**Supplementary Material for Harrison *et al*. 2020**

1. *Keyword search terms*

**Table S23.** Primary and secondary keyword searches used for our literature search conducted on 11 December 2018. Keyword searches differed slightly to account for how operator terms are employed by each database. The searches were further refined by using relevant biological categories in ISI or SCOPUS.

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

1. *Data collection – mating system and parental care moderator terms*

We searched *ISI Web of Science, Scopus* and *Google Scholar* using the search terms: “species name” AND “mating system” for mating system and “species name” AND “parental care” for parental care. For birds, we also searched the *CRC Handbook of Avian Body Masses* (Dunning Jr 2007) and the online reference database *Birds of the World* (birdsoftheworld.org; accessed via an ANU library subscription in 2019) by searching “species name”. We noted whether the mating system of the species was characterised by ‘multiple mating’ or ‘monogamy’, and whether the species provided ‘maternal’, ‘paternal’, ‘biparental’, ‘cooperative’ or ‘no care’. However, after data collection we decided to drop parental care from subsequent analysis because we did not have enough data for enough species to run our proposed meta-regression models. The location of data collected for both parental care and mating system (and body sizes for SSD) are provided in the accompanying Supplementary data file ‘sexual\_selection.xlsx’.

1. *Excluding studies*

*Expand on exclusion terms, what was excluded (i.e. papers that were missing sample sizes, error, means were 0 so we couldn’t use them in our models etc). lots of ns for transparency…*

While we decided not to email authors for missing data (i.e. sample sizes, error type, additional information), we did obtain sample sizes for males and females in study P077 via email correspondence with a co-author of the study.

Additionally, there are 3 studies in our dataset (P172, P210 and P231) that we were concerned might have issues with data duplication/unreliability. LMH checked the retraction database *retractiondatabase.org* regularly during data analysis to check if these studies had been either retracted or flagged for concern. As of 31 August 2020, none of the 3 studies had been retracted, nor had any other issues, so we kept these studies in our final dataset.

1. *Data transformations*

*Score data*

We performed data transformations on latency data and proportional data in order to meet assumptions about normality. However, we were unable to adjust score data, and therefore assumed that scores were normally distributed. Scores did not make up a large proportion of our effect size dataset, but we decided to run contrast meta-analysis models to check whether data composed of scores were significantly different from the rest of our dataset. Summaries of these models are shown in Supplementary Table S13.

*Score data - invertebrates*

Effect sizes calculated from scores were significantly different from the rest of the dataset (including transformed latency and proportion data) for mean differences in personality for invertebrates only (intercept: =0.30, 95% CIs: 0.02, 0.57, *t* = 2.12, *p*=0.03; score: = -0.29, 95% CIs: -0.57, -0.02, *t* = -2.11, *p*=0.04). Invertebrates also had the most score data of any taxonomic group (n=61 effect sizes). We therefore decided to rerun our main three meta-analysis models for invertebrates after removing score data to check for significant differences.

When invertebrate score data was removed, there was 1) no difference in either the mean or the variability between males and females (intercept only model, see Supplementary Table S6.1); 2) no significant sex difference in either the mean or the variability for any of the five personality traits (but males were marginally significantly more active and bolder than females, see Supplementary Table S6.2); and 3) as males became larger than females there was no difference in either the mean or variability of active or boldness behaviour (see Supplementary Table S6.3). As such, removing the score data from the invertebrate dataset likely reduced model power and reduced previously significant results to marginal or no significance. The direction of these effects, however, did not change after removing score data.

1. *Calculating I2*

We extracted *I*2 from our meta-analytic intercept-only models (see Supplementary Table S1) using the following equation:

Where is the total variance, is the phylogenetic variance, is the between-study variance, is the species-specific variance, is the study-specific variance (observation-level random effect), and is the remaining within-study sampling variance (random effects) (Nakagawa & Santos 2012).

Further, we can then partition *I*2 to calculate study-level *I*2 and species-level *I*2 (*I*2s and *I*2u, respectively) (Nakagawa & Santos 2012):

*I*2s = / ,

*I*2u = /

Finally, we can determine the strength of phylogenetic variance using the equation:

Where = 0 there is no phylogenetic signal, but when = 1 there is a strong effect of phylogeny on heterogeneity (Nakagawa & Santos 2012).

1. *Exploratory analysis*

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. For mating system, we followed methods similar to those used to obtain SSD measures. Where mating system was not reported in the research article, we searched *Web of Science*, *Scopus* and *Google Scholar* using the search terms: “species name” AND “mating system”. For birds, we first searched the online reference database *Birds of the World* (birdsoftheworld.org; accessed via an ANU library subscription in 2019). Initially we wanted to categorise mating system into the following: “monogamous”, “polyandrous”, “polygynous”, or “promiscuous”. However, it was difficult to find studies that agreed on mating system definitions for many species, so we collapsed our mating system categories into just “monogamous” or “multiple mating”. The location of data collected for mating system is provided in the Supplementary Material data files (see sexual\_selection.xlsx).

1. *Exploratory analysis - results*

*Mating System*

Monogamous and multiple mating systems were not significantly different from each other for means or variability for any of the taxonomic groups (Supplementary Table S7).

*Age*

Mean personality effect sizes for adults were marginally significantly different to juveniles for invertebrates (intercept: =0.24, 95% CIs: -0.03, 0.51, *t* = 1.74, *p*=0.08; juvenile: =-0.03, 95% CIs: -0.34, 0.28, *t* = -0.18, *p*=0.86), but not any other taxonomic group, and not for variability (Supplementary Table S8).

*Population*

Fish from the wild had greater differences in variability than fish from lab populations (intercept: =-0.09, 95% CIs: -0.18, -0.01, *t* = -2.11, *p*=0.04; lab: =0.08, 95% CIs: -0.02, 0.19, *t* = 1.55, *p*=0.12), but not for mean personality differences, and not for any other taxonomic group (Supplementary Table S9).

*Study environment*

Studies conducted in the lab were significantly different to field studies for mammals (intercept: =0.24, 95% CIs: -0.10, 0.57, *t* = 1.38, *p*=0.17; lab: = -0.30, 95% CIs: -0.56, -0.04, *t* = -2.26, *p*=0.02), and marginally significantly different for reptilia (intercept: =-0.29, 95% CIs: -0.77, 0.18, *t* = -1.24, *p*=0.22; lab: = 0.42, 95% CIs: -0.07, 0.91, *t* = -1.70, *p*=0.09), but not for variability, and not for any other taxonomic group (Supplementary Table S10).

*Study type*

Effect sizes from observational studies were significantly different from experimental studies for mammals (intercept: =0.00, 95% CIs: -0.21, 0.22, *t* = 0.04, *p*=0.97; observation: = 0.38, 95% CIs: 0.12, 0.64, *t* = 2.81, *p*=0.005) and marginally significantly different for reptilia (intercept: =0.13, 95% CIs: -0.09, 0.34, *t* = 1.14, *p*=0.26; observation: =-0.42, 95% CIs: -0.91, 0.07, *t* = -1.70, *p*=0.09), but not for variability, and not for any other taxonomic group (Supplementary Table S11).