**Materials and Methods**

*Literature search and data collection*

We conducted a systematic review and meta-analysis by systematically searching the literature for empirical studies quantifying animal personality traits in males and females (as classified by Sih *et al*. 2004 & Réale *et al*. 2007). We collected 10,991 papers from both the Web of Science and Scopus databases in December 2018, using a primary and secondary keyword searches (see Figure 3). We also included every paper that cited one of five key animal personality reviews: (Dall *et al.* 2004; Sih *et al.* 2004; Réale *et al.* 2007; Schuett *et al.* 2010; Dingemanse & Wolf 2010). A secondary search was used to find and include studies that studied ‘personality-like behaviours’ that were not defined as personalities.

After removing duplicates, 3,737 articles were left for title and abstract screening by LH. A total of XX papers were excluded based on title and abstract because they ….. This left 2,950 papers for full text screening. We included papers only if they met the following inclusion criteria: 1) provided raw means and some measurement of error for both males and females given raw summary statistics are required for the derivation of variance-based effect size measures and provide greater opportunities to control for sources of non-independence (Nakagawa *et al.* 2015; Noble *et al.* 2017). Studies that only report PCAs and factor loadings were excluded as their interpretation can be challenging (see Figure 2); 2)……

In total, this provided 245 eligible studies with relevant data that could be included in our meta-analysis (see Figure 3). We extracted means, error (standard deviation, standard error) and sample sizes from text, tables and figures for both males and females on all personality traits. We used the R package *metaDigitise* (v1.0.0, Pick et al. 2019) to extract summary statistics from figures where needed.

*Effect size and sampling error*

To understand how male and females differ in their personality and variability in personality trait we used two contrast-based effect size measures. To understand changes in mean personality between the sexes we calculated Hedges’ g (sample size adjusted standardised mean difference) and its associated sampling error as follows:

,

Where is the mean, is the pooled standard deviation and N and SD are the sample size and standard deviation for males (M) and females (F), respectively. We uses Hedges g instead of log response ratios as our effect size given much of the data was not on a ratio scale and as such a response ratio could not be calculated. To understand differences in variance in personality traits across the sexes we used the log coefficient of variation (lnCVR) (Nakagawa *et al.* 2015) calculating the effect size and samping variance as follows:

,

where CV*M* and CV*F* are *SDM /M*  and *SDF /F*  , respectively blah blah. The lnCVR allows us to quantify the difference in variance between the sexes independent of changes in the mean (Nakagawa *et al.* 2015)

*Moderator variables*

We collected a series of moderator variables that we *a priori* expected to moderate the effect size magnitude and direction. Personality traits extracted for males and females were categorised into one of five types suggested by Réale *et al.* (2007); ‘boldness’, ‘aggression’, ‘activity’, ‘sociality’ and ‘exploration’. We also recorded the taxonomic group (‘invertebrates’, ‘fish’, ‘amphibians’, ‘reptiles’, ‘birds’, ‘mammals’), the age of the sample (‘juveniles’ or ‘adults’), whether the study was done in the lab or field given this is expected to affect behaviour (Tarka et al. 2018). We also generated two moderator variables that attempt to capture the strength of sexual selection. First, we quantified the degree of sexual size dimorphism (SSD) between males and females as this is strongly correlated with the strength of sexual selection (REFS). We created a SSD index by taking the ratio of male and female mean body size, mass or some other sexually dimorphic trait (e.g.). Our second measure attempting to capture the strength of sexual selection was the type of mating system (‘polygynous’, polygamous’, ‘monogamous’, polyandrous’). While mating system is a more indirect measure a greater amount of data was available to categorise species compared with measures of direct SSD between the sexes. Lastly, we also categorised the parental care strategies exhibited by each species within and across studies (PROVIDE LEVELSHERE). Where body size measures for males and females, mating system and parental care were not reported in the research article itself, we obtained these data by searching Web of Science, Scopus and Google Scholar. We used the search terms: “species name” AND male AND female AND body size OR length (for body size measures), “species name” AND parental care OR mating system for parental care and mating system.

*Meta-analyses*

Finally, if the final dataset contains moderators with fewer than 2 effect sizes per level (or a number that is inappropriate for analysis), they might be removed from further analysis.



3,734

784

245

245

2,705

2,950

10,991

7,257

182

**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.