**Results**

*The effect size dataset*

Looking at the format for Nat Comm (in Cally’s paper), this first paragraph should outline/breakdown the dataset – what kinds of numbers do we have overall? Per personality trait? Per taxo group? What did we get rid of? Lots of n’s in here… Mention that because our dataset covers a broad range of taxonomic groups, and because these animal groups are all so different (i.e. ectotherms and endotherms, invertebrates and vertebrates, heterogametic males and homogametic males) we thought it best to split the dataset by taxonomic group for all analysis.

*Personalities do not differ between the sexes across taxonomic groups*

The overall means across all personality traits were not significantly different from zero in any of the taxonomic groups (restricted maximum likelihood (REML) birds: *β* = -0.09, 95% CIs: -0.33, 0.14; fish: *β* = -0.02, 95% CIs: -0.38, 0.33; invertebrates: *β* = 0.20, 95% CIs: -0.08, 0.48; mammals: *β* = 0.09, 95% CIs: -0.25, 0.43; reptiles: *β* =0.06, 95% CIs: -0.11, 0.22), indicating an absence of sex differences in mean personality behaviours.

There were also no significant sex differences in variability, across all personality traits and all taxonomic groups (restricted maximum likelihood (REML) birds: *β* = -0.16, 95% CIs: -0.67, 0.34; fish: *β* = -0.02, 95% CIs: -0.08, 0.04; invertebrates: *β* = -0.04, 95% CIs: -0.12, 0.04; mammals: *β* = 0.07, 95% CIs: -0.20, 0.33; reptiles: *β* =0.02, 95% CIs: -0.07, 0.12). We observed high heterogeneity (total I2) within each taxonomic group for both SMD and lnCVR datasets (see Table ??). For the SMD dataset, heterogeneity was mostly explained by between study variance by including Study ID in our models (see Table ??), with exception for fish; heterogeneity was attributed mostly to phylogenetic relatedness (total I2 = 72%, 95% CIs: 68-75%; phylogenetic I2 = 23%, 95% CIs: 13-35%). Interestingly, heterogeneity within the lnCVR dataset for birds and mammals was explained by phylogenetic relatedness (birds: total I2 = 94%, 95% CIs: 93-94%; phylogenetic I2 = 46%, 95% CIs: 39-54%; mammals: total I2 = 59%, 95% CIs: 54-64%; phylogenetic I2 = 41%, 95% CIs: 30-51%), while heterogeneity remained mostly unexplained for the remaining taxonomic groups (see Table ??).

*Sex differences and the animal ‘Big Five’*

When personality traits were broken down into the Big Five (Activity, Aggression, Boldness, Exploration and Sociality/Sociability) we found significant sex differences in both the means and the variability for some personality traits within some taxonomic groups.

*SMD*

Female birds were more sociable than males (*β* =-0.48, 95% CIs: -0.93, 0.03, *t* = -2.07, *p*=0.04), male invertebrates were more aggressive than females (*β* =0.46, 95% CIs: 0.07, 0.852, *t* = 2.31, *p*=0.02), and male reptiles were more explorative than females (*β* =0.25, 95% CIs: 0.03, 0.47, *t* = 2.21, *p*=0.03), although it is important to note that some of these results might be driven by relatively small sample sizes (see Table ??). There were no strong sex differences in mean personality for any of the Big Five traits for mammals or fish (see Table ??).

*lnCVR*

Female birds had greater variability in their exploratory behaviour than did males (*β* =-0.31, 95% CIs: -0.56, -0.05, *t* = -2.38, *p*=0.02), but there were no sex differences in variability for any other personality trait within birds. Interestingly, female fish had greater variability in aggressive behaviour than did males (*β* =-0.13, 95% CIs: -0.25, -0.01, *t* = -2.08, *p*=0.04), while male fish had somewhat greater variability in their social behaviour than females (*β* =0.17, 95% CIs: -0.01, 0.35, *t* = 1.91, *p*=0.06). Female invertebrates showed greater variability than males in activity traits (*β* =-0.14, 95% CIs: -0.25, -0.03, *t* = -2.44, *p*=0.01), while there were no sex differences in variability within any other personality trait. Male reptiles had greater variability in their aggressive behaviour than females (*β* =0.16, 95% CIs: -0.01, 0.32, *t* = 1.89, *p*=0.06), although this effect was likely driven by a small sample size (n= 2 studies, 2 species, 3 effect sizes). We found no sex differences in the variability of mammals for any of the five personality traits.

*Sexual size dimorphism alone doesn’t explain sex differences in the mean or variability of personalities, but interacts with some traits to produce strong sex differences.*

Sexual size dimorphism (SSD) alone did not explain sex differences in personality for any taxonomic group (see Table ??), but did interact significantly with some personality traits after categories with fewer than 10 species with SSD data were dropped from the meta-regression model. There was not enough data on reptiles to include them in these models.

*SMD*

The degree of sexual size dimorphism between males and females did not explain boldness behaviour in birds (*β* =-0.17, 95% CIs: -0.71, 0.36, *t* = -0.63, *p*=0.53), nor did it explain aggressive or boldness behaviour in fish (aggression: *β* =-0.12, 95% CIs: -0.66, 0.42, *t* = -0.44, *p*=0.66; boldness: *β* =-0.04, 95% CIs: -0.58, 0.51, *t* = -0.13, *p*=0.90; SSD\*aggression: *β* =0.26, 95% CIs: -0.72, 1.24, *t* = 0.52, *p*=0.60; SSD\*boldness: *β* =-0.37, 95% CIs: -1.43, 0.69, *t* = -0.69, *p*=0.49). For invertebrates, SSD interacted with boldness behaviour so that as males became larger than females (SSD became more male-biased) males also became much bolder than females (boldness: *β* =0.28, 95% CIs: -0.10, 0.66, *t* = 1.46, *p*=0.14; SSD\*boldness: *β* =1.07, 95% CIs: 0.25, 1.90, *t* = 2.56, *p*=0.01). SSD did not interact with activity behaviour in invertebrates, however as SSD became more male-biased females became slightly more active than males (activity: *β* =0.00, 95% CIs: -0.37, 0.38, *t* = 0.02, *p*=0.98; SSD\*activity: *β* =-0.85, 95% CIs: -2.02, 0.32, *t* = -1.43, *p*=0.15). Finally, SSD interacted strongly with all personality traits (except sociality for which there was not enough data) within mammals. When SSD becomes more male-biased, females are much more active than males (SSD\*activity: *β* =-1.98, 95% CIs: -2.89, -1.07, *t* = -4.26, *p*<0.001), but males are much more aggressive (SSD\*aggression: *β* =3.49, 95% CIs: 2.12, 4.86, *t* = 4.99, *p*<0.001), are much bolder (SSD\*boldness: *β* =1.94, 95% CIs: 0.95, 2.92, *t* = 3.86, *p*<0.001), and are more explorative than females (SSD\*exploration: *β* =1.94, 95% CIs: 0.94, 2.94, *t* = 3.82, *p*<0.001).

*lnCVR*

For birds, SSD interacted with boldness behaviour so that as SSD became more male-biased males had somewhat more variability in their boldness behaviour than did females (*β* =0.06, 95% CIs: -0.01, 0.12, *t* = 1.80, *p*=0.07). SSD did not explain variability in activity or aggression behaviour in invertebrates, aggression or boldness behaviour in fish, or activity, aggression, boldness or exploratory behaviour in mammals (see Table ??).

*Sensitivity analyses and robustness*

When I do them, put them in here …

*The effects of mating system, age, and other study characteristics on personality means and variability*

Secondary models were conducted to understand how mating system (monogamy or multiple mating), age (juvenile or adult), study population (wild or lab/captive), study environment (field or lab) and study type (observational or experimental) might moderate sex differences in means and variability for personalities in each of the five taxonomic groups. Some of these categories have previously been implicated to influence sex differences in mean and variability of animal personalities (e.g. Tarka *et al*. 2018), so they were also included in our analysis for confirmation. It is important to note that there was inadequate data to compare any secondary model categories at the personality trait level within taxa.

Surprisingly, neither mating system nor age explained sex differences in either mean or variability of personality in any of the taxonomic groups (see Table ?? for slope estimates and 95% CIs). However, female fish from wild populations had significantly greater variability in their personality compared to males from the wild (wild: *β* = -0.10, 95% CIs: -0.19, -0.01, *p*=0.03), while lab-reared fish populations had no sex differences in variability (lab/captive: *β* = 0.03, 95% CIs: -0.05, 0.10, *p*=0.68). Such a sex difference was absent in the mean personalities of both wild and lab fish populations (wild: *β* = -0.11, 95% CIs: -0.47, 0.25; lab/captive: *β* = 0.03, 95% CIs: -0.33, 0.38), and there was no effect of population on sex differences in the mean or variability of personalities in any other taxonomic group (see Table ??). Study environment had a marginally significant effect on mean personality in mammals; males in the wild had somewhat larger overall means than females (*β* = 0.27, 95% CIs: -0.02, 0.56, *p*=0.06), a difference that was absent from mammals in a laboratory environment (*β* = -0.06, 95% CIs: -0.35, 0.23, *p*=0.66). Study environment did not influence sex differences in mean personality in any other taxonomic group, or the variability of any taxonomic group (including mammals; see Table ??). Finally, whether a study was observational (no manipulation) or experimental (any kind of experimental manipulation) explained sex differences in the mean personality of mammals; males had greater means than females in observational studies compared to mammals in experimental studies (experimental: *β* = 0.01, 95% CIs: -0.12, 0.14, *p*=0.90; observational: *β* = 0.43, 95% CIs: 0.21, 0.65, *p*<0.001). Again, there was no effect of study type on variability for mammals in either experimental or observational studies, nor for the means or variability in any other taxonomic group (see Table ??).

*Publication bias*

Funnel plots and Egger’s regression test were used to look for potential publication bias in our standardised mean (Hedge’s *g*) and variability (lnCVR) effect size datasets. The Egger’s test suggested the presence of publication bias within the standardised means dataset (*z* = -2.79, p= 0.005), but not the variability dataset (*z* = -0.03, p = 0.97). However, significance generated from the Egger’s test can also be an indicator of high heterogeneity, which is the case for our effect sizes. *As an additional check, we looked at the correlation between effect size (Hedge’s g) and journal impact factor, and the correlation between effect size and publication year…*