex difference in the mean or variance in behaviour.

(2) Literature search and data collection

Our meta-analysis was pre-registered with the Open Science Foundation (OSF; study details available at: https://osf.io/b9ju6/). 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 key word searches using the search strings shown in Table S1 in the online Supporting Information. 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 (Sih et al., 2004; Réale et al., 2007). Our secondary searches were designed to locate studies of ‘personality-like’ behaviours that were alternative 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, Houston & McNamara, 2004; Sih et al., 2004; Réale et al., 2007; Dingemans & Wolf, 2010; Schuett et al., 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 9698 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. 1.

After removing obvious duplicates, 3739 articles were available for title and abstract screening by LMH. Of these, 942 were excluded at the title/abstract screening stage because they were: (i) duplicates (N = 130), (ii) human studies (N = 256), (iii) reviews or non-empirical studies (N = 516), (iv) not in English (N = 8), or (v) inaccessible (N = 32). We then carried out a full-text screening of the remaining 2797 articles. We included articles in the final data set 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 reported principle components or factor loadings were excluded (following the methods of Tarka et al., 2018). We excluded articles during
full-text screening using the following criteria: (i) 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’), (ii) the study was on domesticated or agricultural animals, (iii) the study did not report separate data for males and females (i.e. the study was on simultaneous 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 (iv) data were unsuitable/missing (i.e. raw data were missing, missing sample sizes, sampling error type was not reported, or the mean of one sex was zero leading to effect size calculation issues) (Fig. 1 and Appendix S1).

In total, we identified 209 eligible articles with suitable data to calculate effect sizes. 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 version 1.0.0 (Pick, Nakagawa & Noble, 2019) to extract summary statistics from figures. The location of the data in the original article is provided in our raw data file (see Datafile S1: data.xlsx). We received an additional $N=2$ unpublished data sets from colleagues (eastern grey kangaroos Macropus giganteus: W. Menario-Costa; white-winged choughs Cororax melanorhamphos: C. Leon) to bring the total number of eligible studies to 211.

From our final 211 eligible studies, a further $N=7$ invertebrate studies were removed from analysis because they used score data (i.e. ranked behaviours). We ran contrast-based meta-analytic models with score data as a categorical moderator to compare data calculated from scores (which could not be transformed; see below) to the rest of the data within each taxonomic group. Our model sensitivity checks found that effect sizes calculated from scores were significantly different from the rest of the invertebrate data set (but not for any other taxonomic group, see Table S2 and Appendix S1). Thus, we decided to remove these effect sizes, which reduced our final data set to $N=2162$ effect sizes, $N=204$ studies and $N=220$ species (see Datafile S1: data.xlsx).

(3) Data transformations

Our data set contained some means, and associated variances, that had to be transformed to meet distribution assumptions about normality before we could calculate the effect sizes Hedges’ $g$ or ln coefficient of variation (lnCVR) ($N=200$ latency samples; $N=74$ proportion samples). 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:

\[ \ln \bar{X} = \log(\bar{X}) - \log \left(1 + \frac{SD^2}{\bar{X}^2}\right), \]  
(1)

\[ \ln SD = \sqrt{\log \left(1 + \frac{SD^2}{\bar{X}^2}\right)}, \]  
(2)

where \( \bar{X} \) is the mean and \( SD \) is the standard deviation.

Second, some behavioural measures were presented as proportions, which constrains their distribution, so we converted them to the logit scale using \( \ln \left(\frac{p}{1-p}\right) \);

\[ SD = \sqrt{SD^2 \left(\frac{1}{p^2} + \frac{1}{(1-p)^2}\right)}, \]  
where \( p \) is the proportion, to meet normality assumptions before calculating the relevant effect size. Ignoring these assumptions can result in overestimated effect size estimates (which was the case for our data). Information on which effect sizes are based on transformed values are provided in the raw data file (Datafile S1: data.xlsx) supplied as Supporting Information.

(4) 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, SMD; Hedges & Olkin, 1985) as follows:

\[ g = \frac{\bar{X}_M - \bar{X}_F}{SD_p}, \]  
(3)

\[ SD_p = \sqrt{\frac{(N_M - 1)SD_M^2 + (N_F - 1)SD_F^2}{N_M + N_F - 2}}, \]  
(4)

\[ j = 1 - \frac{3}{4(N_M + N_F - 2) - 1} \]  
(5)

where \( \bar{X} \) is the mean of the behavioural measure, \( SD_p \) is the pooled standard deviation and \( N \) and \( SD \) are the sample size and standard deviation for males (M) and females (F). The associated sampling error variance of Hedges’ \( g \) is:

\[ V_g = \frac{\left(\frac{N_M + N_F}{N_M N_F}\right) + \frac{d^2}{(2(N_M + N_F)j^2)}}{N_M N_F} \]  
(6)

where \( d \) is Hedges’ \( g \) without the correction factor \( j \) (see Eq. 3).

We used Hedges’ \( g \) instead of log response ratios as some behavioural measures were based on ranks or scores (excluding invertebrates for which score data were removed). Thus 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 sex differences in the variance in personality traits we used the ln 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 (Pearson’s correlation: males: \( r = 0.90 \); females: \( r = 0.91 \)). Thus, 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, Viechtbauer & Nakagawa, 2020). We calculated lnCVR and its associated sampling variance (\( \sigma_{lnCVR}^2 \)) as:

\[ \ln CVR = \ln \left(\frac{CV_M}{CV_F}\right) + \frac{1}{2(N_M - 1)} - \frac{1}{2(N_F - 1)}, \]  
(7)

\[ \sigma_{lnCVR}^2 = \frac{SD_M^2}{N_M CV_M^2} + \frac{1}{2(N_M - 1)} - 2\rho_{\ln CV_M, \ln CV_F}\sqrt{\frac{SD_M^2}{N_M CV_M^2} + \frac{SD_F^2}{N_F CV_F^2}} \frac{1}{2(N_M - 1)} + \frac{SD_F^2}{N_F CV_F^2} + \frac{1}{2(N_F - 1)} - 2\rho_{\ln CV_M, \ln CV_F}\sqrt{\frac{SD_M^2}{N_M CV_M^2} + \frac{SD_F^2}{N_F CV_F^2}} \frac{1}{2(N_M - 1)} \]  
(8)

where \( CV_M \) and \( CV_F \) are \( SD_M / \bar{X}_M \) and \( SD_F / \bar{X}_F \), respectively; \( \rho_{\ln CV_M, \ln CV_F} \) are the correlation between the natural logarithms of means and standard deviations of males and females respectively. A positive value of lnCVR indicates that males are more variable than females.

(5) 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 laboratory (‘captive bred’) or the wild, whether the behaviours were measured in the laboratory or field, and whether the data were 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 SSD, which is often strongly correlated with indicators of the strength of sexual selection (Fairbairn et al., 2007). 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):

$$SSD\ index = \frac{X_{M\,\text{measured}}}{X_{F\,\text{measured}}} - 1$$  \hspace{1cm} (9)

Using the SSD index rather than the sex difference in raw means allows for: (i) a single continuous moderator that is zero when the sexes are the same size, and positive when males are larger than females; and (ii) 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. There are limitations to using SSD as a proxy for male-biased sexual selection because SSD reflects the combination of effects of natural and sexual selection on each sex. First, a female-biased SSD can be found in many animal groups (including birds, fish, reptiles/amphibians and invertebrates). However, a female-biased SSD can often be attributed to natural selection for larger female size (e.g. fecundity selection) rather than stronger sexual selection on males (e.g. Shine, 1989; Krüger, 2005). That noted, even if the SSD is female-biased within a given taxon, a smaller bias is often likely to reflect stronger sexual selection on males (Fairbairn et al., 2007). Second, a male-biased SSD does not reflect sexual selection due to sperm competition, which can arise when competition amongst males to fertilise eggs is greater than the opportunity to monopolise females via mate attraction, mate-guarding or male–male competition (Lüpold et al., 2014). 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 the CRC Handbook of Avian Body Masses (Dunning Jr, 2007) and the online reference database Birds of the World (Billemam et al., 2020) for body size measures. We then used data from the located studies to calculate the SSD index (see Dataset S2: sexual_selection.xlsx).

Second, we also quantified mating system. 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”. Initially we quantified mating system as either ‘monogamous’, ‘polygynous’, ‘polyandrous’, or ‘promiscuous’. However, there were too few species within some mating system categories to run contrasts for any taxonomic group. We therefore simplified mating system to ‘monogamous’ or ‘multiple mating’. The location of data collected for SSD index and mating system are provided in Dataset S2: sexual_selection.xlsx).

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 taxonomic group. The relationships between the other moderators and the effect sizes are presented in Tables S3–S7, but these should be treated as strictly exploratory analyses.

(6) Meta-analyses

We modelled the effect sizes Hedge’s g (also referred to as the standardised mean difference, SMD) 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: 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 taxonomic group as a moderator and would have resulted in an 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 separate phylogenies for each group. Even then, phylogenies were better resolved for some groups than others (e.g. mammal and bird phylogenies were better than those for invertebrates – 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 provides a baseline to investigate sources of heterogeneity in the data. To account for the non-independence of data we included species 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 result solely from sampling variance (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 taxonomic group either from existing phylogenetic tree databases (e.g. BirdTree.org for birds; Jetz et al., 2012) or using TimeTree.org (Kumar et al., 2017). For the bird phylogeny, we used the Ericson tree backbone (Ericson et al., 2006) to generate 1000 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
Table 1. Multi-level meta-analytic (MLMA) models for each taxonomic group for a sex difference in the mean (SMD) and variability (lnCVR) in personality traits. Overall, there were no significant sex differences for either mean personality or for variability. Positive estimates indicate that the mean or variability in personality is greater for males than females. We report false discovery rate adjusted P-values for all tests for SMD and lnCVR respectively (N = 48 tests each).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Overall mean</th>
<th>95% CIs</th>
<th>Prediction interval</th>
<th>t score</th>
<th>P-value (adjusted)</th>
<th>N effect sizes</th>
<th>N studies</th>
<th>N species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMD (Hedge’s g; mean)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammals</td>
<td>0.07</td>
<td>−0.30, 0.44</td>
<td>−1.14, 1.27</td>
<td>0.37</td>
<td>0.87</td>
<td>674</td>
<td>61</td>
<td>45</td>
</tr>
<tr>
<td>Birds</td>
<td>−0.11</td>
<td>−0.35, 0.13</td>
<td>−1.87, 1.64</td>
<td>−0.92</td>
<td>0.85</td>
<td>480</td>
<td>50</td>
<td>106</td>
</tr>
<tr>
<td>Reptiles/amphibians</td>
<td>0.06</td>
<td>−0.11, 0.24</td>
<td>−0.57, 0.70</td>
<td>0.71</td>
<td>0.85</td>
<td>95</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Fish</td>
<td>0.18</td>
<td>−0.23, 0.58</td>
<td>−1.55, 1.91</td>
<td>0.86</td>
<td>0.85</td>
<td>490</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>Invertebrates</td>
<td>0.25</td>
<td>−0.12, 0.62</td>
<td>−2.12, 2.63</td>
<td>1.35</td>
<td>0.85</td>
<td>423</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td><strong>lnCVR (variance)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammals</td>
<td>0.06</td>
<td>−0.21, 0.32</td>
<td>−0.66, 0.77</td>
<td>0.42</td>
<td>0.95</td>
<td>674</td>
<td>61</td>
<td>45</td>
</tr>
<tr>
<td>Birds</td>
<td>−0.14</td>
<td>−0.63, 0.36</td>
<td>−1.89, 1.61</td>
<td>−0.55</td>
<td>0.95</td>
<td>490</td>
<td>50</td>
<td>106</td>
</tr>
<tr>
<td>Reptiles/amphibians</td>
<td>0.04</td>
<td>−0.04, 0.12</td>
<td>−0.04, 0.12</td>
<td>0.93</td>
<td>0.95</td>
<td>95</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Fish</td>
<td>0.00</td>
<td>−0.09, 0.08</td>
<td>−0.75, 0.74</td>
<td>−0.11</td>
<td>0.95</td>
<td>490</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>Invertebrates</td>
<td>−0.01</td>
<td>−0.11, 0.09</td>
<td>−0.74, 0.72</td>
<td>−0.17</td>
<td>0.95</td>
<td>423</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>

Importing a list of the relevant species to TimeTree.org. If a species in our data set had no phylogenetic representation, we substituted the next closest available species (e.g. same genus or family; see Data file S1 for phylogenetic representation). In all cases, we resolved synonymous taxa across our data set so that species were correctly categorised and pruned our trees where needed. We used the ggplot package (Yu et al., 2016) in R to generate our final phylogenetic trees for each taxonomic group as shown in Figs 2–6. 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 (\(I^2_{\text{species}}\)), studies (\(I^2_{\text{study}}\)), and due to phylogeny (\(I^2_{\text{phylo}}\); Nakagawa & Santos, 2012). The total heterogeneity (\(I^2_{\text{Total}}\)) is the proportion of the total variance in effect size estimates excluding total sampling variance (see Appendix S1 for calculations).

Next, we fitted 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 axes (i.e. ‘activity’, ‘aggression’, ‘boldness’, ‘sociability’, ‘exploration’) (Table 2). These personality trait models are shown in Figs 2–6 using the archaRD R package (Nakagawa et al., 2021a). 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. Thus, we ran subset models for each personality trait type and included SSD. We only ran these subset 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 any models that included SSD (Table 3).

The studies included in our meta-analysis varied greatly in their design and there were several 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 of these sources of non-independence on our 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 \(\rho = 0.3\), \(\rho = 0.5\) and \(\rho = 0.8\). We refitted our models replacing the identity matrix (I) that was assumed when estimating the residual error variance with our D matrices (see Appendix S1 and Tables S8–S13). The results presented in Tables 1–3 are robust and not influenced by the level of non-independence due to correlation among effect sizes measured on the same individuals.

For all models we present meta-analytic mean estimates and 95% confidence intervals (Tables 1–3). To prevent potential Type I errors arising from the number of meta-analytic tests conducted, we applied the false discovery rate (FDR) method (Benjamini & Hochberg, 1995) to adjust P-values obtained for SMD and lnCVR models separately (\(N = 48\) tests). We therefore report the adjusted P-values throughout Section III (see Tables S14–S16 for unadjusted P-values). Historically, a focus on statistical significance has likely contributed to the idea that the ‘greater male variability’ hypothesis remains unresolved given the occurrence of studies reporting both significant and non-significant sex differences in variability. By contrast, a meta-analytic approach
that emphasises the mean effect size draws attention to the magnitude of the estimated difference, which if very small, is likely to have little biological meaning (discussed in Fausto-Sterling, 1985; Hyde, 2005). We therefore encourage readers to interpret mean effect size estimates and their associated confidence intervals.

(7) Publication bias

Published studies might disproportionately report certain findings (e.g. greater male variability). To investigate publication bias, we first checked for funnel plot asymmetry for both SMD and lnCVR when plotting the effect size against a measure of sampling error. While we cannot know how...
many studies are missing, we expect that low-powered studies (e.g. low precision and high sampling error) that show effects opposite to what is predicted are more likely to go unpublished. As such, the lack of studies meeting these criteria in a funnel plot is expected to drive funnel plot asymmetry (Sterne et al., 2011). Visual inspection of funnel plots can be misleading, however, as we need to account for additional sources of variation in effect sizes beyond effect size precision (i.e. moderator variables and random factors). In addition, when using standardised mean differences (SMD) such as Hedge’s $g$, the effect size can be correlated with the sampling error (SE) resulting in ‘artefactual’ funnel asymmetry (Nakagawa et al., 2021b). We therefore included the inverse square root of ‘effective sample size’ ($1/n_i$) as a moderator term in our MLMR models to test whether it explained some of the variation in the reported effect sizes (for methodological justification see Nakagawa et al., 2021b). If the inverse of effective sample size has a significant influence, this is suggestive of either

![Diagram](image)

**Fig. 3.** Sex differences in means and variability for the five personality traits for birds. The phylogeny depicts all bird species present in the meta-analysis. Node colours reflect sexual size dimorphism (SSD) for each species (male-biased are blue, female-biased are red, monomorphic species are white and no data are grey) and heatmap data (green bars) shows the number ($N$) of effect sizes for each personality trait type for each species. Orchard plots show total effect sizes ($k$) and mean effect size for (A) SMD (means) and (B) lnCVR (variance) meta-regression models with personality trait as a moderator term. Thick bars are 95% confidence intervals (CIs) and thin bars are prediction intervals. Circle size reflects effect size precision.
unbalanced sampling or publication bias (Nakagawa et al., 2021b). The inverse of $\hat{n}$ is calculated as $(\text{female } n + \text{male } n) / (\text{female } n \times \text{male } n)$ (see Nakagawa et al., 2021b).

### III. RESULTS

1. **Data set summary**

   Our final data set comprised 2162 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 taxonomic group are shown in Table 1. Boldness was the best studied, and sociality the least studied, of the five personality types ($N = 817$ and 165 effect sizes, respectively) (Table 2).

2. **Sex differences**

   Combining all five personality traits, mean effect sizes were generally small in magnitude and males and females did not differ significantly in their mean personality in any of the five taxonomic groups, nor was there a significant sex difference in variability (Table 1 and Table S14). The effect size estimates from our basic meta-analytic intercept models almost all had high heterogeneity ($\hat{I}^2_{\text{Total SMD}} > 0.60; \hat{I}^2_{\text{Total lnCVR}} > 0.60$), although there was moderate to low

---

**Fig. 4.** Sex differences in means and variability for the five personality traits for reptiles/amphibians. The phylogeny depicts all species used in the meta-analysis. Node colours reflect sexual size dimorphism (SSD) for each species (male-biased are blue, female-biased are red, monomorphic species are white and no data are grey) and heatmap data (green bars) shows the number ($N$) of effect sizes for each personality trait type for each species. Orchard plots show total effect sizes ($k$) and mean effect size for (A) SMD (means) and (B) lnCVR (variance) meta-regression models with personality trait as a moderator term. Thick bars are 95% confidence intervals (CIs) and thin bars are prediction intervals. Circle size reflects effect size precision, where larger circles have greater precision.
heterogeneity for reptiles/amphibians ($I^2_{\text{Total SMD}} = 0.47$; $I^2_{\text{Total lnCVR}} = 0.00$) and fish ($I^2_{\text{Total lnCVR}} = 0.59$) (Table S14). Heterogeneity in the sex difference in mean personality mostly came from between-study differences ($I^2_{\text{study}}$), while phylogenetic relationships and among-species differences ($I^2_{\text{phylo}}$ and $I^2_{\text{species}}$, respectively) explained heterogeneity in the variability of effect sizes for mammals, birds and reptiles/amphibians only (Table S14).

The lack of a sex difference in mean and variability in personality when examining all five personality traits simultaneously could arise if the direction of any difference in sex-specific values varied across the traits. When the traits were analysed separately and by taxonomic group, we found two effect size differences in mean values whose $95\%$ confidence intervals only just overlapped or did not overlap zero. Specifically, females were more sociable than males for birds, and males were more exploratory than females for reptiles (Table 2). Although these estimated effect sizes suggest that sex differences might exist for personality traits, none of the 25 taxon-specific tests for sex differences in mean personality were statistically significant after taking into account the false discovery rate (Table 2, Figs 2–6).

Similarly, effect sizes were generally moderate to small with no significant sex differences in variability when the five personality traits were analysed separately across taxa. We found four effect size differences in mean values whose $95\%$ confidence intervals either only just overlapped

Fig. 5. Sex differences in means and variability for the five personality traits for fish. The phylogeny depicts all species used in the meta-analysis. Node colours reflect sexual size dimorphism (SSD) for each species (male-biased are blue, female-biased are red, monomorphic species are white and no data are grey) and heatmap data (green bars) shows the number ($N$) of effect sizes for each personality trait type for each species. Orchard plots show total effect sizes ($k$) and mean effect size for (A) SMD (means) and (B) lnCVR (variance) meta-regression models with personality trait as a moderator term. Thick bars are $95\%$ confidence intervals (CIs) and thin bars are prediction intervals. Circle size reflects effect size precision, where larger circles have greater precision.
(<0.05 standard deviations) or did not overlap zero. Specifically, females were more variable than males in their exploratory behaviour for birds, males were more variable than females in their aggressive behaviour for reptiles/amphibians, and females were more variable than males in their aggressive behaviour, but males were more variable than females in their social behaviour for fish (Table 2). However, the mean effect size difference in the variability of aggression for reptiles/amphibians comes from only two species, so the generality of this sex difference is unclear. Finally, none of the 25 taxon-specific tests were statistically significant after accounting for the false discovery rate (Table 2, Figs 2–6 and Table S15).

3. **Sexual size dimorphism and sex differences in personality**

SSD did not explain the extent of the sex differences in mean personality in any of the four taxonomic groups tested (mammals, birds, fish or invertebrates; Table 3 and Table S16), although some effect sizes were large (e.g. activity and aggression for mammals). When the sexes were the same size (SSD = 0), there were still no significant differences in the estimated mean personality between males and females.

There were also no significant relationships between SSD and the sex differences in personality trait variability in any of the four taxonomic groups in which we could carry out this test (see Table 3 and Table S16). Finally, there were also

---

**Fig. 6.** Sex differences in means and variability for the five personality traits for invertebrates. The phylogeny depicts all species used in the meta-analysis. Node colours reflect sexual size dimorphism (SSD) for each species (male-biased are blue, female-biased are red, monomorphic species are white and no data are grey) and heatmap data (green bars) shows the number ($N$) of effect sizes for each personality trait type for each species. Orchard plots show total effect sizes ($k$) and mean effect size for (A) SMD (means) and (B) InCVR (variance) meta-regression models with personality trait as a moderator term. Thick bars are 95% confidence intervals (CIs) and thin bars are prediction intervals. Circle size reflects effect size precision, where larger circles have greater precision.
Table 2. Multi-level meta-regression (MLMR) model summaries for each taxa with personality trait type as a fixed moderator. Sex differences were evident for some trait types, within some taxa, but there was no significant difference between the sexes. Positive estimates indicate that the mean and variability are greater for males than females. We report false discovery rate adjusted P-values for all tests for standardised mean difference (SMD) and ln coefficient of variation (lnCVR) respectively (N = 48 tests each). Estimates with 95% confidence intervals (CIs) that only just overlapped or did not overlap zero are highlighted in bold. These models are graphically represented in Figs 2–6.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>SMD mean</th>
<th>SMD 95% CIs</th>
<th>lnCVR mean</th>
<th>lnCVR 95% CIs</th>
<th>lnCVR P-value</th>
<th>N effect sizes</th>
<th>N studies</th>
<th>N species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mammals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>−0.16</td>
<td>−0.57, 0.24</td>
<td>0.85</td>
<td>0.11</td>
<td>−0.18, 0.40</td>
<td>0.95</td>
<td>83</td>
<td>14</td>
</tr>
<tr>
<td>Aggression</td>
<td>0.09</td>
<td>−0.30, 0.48</td>
<td>0.87</td>
<td>0.11</td>
<td>−0.18, 0.41</td>
<td>0.95</td>
<td>87</td>
<td>16</td>
</tr>
<tr>
<td>Boldness</td>
<td>0.13</td>
<td>−0.23, 0.49</td>
<td>0.85</td>
<td>0.03</td>
<td>−0.24, 0.30</td>
<td>0.95</td>
<td>193</td>
<td>27</td>
</tr>
<tr>
<td>Exploration</td>
<td>0.03</td>
<td>−0.35, 0.40</td>
<td>0.93</td>
<td>0.05</td>
<td>−0.23, 0.33</td>
<td>0.95</td>
<td>213</td>
<td>19</td>
</tr>
<tr>
<td>Sociality</td>
<td>0.07</td>
<td>−0.33, 0.46</td>
<td>0.87</td>
<td>0.04</td>
<td>−0.25, 0.34</td>
<td>0.95</td>
<td>98</td>
<td>12</td>
</tr>
<tr>
<td><strong>Birds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>−0.06</td>
<td>−0.37, 0.21</td>
<td>0.87</td>
<td>0.06</td>
<td>−0.22, 0.35</td>
<td>0.95</td>
<td>60</td>
<td>14</td>
</tr>
<tr>
<td>Aggression</td>
<td>−0.15</td>
<td>−0.44, 0.14</td>
<td>0.85</td>
<td>−0.08</td>
<td>−0.39, 0.24</td>
<td>0.95</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td>Boldness</td>
<td>−0.14</td>
<td>−0.40, 0.11</td>
<td>0.85</td>
<td>0.04</td>
<td>−0.18, 0.26</td>
<td>0.95</td>
<td>262</td>
<td>24</td>
</tr>
<tr>
<td>Exploration</td>
<td>0.06</td>
<td>−0.21, 0.33</td>
<td>0.87</td>
<td>−0.30</td>
<td>−0.55, −0.05</td>
<td>0.49</td>
<td>77</td>
<td>15</td>
</tr>
<tr>
<td>Sociality</td>
<td>−0.59</td>
<td>−1.06, −0.12</td>
<td>0.48</td>
<td>0.15</td>
<td>−0.36, 0.65</td>
<td>0.95</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td><strong>Reptiles/amphibians</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>−0.05</td>
<td>−0.47, 0.38</td>
<td>0.92</td>
<td>−0.10</td>
<td>−0.38, 0.18</td>
<td>0.95</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Aggression</td>
<td>−0.13</td>
<td>−0.46, 0.21</td>
<td>0.85</td>
<td>0.16</td>
<td>−0.01, 0.32</td>
<td>0.74</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Boldness</td>
<td>0.11</td>
<td>−0.14, 0.35</td>
<td>0.85</td>
<td>0.08</td>
<td>−0.05, 0.21</td>
<td>0.95</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Exploration</td>
<td>0.25</td>
<td>0.01, 0.49</td>
<td>0.48</td>
<td>−0.09</td>
<td>−0.24, 0.06</td>
<td>0.95</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>Sociality</td>
<td>0.00</td>
<td>−0.56, 0.57</td>
<td>0.99</td>
<td>−0.03</td>
<td>−0.53, 0.60</td>
<td>0.95</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><strong>Fish</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>0.15</td>
<td>−0.31, 0.62</td>
<td>0.87</td>
<td>−0.01</td>
<td>−0.16, 0.14</td>
<td>0.95</td>
<td>92</td>
<td>9</td>
</tr>
<tr>
<td>Aggression</td>
<td>0.25</td>
<td>−0.20, 0.70</td>
<td>0.85</td>
<td>−0.12</td>
<td>−0.26, 0.02</td>
<td>0.85</td>
<td>95</td>
<td>17</td>
</tr>
<tr>
<td>Boldness</td>
<td>0.08</td>
<td>−0.36, 0.52</td>
<td>0.87</td>
<td>0.02</td>
<td>−0.08, 0.13</td>
<td>0.95</td>
<td>173</td>
<td>24</td>
</tr>
<tr>
<td>Exploration</td>
<td>0.27</td>
<td>−0.20, 0.73</td>
<td>0.85</td>
<td>−0.04</td>
<td>−0.19, 0.12</td>
<td>0.95</td>
<td>103</td>
<td>10</td>
</tr>
<tr>
<td>Sociality</td>
<td>0.15</td>
<td>−0.34, 0.64</td>
<td>0.87</td>
<td>0.25</td>
<td>0.04, 0.46</td>
<td>0.49</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td><strong>Invertebrates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>0.25</td>
<td>−0.15, 0.65</td>
<td>0.95</td>
<td>−0.04</td>
<td>−0.19, 0.10</td>
<td>0.95</td>
<td>166</td>
<td>18</td>
</tr>
<tr>
<td>Aggression</td>
<td>0.42</td>
<td>−0.28, 1.12</td>
<td>0.95</td>
<td>0.04</td>
<td>−0.28, 0.35</td>
<td>0.95</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td>Boldness</td>
<td>0.24</td>
<td>−0.15, 0.63</td>
<td>0.95</td>
<td>−0.02</td>
<td>−0.15, 0.10</td>
<td>0.95</td>
<td>164</td>
<td>23</td>
</tr>
<tr>
<td>Exploration</td>
<td>0.04</td>
<td>−0.41, 0.49</td>
<td>0.95</td>
<td>0.12</td>
<td>−0.07, 0.31</td>
<td>0.95</td>
<td>54</td>
<td>7</td>
</tr>
<tr>
<td>Sociality</td>
<td>0.41</td>
<td>−0.23, 1.05</td>
<td>0.95</td>
<td>0.11</td>
<td>−0.27, 0.49</td>
<td>0.95</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

(4) Publication bias

Overall, we found no evidence for publication bias for either mean estimates of sex differences or for variability for any of the taxonomic groups (Table S18).

(5) Heterogamety and sex differences in variability in personality

We decided post hoc to conduct an exploratory analysis to compare the direction of effect sizes for birds and mammals, overall and for each personality trait, to test explicitly whether differences between birds and mammals might be due to the heterogametic sex (males for mammals, females for birds). While not significant, the observed sex differences in variability for birds and mammals tended to follow expected patterns due to the arrangement of sex chromosomes (i.e. heterogamy versus homogamy).

We found no significant differences between birds and mammals in the direction of effect sizes for variability overall (contrast lnCVR = 0.24; SE = 0.46; P = 0.347), or when the variability for the five personality traits were estimated separately (see Table S19 and Appendix S1).

IV. DISCUSSION

Prior to the development of sexual selection theory, there was a widespread view among biologists that females tended to be more variable than males (reviewed by Shields, 1975). Intriguingly, this viewpoint seems to have persisted in the biomedical literature where non-human animal studies have more often been conducted on males than females to reduce...
Fish and birds have limited empirical support from our and others’ claim of greater male phenotypic variability in non-human animals. On closer inspection, this empirical finding has been repeatedly stated that males vary more in their appearance than females. Since then, 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 in non-human animals has limited empirical support from our and others’ analyses. One general finding from a small-scale, cross-species study is that sexually selected traits in male animals do show more variation than naturally selected traits (Pomiankowski & Moller, 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 reported 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 (Wyman & Rowe, 2014). Furthermore, traits not linked to reproduction had only marginally greater variation among males than females. In another cross-species study, variation in body size was significantly greater in males than in females in taxa where males are the heterogametic sex, but the pattern was reversed in taxa where males are the homogametic sex (Reinhold & Engevist, 2013). After correctly scaling for sex differences in body size, the evidence for widespread sex differences in variation in non-human animals remains limited.

The three cross-species studies noted above 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 were partitioned by breeding system, mating system, study environment or trait type (which included the category “behaviour”). Recently, another meta-analysis investigated a vast data set on sex differences in 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 trait types there is, however, a clear bias towards either females (e.g. immunological traits, eye

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mammals</th>
<th>Birds</th>
<th>Invertebrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>0.50</td>
<td>−0.22</td>
<td>0.35</td>
</tr>
<tr>
<td>SMD</td>
<td>95% CI</td>
<td>SMD</td>
<td>SMD</td>
</tr>
<tr>
<td>1.99, 3.00</td>
<td>0.30</td>
<td>95% CI</td>
<td>0.20</td>
</tr>
<tr>
<td>lnCVR</td>
<td>0.05</td>
<td>lnCVR</td>
<td>lnCVR</td>
</tr>
<tr>
<td>−0.14, 0.25</td>
<td>0.95</td>
<td>lnCVR</td>
<td>0.95</td>
</tr>
<tr>
<td>lnCVR</td>
<td>95% CI</td>
<td>P-value</td>
<td>95% CI</td>
</tr>
<tr>
<td>0.48</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>P-value</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>lnCVR</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>lnCVR</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>N studies</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>N species</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>N effect</td>
<td>83</td>
<td>83</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 3. Subset multi-level meta-analytic model with personality trait type and SSD as moderator terms for four of the five taxonomic groups. Sexual size dimorphism (SSD) did not explain sex differences in either mean personality or for variability for any of the taxonomic groups tested. We report false discovery rate adjusted P-values for all tests for standardised mean difference (SMD) and ln coefficient of variation (lnCVR) respectively (N = 48 tests each). Positive estimates indicate that the mean and variability are greater for males than females. Estimates highlighted in bold show mean effect sizes with 95% CIs that do not overlap zero. Separate models were run for each trait type with SSD as a moderator, therefore estimates show the personality trait means when males and females are the same size (SSD = 0), and when males are larger than females (SSD > 0). Only personality types with 10 or more species, with SSD data, were estimated.
morphology) or males (e.g., morphological traits) being more variable in mice.

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 non-human animals. In humans and chimpanzees, evidence of greater male variability is similarly equivocal and remains controversial. While greater male variability has been shown for a range of morphological and physiological traits including brain structure (humans; van der Linden, Dunkel & Madison, 2017; chimpanzees Pan troglodytes: DeCasien et al., 2020) and, perhaps more notably, for behavioural traits like personality (Archer & Mehdikhani, 2003; Borkenau et al., 2013a; Karwowski et al., 2016), cognitive ability (Halpern & LaMay, 2000; Jones, Braithwaite & Healy, 2003; Arden & Plomin, 2006; Johnson et al., 2008) and academic achievement (Lehre et al., 2009; Baye & Monseur, 2016; O'Dea et al., 2018), there are also many studies reporting no sex differences in variability for those same traits. For example, the association between brain structural variation and behavioural differences between the sexes can partially be explained by failing to take into account absolute differences in mean brain size (e.g., van Eijk et al., 2021), and tests of cognitive ability and personality often reveal greater female variability (e.g., Feingold, 1994; Irwing & Lynn, 2005; Taylor & Barbot, 2021). There is also evidence that the extent of any sex difference is context dependent. For example, the gap between girls' and boys' mathematics scores becomes smaller as gender equality in society increases (Hyde & Mertz, 2009).

Our current findings are therefore intriguing but depending on one's view also unsurprising: we show using a larger database of 220 species that personality-like behavioural traits are, in general, not more variable in males than females. Additionally, any support for the greater male variability hypothesis (if it exists) is likely highly dependent on the traits and taxa in question; especially given the high heterogeneity of sex differences in variability for those same traits. For example, the gap between girls' and boys' mathematics scores becomes smaller as gender equality in society increases (Hyde & Mertz, 2009).

We should, of course, be clear that a lack of evidence for greater male variability in non-human 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 (\( F^\text{phylo} \)) explained a large proportion of variance in sex-specific differences in personality variability in males when inspecting the five personality traits (Table 2). However, the direction of the sex differences in variability between mammals and birds were not significant, even though males are heterogametic in mammals and homogametic in birds (Table S19). One explanation is that genes on sex chromosomes do not affect, or only weakly affect, behaviours associated with personality in non-human animals.

Another major biological explanation for greater variation among men than women in behavioural traits is direct sexual selection on these traits. This is especially relevant for personality as certain behavioural traits elevate the likelihood of reproducing. For example, extramort men tend to have more mating opportunities in some societies (Nettle, 2005). Similarly, there is evidence that consistent, above-average levels of aggression elevate success during male–male competition in some non-human animals (e.g., killifish Lucania goodei; McGhee & Travis, 2010). These behavioural traits are therefore sexually selected. Sexual selection is expected to result in condition-dependent expression of traits which should increase phenotypic variation because these traits ‘capture’ and magnify any difference in resource acquisition among individuals (Rowe & Houle, 1996; Wolf et al., 2007). Indeed, there is evidence in humans that some traits under sexual selection in men show a greater sex difference in variability than other traits (e.g., aggression; Archer & Mehdikhani, 2003). It is therefore relevant that we found no moderating effect of SSD, which is a proxy for the level of sexual selection on males (e.g., Rohner, Blanckenhorn & Puniamoorthy, 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 on the focal behavioural traits. 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 natural selection on females and males to be aggressive. Likewise, in group-living animals, females often establish social hierarchies where dominance is maintained through aggressive interactions (Kappler, 2017). Additionally, we might find greater or lower female variability, especially in traits like aggression or sociability, depending on when measurements of personality are taken over the reproductive cycle. That we did not find evidence of greater male variability in personality in non-human animals could also be due to many of the measures of personality involving behaviours likely to be under equivalent natural selection in both sexes. Many animal personality traits are likely to affect survival; notably boldness, which is linked to anti-predator responses, and activity or exploration, which is linked to foraging.

No evidence for greater male than female variation
mammals it would be valuable to conduct a more focussed meta-analysis looking at sex-specific variability in behaviour in catarrhine primates. Nonetheless, our findings for personality-like behaviours in non-human 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 widely accepting explanations for greater behavioural variability in men than women based on biological differences that have evolved under sexual selection 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 that yield greater behavioural variability in men than women (Gray et al., 2019).

V. CONCLUSIONS

(1) Overall, we find no evidence for male–female differences in personality in non-human animals, either for mean values or levels of variation. Crucially, there is no evidence to support the ‘greater male variability’ hypothesis in any taxonomic group for any of the five personality axes.

(2) The magnitude of sexual size dimorphism (SSD), our proxy for sexual selection, did not explain sex differences in mean personality or variability in any of the taxa–personality type combinations that we tested. Given that phylogeny (\(F_p\)) explained a large proportion of variance in sex-specific differences in personality variability in mammals it would, however, be valuable to conduct a more focussed meta-analysis looking at sex-specific variability in behaviour in catarrhine primates to explore possible evolved sex differences in variability in humans that are likely to be expressed in a broad range of environments.

(3) Our findings for non-human animals, alongside rather weak evidence for greater male variability for other traits, suggests that accepting explanations for greater behavioural variability in men than women based on biological differences that have evolved under past sex differences in sex-specific sexual or natural selection is premature. More broadly, researchers should not assume that males or men are the more variable sex when measuring traits expressed in both sexes.

VI. ACKNOWLEDGEMENTS, DATA AVAILABILITY AND AUTHOR CONTRIBUTIONS

We would like to thank Welton Menario-Costa and Constanza Leon for providing their unpublished data to include in our meta-analysis, and Timothée Bonnet, Rose O’Dea, Fonti Kar and Pieter Arnold for helpful meta-analysis discussions and advice. We also wish to thank Tim Janicke and a second, anonymous reviewer for their insightful comments.

on an earlier version of this review. L.M.H. was supported by an Australian Government Research Training Program PhD scholarship. M.D.J. and D.W.A.N. were supported by the Australian Research Council (Discovery Grant DP190100279 awarded to M.D.J., DECRA Fellowship DE150101774 awarded to D.W.A.N.). All authors declare no conflict of interest.

Data availability: all data and code used in this study have been provided as Supporting Information, and have also been made available at the Open Science Foundation: https://osf.io/b9ju6/.

Author contributions: M.D.J. and L.M.H. conceived the study, L.M.H. collected all data and conducted data analyses with D.W.A.N. L.M.H. wrote the first version of the manuscript, and D.W.A.N. and M.D.J. provided critical edits and revisions. All authors have read and approved the final version.

VII. REFERENCES

An asterisk (*) indicates that a study was used in the meta-analysis. A † indicates that a study was used for our measures of sexual selection (SSD and mating system).


No evidence for greater male than female variation


The evolution of chromosomal sex determination and genetic constraints on adaptive evolution. 


Hedges, L. V.

*Biological Reviews of the Cambridge Philosophical Society* 70, 252–259.

Guay, P. J.


morphological traits in house sparrows (Passer domesticus). *Journal of Evolutionary Biology* 16, 1296–1307.


No evidence for greater male than female variation
...
Stockley, P.


VIII. Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Supporting information and methods.

Table S1. Key word search terms.

Table S2. Sensitivity contrast models with score data.

Table S3. Exploratory analyses – mating system.

Table S4. Exploratory analyses – age.

Table S5. Exploratory analyses – study population.

Table S6. Exploratory analyses – study environment.

Table S7. Exploratory analyses – study type.

Table S8. Sensitivity analyses – intercept models with D matrix ($\rho = 0.3$).

Table S9. Sensitivity analyses – intercept models with D matrix ($\rho = 0.5$).

Table S10. Sensitivity analyses – intercept models with D matrix ($\rho = 0.8$).

Table S11. Sensitivity analyses – multi-level meta-regression personality trait models with D matrix ($\rho = 0.3$).

Table S12. Sensitivity analyses – multi-level meta-regression personality trait models with D matrix ($\rho = 0.5$).

Table S13. Sensitivity analyses – multi-level meta-regression personality trait models with D matrix ($\rho = 0.8$).

Table S14. Intercept-only random effects meta-analysis model output for each of the five taxonomic groups comparing males and females for mean differences (SMD) and for variability (lnCVR).

Table S15. Multi-level meta-regression model output for each of the five taxonomic groups with personality trait type as a moderator.

Table S16. Subset meta-analysis models for each of the taxonomic groups/personality trait types where we could include sexual size dimorphism (SSD) as a moderator.

Table S17. Multi-level meta-regression model output for each of the five taxonomic groups with personality trait type and sexual size dimorphism (SSD), and their interaction, as moderator terms.

Table S18. Publication bias.

Table S19. Exploratory analysis – multi-level meta-regression model comparing (a) mean overall variability (lnCVR) and (b) personality trait effect sizes for variability for birds and mammals.

Data file S1. Final data file (data.xlsx).

Data file S2. Sexual selection data location file (sexual_selection.xlsx).

Data file S3. R code used for data analysis (code.R).

Appendix S2. Supporting Information

(Received 11 January 2021; revised 13 November 2021; accepted 17 November 2021; published online 14 December 2021)