

1 **Reliability of meta-analyses in ecology and evolution: (mostly) good news from a case**
2 **study on sexual signals**

3 **Pietro Pollo^{1*}, Malgorzata Lagisz¹, Renato Chaves Macedo-Rego², Ayumi Mizuno^{1,3},**
4 **Yefeng Yang¹, Shinichi Nakagawa^{1,3}**

5 ¹ Evolution & Ecology Research Centre, School of Biological, Earth & Environmental
6 Sciences, University of New South Wales, Kensington, NSW, 2052, Australia

7 ² Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. Peter Henry Rolfs,
8 Viçosa, Minas Gerais, 36570-000, Brazil

9 ³ Department of Biological Sciences, University of Alberta, CW 405, Biological Sciences
10 Building, Edmonton, AB T6G 2E9, Canada

11 * Corresponding author: pietro_pollo@hotmail.com

12

13 **Abstract**

14 Meta-analyses are powerful synthesis tools that are popular in ecology and evolution due to the
15 rapidly growing literature of this field. Although the usefulness of meta-analyses depends on
16 their reliability, such as the precision of individual and mean effect sizes, attempts to reproduce
17 meta-analyses' results remain rare in ecology and evolution. Here, we assess the reliability of
18 41 meta-analyses on sexual signals by evaluating the replicability and reproducibility of their
19 results. We attempted to (1) replicate meta-analyses' mean effect sizes using the datasets they
20 provided, (2) replicate meta-analyses' effect sizes by re-extracting 5,703 effect sizes from 246
21 empirical studies they used as sources, (3) assess the extent of relevant data missed by original
22 meta-analyses, and (4) reproduce meta-analyses' mean effect sizes after incorporating re-
23 extracted and relevant missing data. We found many discrepancies between meta-analyses'
24 reported results and those generated by our analyses for all replicability and reproducibility
25 attempts. Nonetheless, we argue that the meta-analyses we evaluated are largely reproducible

26 because the differences we found were small in magnitude, leaving the original interpretation
27 of these meta-analyses' results unchanged. Still, we highlight issues we observed in these meta-
28 analyses that affected their reliability, providing recommendations to ameliorate them.

29

30 **Introduction**

31 The literature of ecology and evolution, like of other fields of study, is expanding rapidly^{1,2}.
32 Consequently, synthesising this growing body of work becomes increasingly necessary to
33 identify patterns across individual studies. Meta-analyses, which are perceived as the gold
34 standard for evidence synthesis, can contribute to this endeavour because they aim to detect
35 and retrieve all relevant studies on a given topic, extract data from these studies, and quantify
36 an average effect of interest using the extracted data^{3,4}. Nonetheless, many challenges arise
37 during the execution of meta-analytical studies, requiring researchers to be vigilant to ensure
38 their reliability.

39 Foremost, meta-analyses need to be transparent, which means that they must provide
40 details on decisions and resources regarding all steps of their implementation (e.g. searches,
41 screening, data extraction, analysis code, etc). Yet, recent appraisals of the secondary literature
42 in ecology and evolution show that meta-analyses in this field are often poorly transparent⁵⁻⁷.
43 Reporting guidelines, such as MOOSE⁸ and PRISMA^{6,9}, were developed to address
44 transparency issues in evidence syntheses. Researchers who adhere to these guidelines can thus
45 enhance the quality of their meta-analyses. Still, transparency represents only the first step for
46 high-quality syntheses as providing information does not guarantee its reliability. For instance,
47 researchers can make mistakes when describing their methods, provide faulty code for their
48 analyses, or even forge data, impairing the reproducibility of their findings. Therefore, it is
49 imperative to evaluate studies beyond transparency.

50 Meta-analyses possess the advantage of having sources of information that are readily
51 accessible (i.e. individual studies from which data are extracted). This means that meta-
52 analyses can be more easily evaluated than other approaches regarding reproducibility aspects.
53 More specifically, meta-analytical studies that report their decision criteria (i.e. studies and data
54 considered valid for inclusion) and their effect sizes with information on their sources (i.e.
55 studies they were extracted from) should allow evaluations regarding the reproducibility of
56 their dataset and their results. However, to our knowledge, only computational reproducibility
57 of meta-analyses (i.e. reproducibility of their code) has been examined in ecology and
58 evolution¹⁰. By contrast, reproducibility attempts of data extraction and general results have
59 only been conducted in other fields of study (but see ref. ¹¹; e.g. medicine¹²; psychology¹³).
60 Furthermore, other elements connected to the reliability of meta-analyses, such as their efficacy
61 in detecting relevant studies, remain virtually unexplored.

62 Here, using 41 meta-analyses related to sexual signals (refs. ¹⁴⁻⁵⁷), we conduct the
63 largest reproducibility effort for meta-analyses ever done in terms of number of primary studies
64 ($N = 246$) or the number of re-extracted effect sizes ($N = 5,703$). We evaluate multiple reliability
65 aspects related to distinct implementation stages of these meta-analyses (Figure S1). First, we
66 re-analyse the dataset provided by these meta-analyses, comparing the mean effect sizes we
67 obtained with those reported in them. Second, we extract data from their original sources (i.e.
68 individual studies) and compare these extracted data points with the ones reported in these
69 meta-analyses. Third, we assess how many data points from the original sources that we
70 verified should have been extracted and included in these meta-analyses' datasets but were not
71 (i.e. "missed" data). Fourth, because these meta-analyses ask similar questions involving sexual
72 signals, we estimate the minimum number of studies that contained relevant data but were not
73 listed as sources in these meta-analyses (i.e. undetected studies). Fifth, we assess the extent that

74 the results of these meta-analyses change when re-extracted data (along with missed data and
75 undetected studies) is analysed in place of originally reported data.

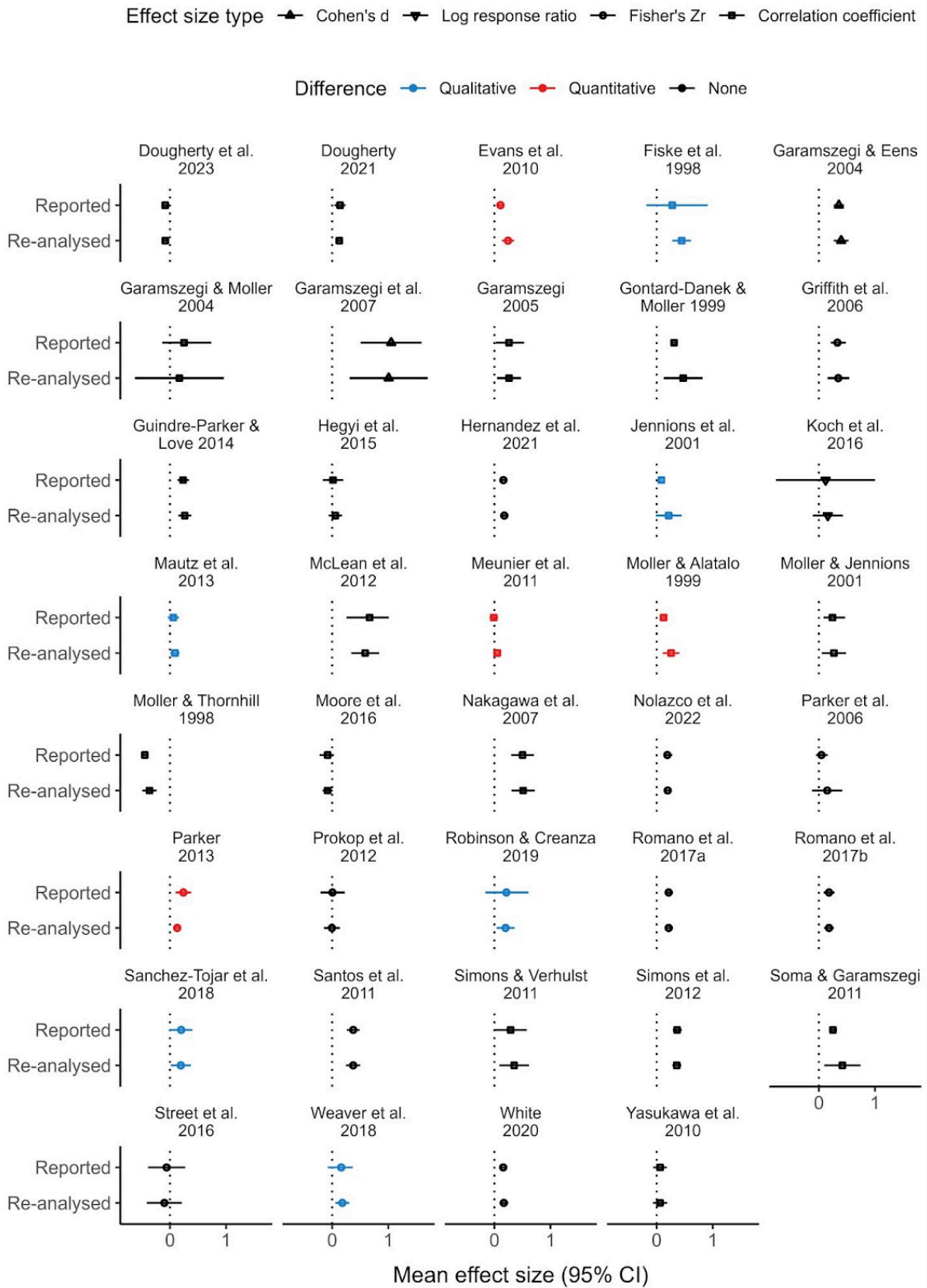
76

77 **Results**

78 *Replicability of mean effect sizes*

79 We found qualitative differences in 15.4% (6 out of 39) of the comparisons between mean
80 effect sizes reported by meta-analyses and mean effect sizes resulting from re-analyses of those
81 meta-analyses' reported datasets (Figure 1). More specifically, five of these meta-analyses
82 reported a mean effect size not different from zero but its re-analysed counterpart was positive,
83 while the opposite happened for another meta-analysis. We also detected quantitative
84 differences between originally reported and re-analysed mean effect sizes in approximately a
85 tenth of comparisons made (4 out of 39), although none of them were qualitative differences
86 (Figure 1). Despite this, the absolute difference in magnitude between originally reported and
87 re-analysed mean effect sizes was always lower than 0.2, even for cases with detected
88 qualitative or quantitative differences (Figure 2). We also found that effect sizes were positively
89 predicted by effective sample size (i.e. evidence of publication bias) in 23.1% (9 out of 39) of
90 meta-analyses. Two-thirds of these meta-analyses (6 out of 9) diligently reported that they
91 detected publication bias (refs. ^{23,24,26,27,39,49}), while the remaining third did not assess
92 publication bias whatsoever (refs. ^{17,20,54}).

93



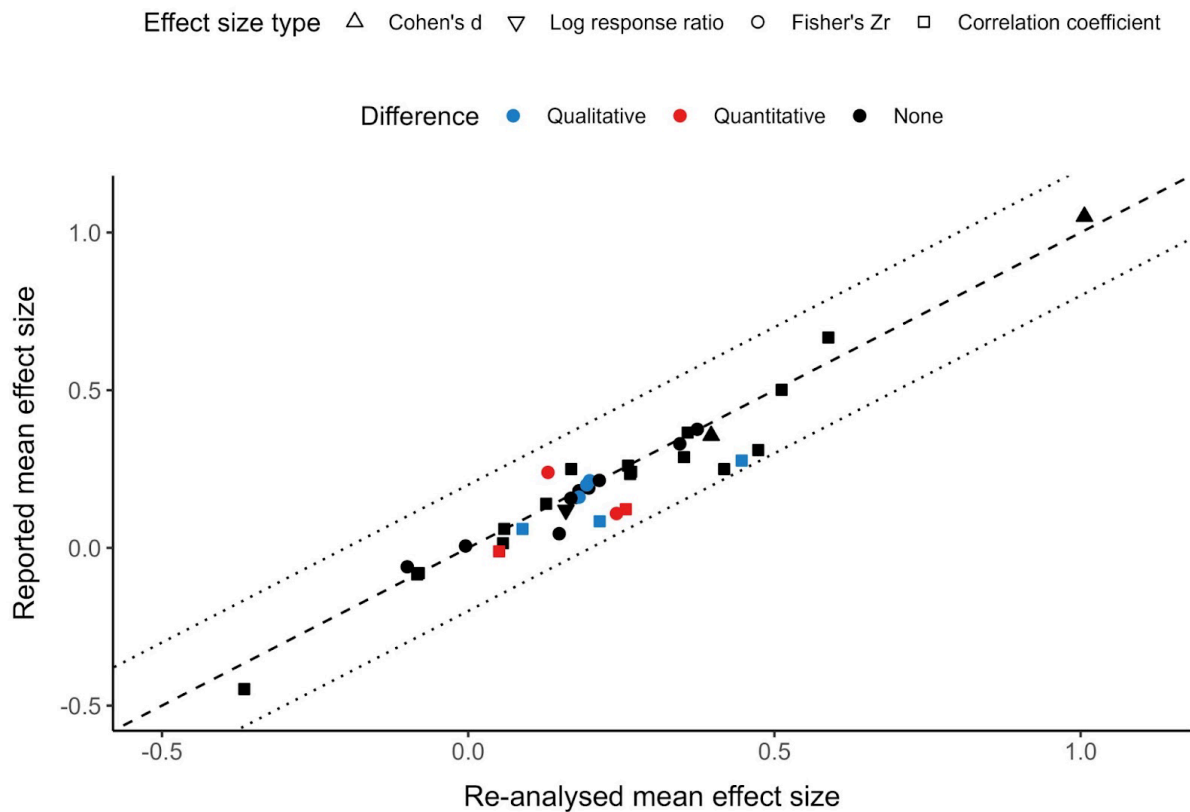
94

95 Figure 1. Originally reported and re-analysed mean effect sizes of 39 meta-analyses.

96 Qualitative differences represent a change in interpretation between pairs (positive vs. not

97 different from zero or vice-versa), while quantitative differences represent statistical
98 differences (absolute z-score greater than 1.96).

99



100

101 Figure 2. Comparison between mean effect sizes originally reported by meta-analyses and
102 mean effect sizes resulting from the re-analysis of the reported dataset of the same meta-
103 analyses. The dashed line highlights a perfect correspondence between variables, while dotted
104 lines highlight a difference of 0.2 and -0.2 between reported and re-analysed mean effect sizes.

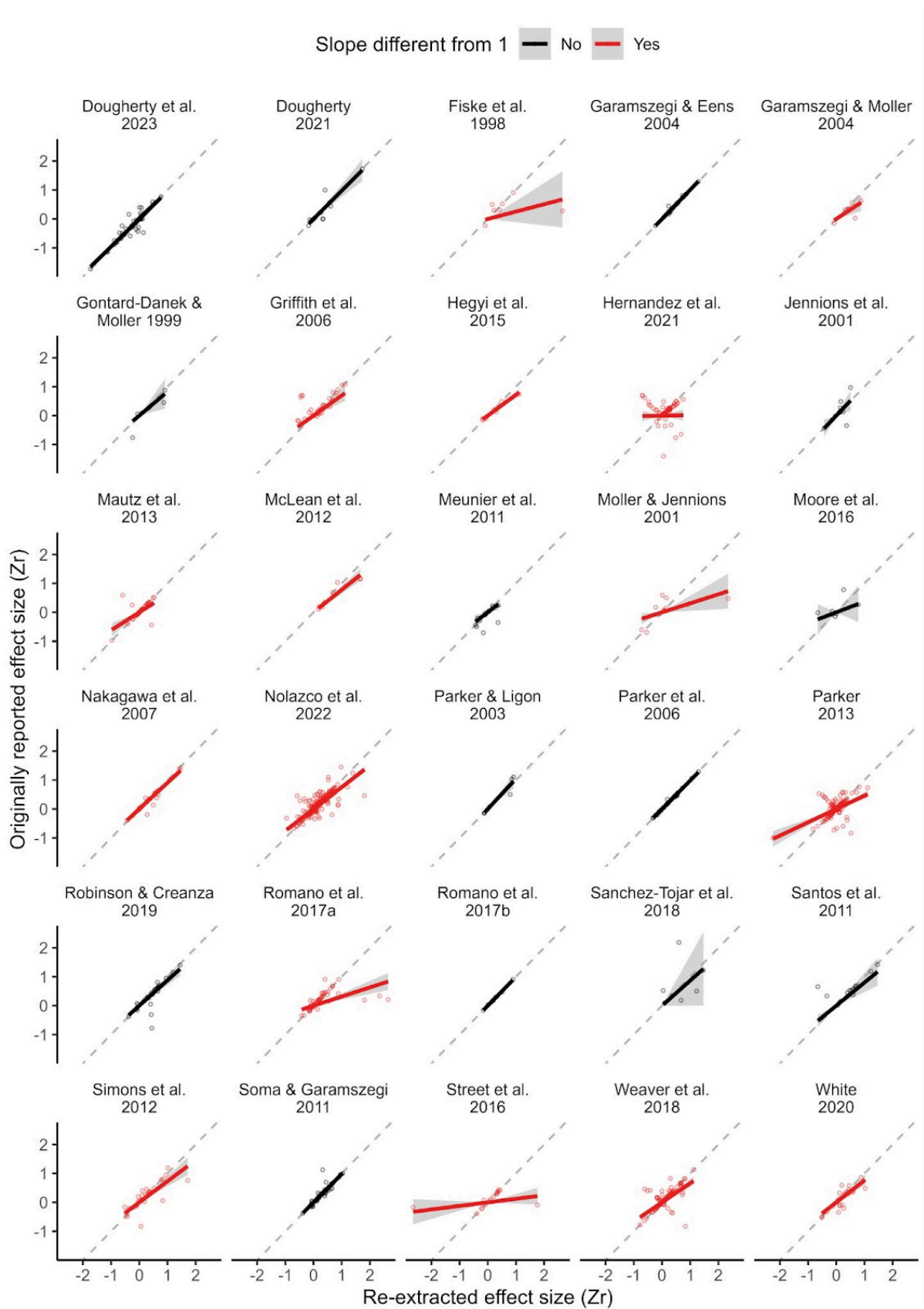
105

106 *Replicability of individual effect sizes*

107 We found that effect sizes originally reported by meta-analyses were, on average, statistically
108 identical to their equivalent re-extracted ones in less than half (14 out of 30) of the meta-
109 analyses evaluated (Figure 3, Table S3). For the remaining meta-analyses, the slope from linear
110 regressions between reported and re-extracted effect sizes was always lower than 1 ($\bar{x} \pm SE =$

111 0.55 ± 0.07), showing that reported effect sizes tended to be smaller or in the opposite direction

112 than re-extracted ones in these cases.



113

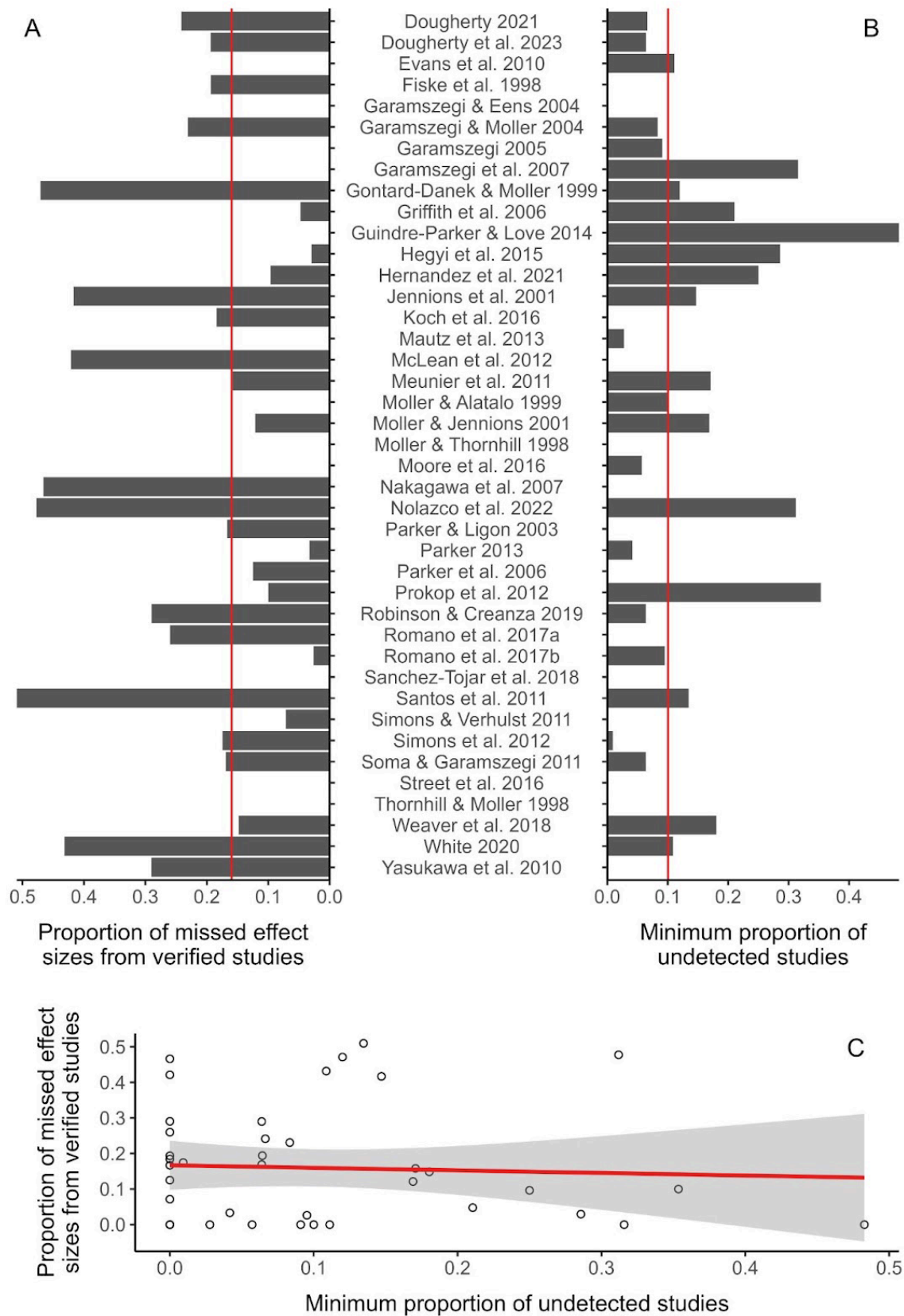
114

115 Figure 3. Relationship between effect sizes originally reported by meta-analyses and effect
116 sizes re-extracted from empirical studies. Solid lines represent a linear regression between
117 reported and re-extracted effect sizes (with the intercept forced to zero), while shaded areas
118 represent their 95% confidence interval. Red lines indicate slopes that significantly differ from
119 1. Dashed lines highlight a perfect relationship between reported and re-extracted effect sizes.

120

121 *Missing data and undetected studies*

122 We found that meta-analyses missed (i.e. failed to extract and report), on average, $16 \pm 2.6\%$
123 ($\bar{x} \pm SE$) of relevant effect sizes from empirical studies they used as sources for data extraction
124 (Figure 4A). Additionally, we found that meta-analyses were unsuccessful in detecting, on
125 average, at least 10% of empirical studies they should have included as data sources (Figure
126 4B). Furthermore, we found no association between the number of missing effect sizes and the
127 minimum proportion of undetected studies across meta-analyses (Figure 4C).



128

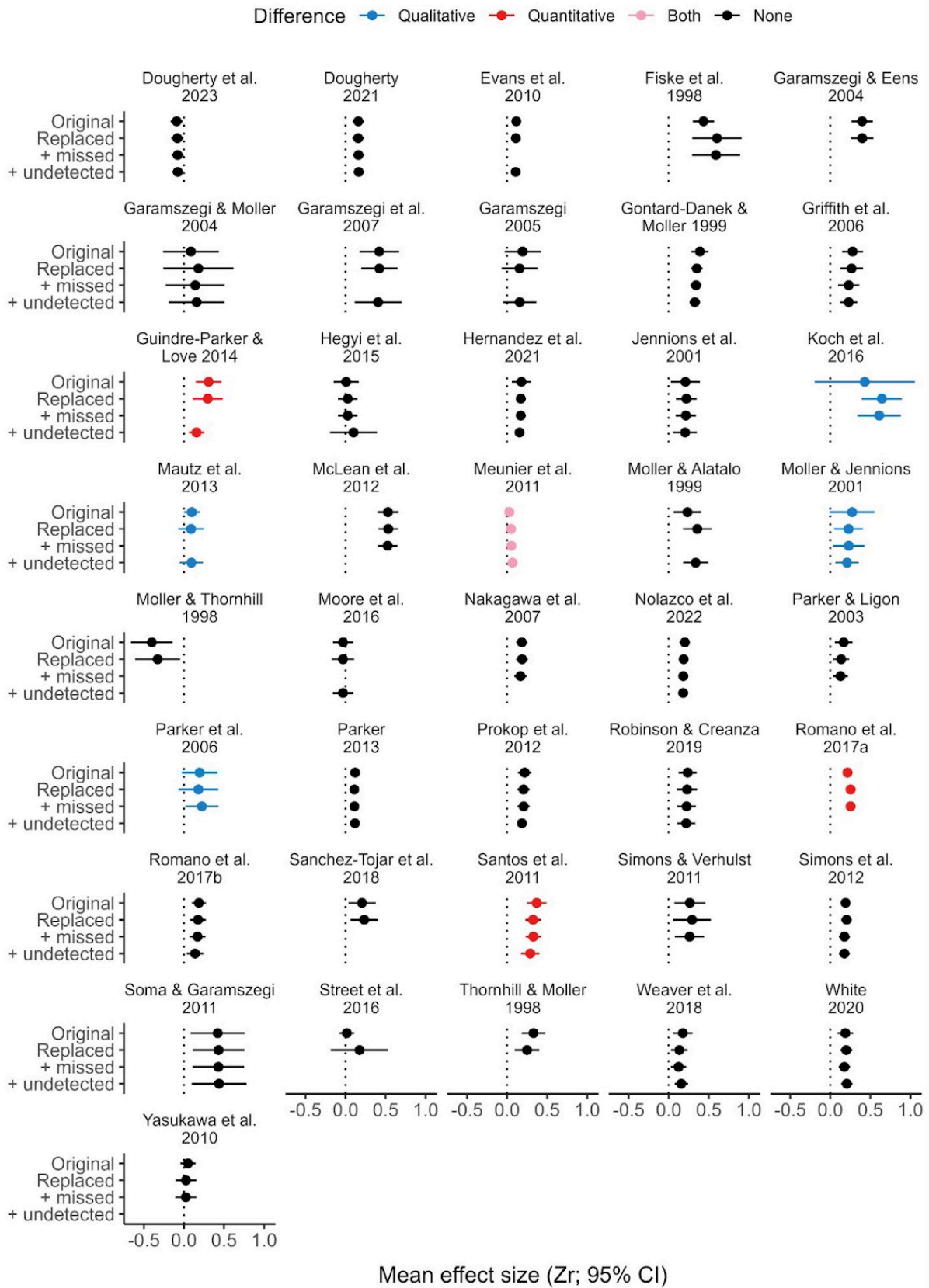
129 Figure 4. Proportion of missed effect sizes from all relevant effect sizes re-extracted from
 130 verified empirical studies (A), proportion of undetected studies from the minimum number of

131 empirical studies that each meta-analysis should have included (B), and the relationship
132 between these variables (C). Vertical red lines in panels A and B represent the average
133 proportion of missed effect sizes from all relevant re-extracted effect sizes and the minimum
134 proportion of undetected studies, respectively. The red line in panel C represents the fit of linear
135 regression between variables with its 95% confidence interval as the shaded area.

136

137 *Reproducibility of results*

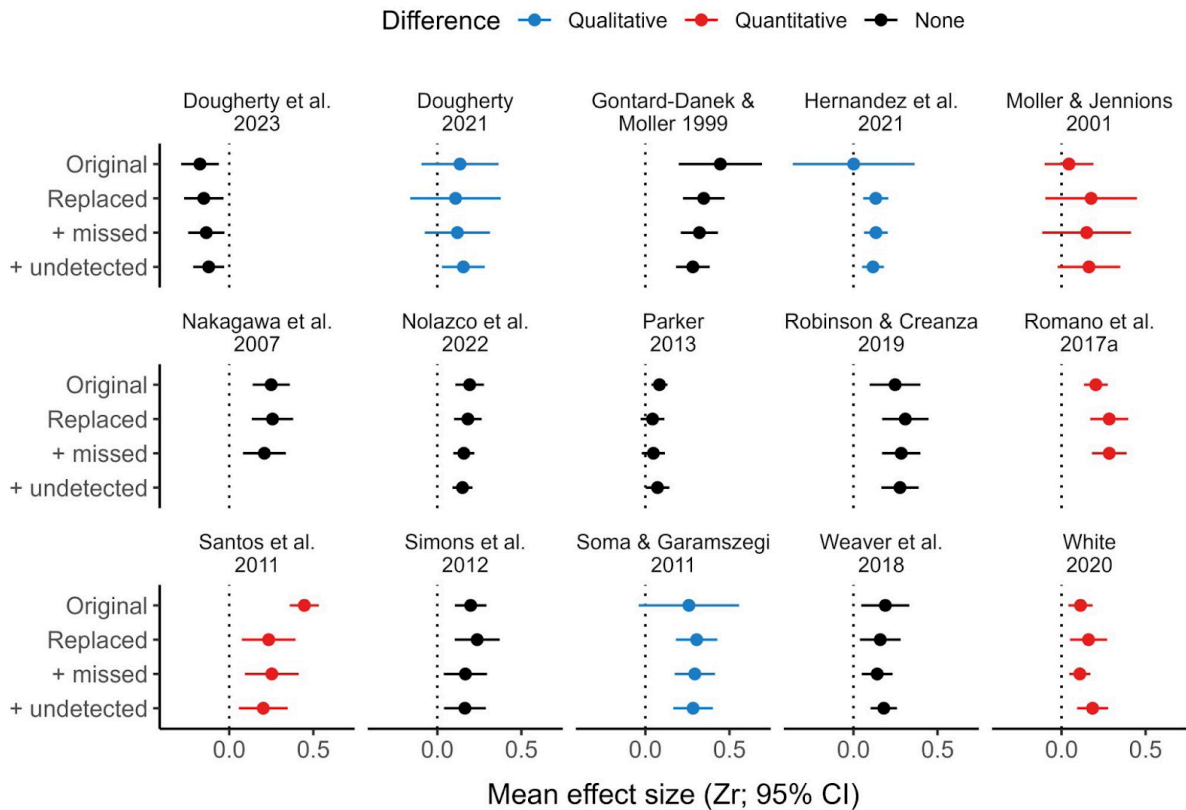
138 Using data from both verified and unverified empirical studies, we found that mean effect sizes
139 from analyses of originally reported datasets and those from analyses incorporating all relevant
140 re-extracted data (matched, missed data, and undetected studies) were qualitative and
141 quantitatively distinct for 12.2% (5 out of 41) and 9.7% (4 out of 41) of meta-analyses,
142 respectively (both types of differences occurred for one meta-analysis; Figure 5, but see Figure
143 S6). Similar comparisons using only data from verified empirical studies increased the
144 occurrence of both qualitative differences (20%, 3 out of 15) and quantitative differences
145 (26.7%, 4 out of 15; Figure 6, but see Figure S7). Nonetheless, differences in magnitude
146 between mean effect sizes were often small (Figure S3). Furthermore, heterogeneity varied
147 only slightly among fitted models (with perhaps one exception; Figure S4).



148

149 Figure 5. Mean effect sizes from up to four distinct meta-analytical models for each of 41 meta-
 150 analyses, using data from all empirical studies (both verified and unverified by us, see details

151 in-text). Dotted lines highlight zero. Comparisons were made between the first and the last
 152 result shown within each subplot, with qualitative differences representing a change in
 153 interpretation (positive *vs.* not different from zero or vice-versa) and quantitative differences
 154 representing statistical differences (absolute *z*-score greater than 1.96).
 155



156
 157 Figure 6. Mean effect sizes from up to four distinct meta-analytical models for each of 15 meta-
 158 analyses, using only data from empirical studies that we verified (see details in-text). Dotted
 159 lines highlight zero. Comparisons were made between the first and the last result shown within
 160 each subplot, with qualitative differences representing a change in interpretation (positive *vs.*
 161 not different from zero or vice-versa) and quantitative differences representing statistical
 162 differences (absolute *z*-score greater than 1.96).
 163

164 **Discussion**

165 As the very first case study in ecology and evolution, we assessed multiple aspects related to
166 the reliability of meta-analyses that focus on the relationship between sexual signals and
167 various proxies. First, we were able to precisely replicate meta-analyses' reported mean effect
168 sizes in most but not all cases. Second, we found that effect sizes reported by meta-analyses
169 and those we re-extracted did not overlap in multiple cases. Third, we observed that some
170 relevant data, both from included and undetected studies, were missing from certain meta-
171 analyses. Fourth, incorporating such missed relevant data into further analyses, in addition to
172 replacing reported data from verified empirical studies with re-extracted data, sometimes
173 produced distinct meta-analytical results. However, the discrepancies we found tended to be
174 small in magnitude and rarely generated distinct interpretations from the ones originally
175 reported. Thus, we argue that the meta-analyses we evaluated are fairly reliable, although this
176 is subjective to the degree of accuracy expected. Below we discuss the impacts of our findings
177 and provide recommendations to further improve the reliability of meta-analyses in ecology
178 and evolution.

179 Most of our results are based on comparisons between mean effect sizes (Figures 2, 5,
180 and 6) using categorical interpretations (i.e. negative, not different from zero, or positive;
181 indicating qualitative differences) and statistical calculations (indicating quantitative
182 differences). We detected qualitative differences in only a handful of these comparisons, most
183 of them resulting from differences in width of mean effect sizes' confidence intervals rather
184 than in their value. Similarly, we detected few quantitative differences when comparing mean
185 effect sizes that ideally would have remained identical across analyses. Still, these
186 discrepancies are only relevant if they impact the interpretation of results from original meta-
187 analyses. Ecology and evolution researchers rarely mention the magnitude of mean effect sizes
188 from meta-analyses, focusing their attention on the existence or absence of effects or
189 relationships (Pollo et al. *in prep.*). This means that quantitative differences alone would

190 probably be unimportant for most researchers (unless hypotheses rely on effect size
191 comparisons, as in ref. ⁵⁸), while qualitative differences would affect how results are perceived.
192 However, even though interpreting results based on the inclusion of zero by confidence
193 intervals is a common practice, it can generate misleading conclusions⁵⁹. For instance, the
194 lower confidence bound for the mean effect size of ref. ⁴⁵ was reported as -0.01 but was slightly
195 above zero in all our analyses, while the magnitude of this study's mean effect size was reported
196 as 0.2 and ranged from 0.19 to 0.21 in our analyses (Figures 2 and 5). Therefore, we conclude
197 that results from ref. ⁴⁵, as well those from meta-analyses with similar minute discrepancies,
198 should be deemed reproducible despite being different to the ones we obtained when comparing
199 the inclusion of zero by confidence intervals. Importantly, statistical definitions of replicability
200 and reproducibility are highly debatable, meaning that conclusions based on result comparisons
201 become inevitably subjective to a certain degree^{60,61}.

202 Some effect sizes reported in meta-analyses were very different from the ones we re-
203 extracted (Figure 3). A portion of these discrepancies could be explained by using distinct effect
204 size calculations and/or from extracting data from distinct sources within empirical studies (e.g.
205 raw data from a figure vs. statistics reported in-text). Yet, we noticed that numerous effect sizes
206 reported were similar in value but opposite in direction to the ones we re-extracted. Although
207 this did not strongly affect differences among estimated mean effect sizes, effect sizes with the
208 wrong direction can be particularly dangerous as they should be more impactful than simply
209 imprecise ones, at least for large effect sizes. This emphasises the importance in meta-analyses
210 of (1) establishing a coherent rationale to ascertain direction of effect sizes, (2) reporting
211 directionality decisions in detail, and (3) remaining vigilant during data extraction.
212 Additionally, we recommend cross-checking data extractions (i.e. independent verification of
213 the data by someone that did not extract them) to increase the chances of spotting and correcting
214 mistakes, including those related to the direction of effect sizes.

215 We observed that not all relevant data from verified empirical studies were used by meta-
216 analyses (Figure 4). Although we cannot ascertain how or why meta-analyses' authors missed
217 relevant data from studies they extracted data from, the reason that some empirical studies
218 remained undetected might be more easily explained. While it is possible that meta-analyses
219 could have employed suboptimal searches, empirical studies commonly neglect to inform their
220 entire scope and results in their title, abstract, and key words, which are used to retrieve and
221 screen studies⁶². Thus, even though authors of meta-analyses should follow certain guidelines
222 to build effective searches (e.g. ref. ⁶³), meta-analysts may unfortunately miss relevant studies
223 despite their best efforts. Conversely, authors of empirical studies should be mindful of how
224 search engines work, crafting their title, abstract, and keywords to enhance the findability of
225 their work^{62,64}.

226 The goal of our study was to evaluate the reliability of meta-analyses beyond
227 transparency, yet our results might be affected by transparency issues found in the meta-
228 analyses we assessed. First, matching reported and re-extracted data points proved to be a
229 difficult task because meta-analyses usually lack details on their extracted data. For instance,
230 the sexual signal and proxy for each data point was only vaguely described in most meta-
231 analyses. Furthermore, no meta-analysis in our dataset reported the location of the extracted
232 data within empirical studies (e.g. which page, table, figure, etc). Second, meta-analyses were
233 often poorly transparent with their inclusion criteria. We tried to comprehend vague
234 information, incorporate omitted criteria, and ignore contradictions when matching re-
235 extracted to reported datasets (see Supplementary information S4). Nonetheless, our decisions
236 might have affected the amount of missing data and of undetected studies by each meta-
237 analysis. Ultimately, this could have influenced our findings related to the reproducibility of
238 mean effect sizes.

239 In addition to the recommendations we already mentioned (e.g. cross-checking of data
240 extracted), we urge meta-analysts to provide all possible details on the data they collect. For
241 instance, mentioning which exact measurements were sought are necessary instead of simply
242 mentioning umbrella terms (e.g. condition-dependence by ref. ²⁴). The location of the
243 information is also crucial: readers should not have to examine datasets to find important
244 details. Instead, summarised details should be in the manuscript or, less preferably, in the
245 supplementary material, but always in a readable format (figures, simplified tables, in-text, not
246 in spreadsheets). Furthermore, we reiterate recommendations by ref. ⁵, such as providing the
247 within-text source for each data point extracted and the equations used to calculate effect sizes
248 (along with assumptions and transformations utilised). We summarise our recommendations in
249 Table S4.

250

251 **Methods**

252 This manuscript is part of a larger research project that uses data from specific meta-analyses
253 (see subsections below). Our methodology, summarised in Figure 1, was described in our pre-
254 registration⁶⁵, and we adhered to it as much as possible (see changes in Supplementary
255 information S1). We report author contributions using MeRIT guidelines⁶⁶ and the CRediT
256 statement⁶⁷.

257

258 *Reported dataset*

259 A recent systematic map identified the existence of 151 meta-analyses on topics related to
260 sexual selection⁷, 59 of them focusing on questions associated with “pre-copulatory sexual
261 traits” (i.e. sexual signals)⁷. In November 2023, PP selected 44 meta-analyses from this set,
262 specifically the ones examining the relationship between sexual signals and distinct conditions,
263 fitness proxies, or individual traits (hereby *proxies*; see Table S1). PP also included another

264 meta-analysis examining the relationship between ornament expression and parasite load¹⁴,
265 which was published after searches were conducted in ref. ⁷. This resulted in a total of 45
266 eligible meta-analyses to fulfil our objectives (refs. ^{14–57,68}).

267 PP gathered the data reported by these meta-analyses from their tables, appendices,
268 supplementary files, and occasionally from direct correspondence with their authors. We could
269 not obtain data for four eligible meta-analyses as their data were unreported^{55,56,68} or poorly
270 described⁵⁷. Altogether, the remaining 41 meta-analyses from which PP extracted data^{14–54}
271 yielded 6,773 data points (see *Analyses* section).

272 PP filtered the collated dataset for most analyses conducted here (see *Analyses* section
273 and Supplementary information S2). This was done because not all data collected from meta-
274 analyses were relevant to objectives of other parts of the larger research project (e.g. meta-
275 meta-analysis of sexual signals; see ref. ⁶⁵). Following this filtering process, the number of data
276 points in the collated dataset decreased to 5,496.

277

278 *Re-extractions*

279 Many meta-analyses included in the collated dataset extracted data from the same empirical
280 studies. More specifically, after filtering, PP detected that a quarter of empirical studies in the
281 dataset (314 out of 1237) were used as sources by at least two different meta-analyses (hereby
282 *duplicates*, Figure S1). PP selected a subset of these duplicates for data re-extraction to reduce
283 sampling effort (generating a greater sample size of data points for each meta-analysis
284 assessed), prioritising certain empirical studies (see details in Supplementary information S3).
285 In total, this process produced a set of 249 empirical studies for re-extraction. However, we
286 could not access the full-text of three of these studies, so our sample of studies for re-extraction
287 was reduced to 246 (hereby *verified* empirical studies; refs. ^{69–314}).

288 We extracted all data on the relationship between sexual signals and proxies from
289 verified empirical studies (see details in Supplementary information S2), blinded to which
290 exact data points were extracted for each of the meta-analyses that included a given study. More
291 specifically, PP extracted data from 59.3% of the selected empirical studies, while SN, YY,
292 AM, RCMR, and ML respectively extracted data from other 15.5%, 7.7%, 6.5%, 5.7%, and
293 5.3% studies. PP then cross-checked all data extractions done by other authors (40.7% of
294 studies). Importantly, authors never re-extracted data from empirical studies they participated
295 in. We extracted data required to calculate effect sizes from text, tables, supplementary
296 material, and figures (using the package metaDigitise 1.0.1³¹⁵). When empirical studies
297 reported similar results in various forms, we prioritised extractions in the following order: (1)
298 raw data (calculating estimates directly) from sources other than figures, (2) raw estimates (i.e.
299 means and correlation coefficients) from sources other than figures, (3) raw data or raw
300 estimates from figures, (4) other estimates (e.g. t , β , χ^2) regardless of their origin. Nonetheless,
301 we first prioritised data sources that showed more details. For instance, if a correlation was
302 given for all individuals in-text but a scatter plot showed the same data with dots separated by
303 age or sex, we collected data from the latter. In total, we extracted 5,703 valid data points. We
304 note that we do not claim that our re-extracted data points are more or less correct than the data
305 originally reported by meta-analyses, yet we do expect them to converge, meaning that
306 mismatches should be taken seriously given our transparent procedures.

307

308 *Matching reported data with re-extracted data*

309 PP carefully examined the inclusion criteria reported in meta-analyses to verify which re-
310 extracted data points should have been included by them. However, PP found several issues
311 with these inclusion criteria. First, the proxies and sexual signals that meta-analyses included
312 were often vague or ambiguous. For instance, ref. ¹⁵ was interested in behavioural sexual

313 signals but their dataset also included extended phenotypes (e.g. domes built by crabs). Even
314 though these extended phenotypes can be considered behavioural products, it was unclear
315 whether other similar structures (e.g. bowers, ornamented nests) were deemed valid for
316 inclusion by ref. ¹⁵. Second, we detected apparent inclusion criteria patterns in datasets of some
317 meta-analyses' that were not mentioned in-text. For example, ref. ³⁶ stated that they included
318 data on reproductive success but the only reproductive success measure in their dataset was the
319 number of fledglings, even though the studies they used for effect size extraction also contained
320 other measures (e.g. number of eggs, number of hatchlings). Third, some meta-analyses
321 outwardly contradicted their own information. For example, ref. ⁵² stated that they included
322 standardised colour metrics (hue, chroma or composite measures of those) for carotenoid-based
323 colours in adult birds, describing specific proxies in their table 2. Yet, they seemed to have
324 included data points in which (1) the sexual signal was the size of a colourful plumage (e.g.
325 patch size), (2) individuals were juveniles (including when data points were separated by age),
326 and (3) proxies other than the ones reported in-text were used (e.g. offspring size). We
327 summarised all meta-analyses' originally reported inclusion criteria, the ambiguities,
328 omissions, and contradictions we detected in them, and how we dealt with these issues for
329 matching purposes in Supplementary information S4.

330 We then attempted to match data points reported in meta-analyses with the ones we re-
331 extracted from empirical sources. PP mainly used the description of sexual signals and proxies
332 of each data point to match them with re-extracted data. When multiple data points from the
333 same empirical study had similar descriptions, we also used sample size and other additional
334 information (e.g. statistics reported, if given) for matching purposes.

335 There were three possibilities for each matching attempt. First, when both original and
336 re-extracted data points had a similar description, PP linked them by labelling the latter with
337 the ID of the former (i.e. successfully matched them). Yet, it was common to find multiple data

338 points in our re-extracted dataset that matched one or many data points from the originally
339 reported dataset (or vice-versa), so this matching was not necessarily exact (see examples in
340 Supplementary information S5). Second, there were cases in which we could not find original
341 data points with an equivalent description to relevant re-extracted ones. We assumed that these
342 data were missed or undetected by meta-analyses' authors. This allowed us to obtain two
343 aspects related to reliability: (1) the number of relevant data points that were absent in meta-
344 analyses' datasets despite being present in empirical studies reported as sources (hereby
345 *missing data*) and (2) the number of empirical studies that contained relevant data points that
346 should have been used as sources but were not (hereby *undetected studies*). Third, there were
347 cases in which data points were shown in meta-analyses' reported datasets but were absent in
348 our re-extractions. We re-checked all of these latter cases: although some data points were not
349 extracted by us because of our criteria (e.g. invalid proxy) or because we considered them
350 repeated data, most of them could simply not be found in empirical studies (see details in Table
351 S2). This could have happened if meta-analyses' authors contacted empirical authors and were
352 thus able to obtain more data than what was shown in the empirical articles (unpublished data).
353 Alternatively, meta-analyses' authors may have made mistakes during data extraction, even
354 though we cannot ascertain when this was truly the case.

355 After PP finalised the matching process, AM, ML, RCMR, SN, and YY cross-checked
356 matching decisions for five different empirical studies each. This resulted in matching
357 decisions for approximately 10% of all verified studies being cross-checked, somewhat
358 attesting the reliability of our process.

359

360 *Effect sizes*

361 Originally reported data points were given as the following effect size types: Cohen's d by refs.
362 ¹⁸ and ²⁰, logarithm of response ratio (logRR) by ref. ²⁸, and Fisher's Z_r or correlation

363 coefficients (r) by remaining meta-analytical studies. One meta-analysis in particular (ref. ³⁸)
364 only provided p -values and sample sizes, so PP calculated effect sizes from this information.
365 We transformed effect sizes reported to Zr for all analyses except the replication of general
366 results (see *Analyses* section). We also calculated Zr (along with its sampling variance) from
367 all re-extracted data points. Additionally, effect sizes reported by ref. ²⁸ did not reflect the raw
368 data that they provided (means and standard deviations), so PP re-calculated their effect sizes.
369 All equations for calculation and conversion of effect sizes are given in the Supplementary
370 information S6, while the direction rationale applied to effect sizes is detailed in Supplementary
371 information S7.

372

373 *Analyses*

374 First, we attempted to replicate meta-analyses' reported mean effect sizes using their original
375 datasets. To do so, we conducted a meta-analytical model for each meta-analytical study using
376 all of their reported effect sizes together (i.e. global model, *sensu* ref. ⁷). However, for meta-
377 analytical studies that only performed subgroup analyses, we only re-analysed the largest or
378 first reported subgroup. For instance, ref. ³⁷ only analysed the relationship between plumage
379 colour and proxies for each sex separately, so we re-analysed only their data related to females.
380 Effect sizes for these replication analyses were of the same type as results reported by meta-
381 analyses (see *Effect sizes* section). Meta-analytical models for all of our analyses contained
382 multiple random factors (see end of this section) but, for this replication analysis, we also
383 included an additional random factor if provided by authors. For example, both refs. ^{42,45} used
384 population ID as an additional random factor in their models. Yet, we were unable to include
385 these additional random factors in meta-analytic models when these variables were not
386 provided by authors with the rest of the data (e.g. experiment ID in ref. ¹⁵ and population ID in
387 ref. ⁴⁶) or when they were redundant (population ID was different for every source in ref. ¹⁶).

388 Moreover, two meta-analytical studies from our collated dataset were excluded from this
389 specific analysis: ref. ³⁸ did not provide a confidence interval for their mean effect size and
390 only data points of interest were extracted from ref. ⁵¹ (i.e. their dataset was not fully extracted).
391 We also tested for signs of publication bias in re-analysed datasets by adding the inverse of the
392 effective sample size as a moderator in meta-analytical models (alternative Egger's
393 regression)³¹⁶.

394 Second, we attempted to replicate part of individual effect sizes reported in meta-
395 analyses. To do so, we compared originally reported effect sizes with the ones we re-extracted
396 using linear models with only exactly matched data (i.e. reported data point matched to a single
397 re-extracted data point, see Supplementary information S5). In these linear models, the
398 intercept was forced to be zero, originally reported effect sizes were the response variable, and
399 re-extracted effect sizes were the predictor variable. We evaluated whether the 95% confidence
400 interval of the estimated slope in each linear model included the value 1, which would represent
401 that reported and re-extracted effect sizes are very similar. We could not compare originally
402 reported effect sizes with re-extracted effect sizes from 11 meta-analyses because they
403 contained less than six exactly matched data points (the arbitrary threshold we established for
404 this analysis), so we only reported results related to the remaining 30 meta-analyses in our
405 dataset.

406 Third, we assessed meta-analyses' reproducibility by comparing the results of four
407 slightly distinct meta-analytical models for each meta-analytical study. In model 1, we analysed
408 the filtered dataset, which contained only reported effect sizes. In model 2, we used the same
409 data as in model 1 but we replaced reported effect sizes that came from verified studies with
410 their equivalent re-extracted effect sizes (i.e. the ones that matched in description). In model 3,
411 we used the same data as in model 2 but we also added other relevant re-extracted data points
412 from verified studies that were originally missed by meta-analyses' authors (i.e. missed data).

413 In model 4, we used the same data as in model 3 but we also added all relevant effect sizes
414 from undetected studies. Differently from our analysis attempting to replicate mean effect sizes
415 (first described in this section), we conducted global models for all meta-analytical studies,
416 even those that had done only subgroup analyses. We note that the datasets used for most meta-
417 analyses remained similar across the four models described above because we did not re-extract
418 data from all empirical studies reported as sources (Figure S2). For example, ref. ¹⁵ extracted
419 data from 197 empirical studies, but we verified only 15 of these studies. This means that results
420 from these distinct models were unlikely to change for meta-analyses with proportionally few
421 verified sources. Thus, we additionally conducted the same four models described above using
422 only data from verified studies (instead of using data from both verified and unverified studies)
423 for meta-analyses with at least 15 verified sources ($N = 15$). Furthermore, we also evaluated
424 whether the inclusion of missing data and undetected studies changed the generalisability of
425 results by examining the σ (a measure of heterogeneity) from each random factor in meta-
426 analytical models.

427 When applicable, we compared estimates from meta-analytic models both qualitatively
428 and quantitatively. We first ascertained whether an estimate was positive, negative, or not
429 different from zero (if its 95% CI overlapped zero), so that estimates with distinct
430 classifications represented a qualitative difference. By contrast, a quantitative difference
431 occurred when the absolute difference between two estimates produced a z -score of more than
432 1.96 (i.e. statistically significant, two-tailed $\alpha = 0.05$), calculated as:

$$433 \quad z = \frac{X_{reported} - X_{re-analysed}}{\sqrt{se_{reported}^2 + se_{re-analysed}^2 - 2rse_{reported}se_{re-analysed}}}$$

434 where X represents the estimated mean effect size, se represents its standard error, and
435 r represents the correlation coefficient between these two groups (set as 0.8 for all main results
436 as we assume this correlation is high but not perfect; but see Figures S5, S6, and S7 for results

437 using alternative methods). Although we showed all mean effect sizes generated in the last set
438 of analyses, where we conducted up to four models per meta-analysis, we only discuss
439 comparisons between the first and last models for simplicity.

440 We conducted all analyses described above in R 4.4.0³¹⁷. Multilevel meta-analytical
441 models were fitted using the *rma.rv* function from the package *metafor* 4.6-0³¹⁸. All meta-
442 analytical models fitted contained empirical study ID, species ID (non-phylogenetic effect),
443 and phylogenetic relatedness as a random factor in meta-analytical models³¹⁹. However, we
444 removed phylogenetic relatedness from meta-analytical models related to refs. ^{18,24,31,32},
445 otherwise some of them would not converge. Phylogenetic trees were built using the packages
446 *ape* 5.8³²⁰ and *rotl* 3.1.0³²¹, which uses data from ref. ³²².

447

448 **Data and code availability**

449 All data and code used in this study are available at:
450 https://osf.io/6njem/?view_only=7b01538fb32e4f78b7130b6e8f303649.

451

452 **Author contributions**

453 Conceptualisation: PP, ML, AM, YY, SN; data curation: PP; formal analysis: PP; funding
454 acquisition: SN; investigation: PP, ML, RCMR, AM, YY, SN; methodology: PP, ML, RCMR,
455 AM, YY, SN; project administration: PP; software: PP; supervision: SN; visualisation: PP;
456 writing – original draft: PP, SN; writing – review & editing: PP, ML, RCMR, AM, YY, SN.

457

458 **Acknowledgements**

459 We thank László Zsolt Garamszegi, Sarah Guindre-Parker, América Hernández Martínez, and
460 Tim Parker for sending us data from their meta-analyses. We also thank Yong Zhi Foo for
461 contributing to the conceptualisation of this project. PP, ML, YY, and SN were supported by

462 ARC (Australian Research Council) Discovery Project grants (DP210100812 and
463 DP230101248). RCMR was supported by a grant from the Coordenação de Aperfeiçoamento
464 de Pessoal de Nível Superior (CAPES) as part of the Programa de Desenvolvimento da Pós-
465 Graduação (PDPG 88887.803974/2023-00).

466

467 **Competing interests**

468 We declare no competing interests.

469

470 **References**

- 471 1. McCallen, E. *et al.* Trends in ecology: shifts in ecological research themes over the past
472 four decades. *Front Ecol Environ* **17**, 109–116 (2019).
- 473 2. Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N. & Fei, S. Automated
474 content analysis: addressing the big literature challenge in ecology and evolution.
475 *Methods Ecol Evol* **7**, 1262–1272 (2016).
- 476 3. Koricheva, J., Gurevitch, J. & Mengersen, K. *Handbook of Meta-Analysis in Ecology*
477 *and Evolution*. (Princeton University Press, Princeton, NJ, 2013).
- 478 4. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. *Introduction to Meta-*
479 *Analysis*. (John Wiley & Sons, Hoboken, NJ, 2021).
- 480 5. Ivimey-Cook, E. R., Noble, D. W. A., Nakagawa, S., Lajeunesse, M. J. & Pick, J. L.
481 Advice for improving the reproducibility of data extraction in meta-analysis. *Res Synth*
482 *Methods* **14**, 911–915 (2023).
- 483 6. O’Dea, R. E. *et al.* Preferred reporting items for systematic reviews and meta-analyses
484 in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews* **96**,
485 1695–1722 (2021).
- 486 7. Pollo, P., Lagisz, M., Yang, Y., Culina, A. & Nakagawa, S. Synthesis of sexual selection:
487 a systematic map of meta-analyses with bibliometric analysis. *Biological Reviews* **4**,
488 (2024).
- 489 8. Stroup, D. F. *et al.* Meta-analysis of observational studies in epidemiology. *JAMA* **283**,
490 2008 (2000).
- 491 9. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting
492 systematic reviews. *BMJ* **372**, n71 (2021).

- 493 10. Kambouris, S., Wilkinson, D. P., Smith, E. T. & Fidler, F. Computationally reproducing
494 results from meta-analyses in ecology and evolutionary biology using shared code and
495 data. *PLoS One* **19**, e0300333 (2024).
- 496 11. Stubblefield, J. W. & Orzack, S. H. Resource transfers and evolution: helpful offspring
497 and sex allocation. *Theor Popul Biol* **83**, 64–81 (2013).
- 498 12. Page, M. J. *et al.* The REPRISE project: protocol for an evaluation of REProducibility
499 and Replicability In Syntheses of Evidence. *Syst Rev* **10**, 112 (2021).
- 500 13. Maassen, E., Van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A. & Wicherts,
501 J. M. Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS*
502 *One* **15**, 1–18 (2020).
- 503 14. Dougherty, L. R., Rovenolt, F., Luyet, A., Jokela, J. & Stephenson, J. F. Ornaments
504 indicate parasite load only if they are dynamic or parasites are contagious. *Evol Lett* **7**,
505 176–190 (2023).
- 506 15. Dougherty, L. R. Meta-analysis reveals that animal sexual signalling behaviour is honest
507 and resource based. *Nat Ecol Evol* **5**, 688–699 (2021).
- 508 16. Evans, S. R., Hinks, A. E., Wilkin, T. A. & Sheldon, B. C. Age, sex and beauty:
509 methodological dependence of age- and sex-dichromatism in the great tit *Parus major*.
510 *Biological Journal of the Linnean Society* **101**, 777–796 (2010).
- 511 17. Fiske, P., Rintamaki, P. T. & Karvonen, E. Mating success in lekking males: a meta-
512 analysis. *Behavioral Ecology* **9**, 328–338 (1998).
- 513 18. Garamszegi, L. Z. & Eens, M. Brain space for a learned task: strong intraspecific
514 evidence for neural correlates of singing behavior in songbirds. *Brain Res Rev* **44**, 187–
515 193 (2004).
- 516 19. Garamszegi, L. Z. & Møller, A. P. Extrapair paternity and the evolution of bird song.
517 *Behavioral Ecology* **15**, 508–519 (2004).
- 518 20. Garamszegi, L. Z. *et al.* Age-dependent expression of song in the collared flycatcher,
519 *Ficedula albicollis*. *Ethology* **113**, 246–256 (2007).
- 520 21. Garamszegi, L. Z. Bird song and parasites. *Behav Ecol Sociobiol* **59**, 167–180 (2005).
- 521 22. Gontard-Danek, M.-C. & Møller, A. P. The strength of sexual selection: a meta-analysis
522 of bird studies. *Behavioral Ecology* **10**, 476–486 (1999).
- 523 23. Griffith, S. C., Parker, T. H. & Olson, V. A. Melanin- versus carotenoid-based sexual
524 signals: is the difference really so black and red? *Anim Behav* **71**, 749–763 (2006).
- 525 24. Guindre-Parker, S. & Love, O. P. Revisiting the condition-dependence of melanin-based
526 plumage. *J Avian Biol* **45**, 29–33 (2014).
- 527 25. Hegyi, G., Kötél, D. & Laczi, M. Direct benefits of mate choice: a meta-analysis of
528 plumage colour and offspring feeding rates in birds. *Science of Nature* **102**, 62 (2015).

- 529 26. Hernández, A., Martínez-Gómez, M., Beamonte-Barrientos, R. & Montoya, B.
530 Colourful traits in female birds relate to individual condition, reproductive performance
531 and male-mate preferences: a meta-analytic approach. *Biol Lett* **17**, 20210283 (2021).
- 532 27. Jennions, M. D., Møller, A. P. & Petrie, M. Sexually selected traits and adult survival: a
533 meta-analysis. *Q Rev Biol* **76**, 3–36 (2001).
- 534 28. Koch, R. E., Wilson, A. E. & Hill, G. E. The importance of carotenoid dose in
535 supplementation studies with songbirds. *Physiological and Biochemical Zoology* **89**,
536 61–71 (2016).
- 537 29. Mautz, B. S., Møller, A. P. & Jennions, M. D. Do male secondary sexual characters
538 signal ejaculate quality? A meta-analysis. *Biological Reviews* **88**, 669–682 (2013).
- 539 30. McLean, M. J., Bishop, P. J. & Nakagawa, S. Male quality, signal reliability and female
540 choice: assessing the expectations of inter-sexual selection. *J Evol Biol* **25**, 1513–1520
541 (2012).
- 542 31. Meunier, J., Pinto, S. F., Burri, R. & Roulin, A. Eumelanin-based coloration and fitness
543 parameters in birds: a meta-analysis. *Behav Ecol Sociobiol* **65**, 559–567 (2011).
- 544 32. Møller, A. P. & Alatalo, R. V. Good-genes effects in sexual selection. *Proceedings of the*
545 *Royal Society B: Biological Sciences* **266**, 85–91 (1999).
- 546 33. Møller, A. P. & Jennions, M. How important are direct fitness benefits of sexual
547 selection? *Naturwissenschaften* **88**, 401–415 (2001).
- 548 34. Møller, A. P. & Thornhill, R. Bilateral symmetry and sexual selection: a meta-analysis.
549 *Am Nat* **151**, 174–192 (1998).
- 550 35. Moore, F. R., Shuker, D. M. & Dougherty, L. Stress and sexual signaling: a systematic
551 review and meta-analysis. *Behavioral Ecology* **27**, 363–371 (2016).
- 552 36. Nakagawa, S., Ockendon, N., Gillespie, D. O. S., Hatchwell, B. J. & Burke, T. Assessing
553 the function of house sparrows' bib size using a flexible meta-analysis method.
554 *Behavioral Ecology* **18**, 831–840 (2007).
- 555 37. Nolzco, S., Delhey, K., Nakagawa, S. & Peters, A. Ornaments are equally informative
556 in male and female birds. *Nat Commun* **13**, 5917 (2022).
- 557 38. Parker, T. H. & Ligon, J. D. Female mating preferences in red junglefowl: a meta-
558 analysis. *Ethol Ecol Evol* **15**, 63–72 (2003).
- 559 39. Parker, T. H. What do we really know about the signalling role of plumage colour in
560 blue tits? A case study of impediments to progress in evolutionary biology. *Biological*
561 *Reviews* **88**, 511–536 (2013).
- 562 40. Parker, T. H., Barr, I. R. & Griffith, S. C. The blue tit's song is an inconsistent signal of
563 male condition. *Behavioral Ecology* **17**, 1029–1040 (2006).

- 564 41. Prokop, Z. M., Michalczyk, Ł., Drobniak, S. M., Herdegen, M. & Radwan, J. Meta-
565 analysis suggests choosy females get sexy sons more than ‘good genes’. *Evolution* **66**,
566 2665–2673 (2012).
- 567 42. Robinson, C. M. & Creanza, N. Species-level repertoire size predicts a correlation
568 between individual song elaboration and reproductive success. *Ecol Evol* **9**, 8362–8377
569 (2019).
- 570 43. Romano, A., Costanzo, A., Rubolini, D., Saino, N. & Møller, A. P. Geographical and
571 seasonal variation in the intensity of sexual selection in the barn swallow *Hirundo*
572 *rustica*: a meta-analysis. *Biological Reviews* **92**, 1582–1600 (2017).
- 573 44. Romano, A., Saino, N. & Møller, A. P. Viability and expression of sexual ornaments in
574 the barn swallow *Hirundo rustica*: a meta-analysis. *J Evol Biol* **30**, 1929–1935 (2017).
- 575 45. Sánchez-Tójar, A. *et al.* Meta-analysis challenges a textbook example of status
576 signalling and demonstrates publication bias. *Elife* **7**, 1–26 (2018).
- 577 46. Santos, E. S. A., Scheck, D. & Nakagawa, S. Dominance and plumage traits: meta-
578 analysis and metaregression analysis. *Anim Behav* **82**, 3–19 (2011).
- 579 47. Simons, M. J. P. & Verhulst, S. Zebra finch females prefer males with redder bills
580 independent of song rate—a meta-analysis. *Behavioral Ecology* **22**, 755–762 (2011).
- 581 48. Simons, M. J. P., Cohen, A. A. & Verhulst, S. What does carotenoid-dependent
582 coloration tell? Plasma carotenoid level signals immunocompetence and oxidative stress
583 state in birds—a meta-analysis. *PLoS One* **7**, e43088 (2012).
- 584 49. Soma, M. & Garamszegi, L. Z. Rethinking birdsong evolution: meta-analysis of the
585 relationship between song complexity and reproductive success. *Behavioral Ecology* **22**,
586 363–371 (2011).
- 587 50. Street, S. E., Cross, C. P. & Brown, G. R. Exaggerated sexual swellings in female
588 nonhuman primates are reliable signals of female fertility and body condition. *Anim*
589 *Behav* **112**, 203–212 (2016).
- 590 51. Thornhill, R. & Møller, A. P. The relative importance of size and asymmetry in sexual
591 selection. *Behavioral Ecology* **9**, 546–551 (1998).
- 592 52. Weaver, R. J., Santos, E. S. A., Tucker, A. M., Wilson, A. E. & Hill, G. E. Carotenoid
593 metabolism strengthens the link between feather coloration and individual quality. *Nat*
594 *Commun* **9**, 73 (2018).
- 595 53. White, T. E. Structural colours reflect individual quality: a meta-analysis. *Biol Lett* **16**,
596 20200001 (2020).
- 597 54. Yasukawa, K., Enstrom, D. A., Parker, P. G. & Jones, T. C. Male Red-winged Blackbirds
598 with experimentally dulled epaulets experience no disadvantage in sexual selection. *J*
599 *Field Ornithol* **81**, 31–41 (2010).

- 600 55. Hamilton, W. J. & Poulin, R. The Hamilton and Zuk hypothesis revisited: a meta-
601 analytical approach. *Behaviour* **134**, 299–320 (1997).
- 602 56. Macedo-Rego, R. C., Jennions, M. D. & Santos, E. S. A. Larger females and larger males
603 produce more offspring, regardless of the social mating system: a meta-analysis on
604 animals. in *When and how do episodes of sexual selection occur in animal species? A*
605 *meta-analytical investigation* (ed. Macedo-Rego, R. C.) 148–187 (Universidade de São
606 Paulo, São Paulo, Brazil, 2020).
- 607 57. Møller, A. P., Christe, P. & Lux, E. Parasitism, host immune function, and sexual
608 selection. *Q Rev Biol* **74**, 3–20 (1999).
- 609 58. Pollo, P., Nakagawa, S. & Kasumovic, M. M. The better, the choosier: a meta-analysis
610 on interindividual variation of male mate choice. *Ecol Lett* **25**, 1305–1322 (2022).
- 611 59. Amrhein, V., Greenland, S. & McShane, B. Scientists rise up against statistical
612 significance. *Nature* **567**, 305–307 (2019).
- 613 60. Hedges, L. V. & Schauer, J. M. More than one replication study is needed for
614 unambiguous tests of replication. *Journal of Educational and Behavioral Statistics* **44**,
615 543–570 (2019).
- 616 61. Schauer, J. M. & Hedges, L. V. Assessing heterogeneity and power in replications of
617 psychological experiments. *Psychol Bull* **146**, 701–719 (2020).
- 618 62. Pottier, P. *et al.* Title, abstract and keywords: a practical guide to maximize the visibility
619 and impact of academic papers. *Proceedings of the Royal Society B: Biological Sciences*
620 **291**, (2024).
- 621 63. Foo, Y. Z., O’Dea, R. E., Koricheva, J., Nakagawa, S. & Lagisz, M. A practical guide to
622 question formation, systematic searching and study screening for literature reviews in
623 ecology and evolution. *Methods Ecol Evol* **12**, 1705–1720 (2021).
- 624 64. Schilhan, L., Kaier, C. & Lackner, K. Increasing visibility and discoverability of
625 scholarly publications with academic search engine optimization. *Insights the UKSG*
626 *journal* **34**, 1–16 (2021).
- 627 65. Pollo, P. *et al.* Pre-registration for a methodological evaluation and a meta-meta-analysis
628 of sexual signals. Preprint at <https://doi.org/10.17605/OSF.IO/HR2YK> (2024).
- 629 66. Nakagawa, S. *et al.* Method Reporting with Initials for Transparency (MeRIT) promotes
630 more granularity and accountability for author contributions. *Nat Commun* **14**, 1788
631 (2023).
- 632 67. McNutt, M. K. *et al.* Transparency in authors’ contributions and responsibilities to
633 promote integrity in scientific publication. *Proceedings of the National Academy of*
634 *Sciences* **115**, 2557–2560 (2018).
- 635 68. Leung, B. & Forbes, M. R. Fluctuating asymmetry in relation to stress and fitness:
636 effects of trait type as revealed by meta-analysis. *Écoscience* **3**, 400–413 (1996).

- 637 69. Postma, E. & Gienapp, P. Origin-related differences in plumage coloration within an
638 island population of great tits (*Parus major*). *Can J Zool* **87**, 1–7 (2009).
- 639 70. Gil, D. & Slater, P. J. B. Multiple song repertoire characteristics in the willow warbler
640 (*Phylloscopus trochilus*): correlations with female choice and offspring viability. *Behav*
641 *Ecol Sociobiol* **47**, 319–326 (2000).
- 642 71. Howard, R. D. The influence of sexual selection and interspecific competition on
643 mockingbird song (*Mimus polyglottos*). *Evolution* **28**, 428–438 (1974).
- 644 72. Horn, A. G., Dickinson, T. E. & Falls, J. B. Male quality and song repertoires in western
645 meadowlarks (*Sturnella neglecta*). *Can J Zool* **71**, 1059–1061 (1993).
- 646 73. Ferns, P. N. & Hinsley, S. A. Carotenoid plumage hue and chroma signal different
647 aspects of individual and habitat quality in tits. *Ibis* **150**, 152–159 (2008).
- 648 74. Searcy, W. A. Song repertoire size and female preferences in song sparrows. *Behav Ecol*
649 *Sociobiol* **14**, 281–286 (1984).
- 650 75. Forstmeier, W., Kempenaers, B., Meyer, A. & Leisler, B. A novel song parameter
651 correlates with extra-pair paternity and reflects male longevity. *Proceedings of the Royal*
652 *Society B: Biological Sciences* **269**, 1479–1485 (2002).
- 653 76. Hiebert, S. M., Stoddard, P. K. & Arcese, P. Repertoire size, territory acquisition and
654 reproductive success in the song sparrow. *Anim Behav* **37**, 266–273 (1989).
- 655 77. Lampe, H. M. & Espmark, Y. O. Mate choice in pied flycatchers *Ficedula hypoleuca*:
656 can females use song to find high-quality males and territories? *Ibis* **145**, E24–E33
657 (2002).
- 658 78. Bouwman, K. M., van Dijk, R. E., Wilmenga, J. J. & Komdeur, J. Older male reed
659 buntings are more successful at gaining extrapair fertilizations. *Anim Behav* **73**, 15–27
660 (2007).
- 661 79. Hansen, A. J. & Rohwer, S. Coverable badges and resource defence in birds. *Anim Behav*
662 **34**, 69–76 (1986).
- 663 80. Røskaft, E. & Rohwer, S. An experimental study of the function of the red epaulettes
664 and the black body colour of male red-winged blackbirds. *Anim Behav* **35**, 1070–1077
665 (1987).
- 666 81. Yasukawa, K., Butler, L. K. & Enstrom, D. A. Intersexual and intrasexual consequences
667 of epaulet colour in male red-winged blackbirds: an experimental approach. *Anim Behav*
668 **77**, 531–540 (2009).
- 669 82. Aweida, M. K. Repertoires, territory size and mate attraction in western meadowlarks.
670 *Condor* **97**, 1080–1083 (1995).
- 671 83. Mora, A. R., Meniri, M., Glauser, G., Vallat, A. & Helfenstein, F. Badge size reflects
672 sperm oxidative status within social groups in the house sparrow *Passer domesticus*.
673 *Front Ecol Evol* **4**, 1–10 (2016).

- 674 84. Garamszegi, L. Z., Heylen, D., Møller, A. P., Eens, M. & De Lope, F. Age-dependent
675 health status and song characteristics in the barn swallow. *Behavioral Ecology* **16**, 580–
676 591 (2005).
- 677 85. Balsby, T. J. S. Song activity and variability in relation to male quality and female choice
678 in whitethroats *Sylvia communis*. *J Avian Biol* **31**, 56–62 (2000).
- 679 86. Eens, M., Pinxten, R. & Verheyen, R. F. Male song as a cue for mate choice in the
680 European starling. *Behaviour* **116**, 210–238 (1991).
- 681 87. Dreiss, A. *et al.* Sex ratio and male sexual characters in a population of blue tits, *Parus*
682 *caeruleus*. *Behavioral Ecology* **17**, 13–19 (2006).
- 683 88. Foerster, K., Poesel, A., Kunc, H. & Kempenaers, B. The natural plasma testosterone
684 profile of male blue tits during the breeding season and its relation to song output. *J*
685 *Avian Biol* **33**, 269–275 (2002).
- 686 89. Poesel, A., Foerster, K. & Kempenaers, B. The dawn song of the blue tit *Parus caeruleus*
687 and its role in sexual selection. *Ethology* **107**, 521–531 (2001).
- 688 90. Doutrelant, C., Blondel, J., Perret, P. & Lambrechts, M. M. Blue tit song repertoire size,
689 male quality and interspecific. *J Avian Biol* **31**, 360–366 (2000).
- 690 91. Piersma, T. *et al.* Breeding plumage honestly signals likelihood of tapeworm infestation
691 in females of a long-distance migrating shorebird, the bar-tailed godwit. *Zoology* **104**,
692 41–48 (2001).
- 693 92. Conner, R. N., Anderson, M. E. & Dickson, J. G. Relationships among territory size,
694 habitat, song, and nesting success of northern cardinals. *Auk* **103**, 23–31 (1986).
- 695 93. Bijmens, L. Blue tit *parus caeruleus* song in relation to survival, reproduction and
696 biometry. *Bird Study* **35**, 61–67 (1988).
- 697 94. Freeman-Gallant, C. R. *et al.* Oxidative damage to DNA related to survivorship and
698 carotenoid-based sexual ornamentation in the common yellowthroat. *Biol Lett* **7**, 429–
699 432 (2011).
- 700 95. Freeman-Gallant, C. R. *et al.* Sexual selection, multiple male ornaments, and age- and
701 condition-dependent signaling in the common yellowthroat. *Evolution* **64**, 1007–1017
702 (2010).
- 703 96. Byers, B. E. Extrapair paternity in chestnut-sided warblers is correlated with consistent
704 vocal performance. *Behavioral Ecology* **18**, 130–136 (2007).
- 705 97. Rehsteiner, U., Geisser, H. & Reyer, H. U. Singing and mating success in water pipits:
706 one specific song element makes all the difference. *Anim Behav* **55**, 1471–1481 (1998).
- 707 98. Marshall, R. C., Buchanan, K. L. & Catchpole, C. K. Song and female choice for
708 extrapair copulations in the sedge warbler, *Acrocephalus schoenobaenus*. *Anim Behav*
709 **73**, 629–635 (2007).

- 710 99. Buchanan, K. L. & Catchpole, C. K. Female choice in the sedge warbler, *Acrocephalus*
711 *schoenobaenus*: multiple cues from song and territory quality. *Proceedings of the Royal*
712 *Society B: Biological Sciences* **264**, 521–526 (1997).
- 713 100. Bell, B. D., Borowiec, M., Lontkowski, J. & Pledger, S. Short records of marsh warbler
714 (*Acrocephalus palustris*) song provide indices that correlate with nesting success. *J*
715 *Ornithol* **145**, 8–15 (2004).
- 716 101. Hadfield, J. D. *et al.* Direct versus indirect sexual selection: Genetic basis of colour, size
717 and recruitment in a wild bird. *Proceedings of the Royal Society B: Biological Sciences*
718 **273**, 1347–1353 (2006).
- 719 102. Catchpole, C. K. Song repertoires and reproductive success in the great reed warbler
720 *Acrocephalus arundinaceus*. *Behav Ecol Sociobiol* **19**, 439–445 (1986).
- 721 103. Zirpoli, J. A., Black, J. M. & Gabriel, P. O. Parasites and plumage in Steller’s jays: an
722 experimental field test of the parasite-mediated handicap hypothesis. *Ethol Ecol Evol*
723 **25**, 103–116 (2013).
- 724 104. Senar, J. C., Figuerola, J. & Pascual, J. Brighter yellow blue tits make better parents.
725 *Proceedings of the Royal Society B: Biological Sciences* **269**, 257–261 (2002).
- 726 105. Zuk, M., Kim, T., Robinson, S. I. & Johnsen, T. S. Parasites influence social rank and
727 morphology, but not mate choice, in female red junglefowl, *Gallus gallus*. *Anim Behav*
728 **56**, 493–499 (1998).
- 729 106. Grindstaff, J. L., Lovern, M. B., Burtka, J. L. & Hallmark-Sharber, A. Structural
730 coloration signals condition, parental investment, and circulating hormone levels in
731 Eastern bluebirds (*Sialia sialis*). *Journal of Comparative Physiology A* **198**, 625–637
732 (2012).
- 733 107. Setchel, J. M. & Wickings, E. J. Sexual swelling in mandrills (*Mandrillus sphinx*): a test
734 of the reliable indicator hypothesis. *Behavioral Ecology* **15**, 438–445 (2004).
- 735 108. Rigaille, L., Higham, J. P., Lee, P. C., Blin, A. & Garcia, C. Multimodal sexual signaling
736 and mating behavior in olive baboons (*Papio anubis*). *Am J Primatol* **75**, 774–787
737 (2013).
- 738 109. Möhle, U., Heistermann, M., Dittami, J., Reinberg, V. & Hodges, J. K. Patterns of
739 anogenital swelling size and their endocrine correlates during ovulatory cycles and early
740 pregnancy in free-ranging barbary macaques (*Macaca sylvanus*) of Gibraltar. *Am J*
741 *Primatol* **66**, 351–368 (2005).
- 742 110. Huchard, E. *et al.* Can fertility signals lead to quality signals? Insights from the evolution
743 of primate sexual swellings. *Proceedings of the Royal Society B: Biological Sciences*
744 **276**, 1889–1897 (2009).
- 745 111. Brenowitz, E. A., Nalls, B., Wingfield, J. C. & Kroodsma, D. E. Seasonal changes in
746 avian song nuclei without seasonal changes in song repertoire. *Journal of Neuroscience*
747 **11**, 1367–1374 (1991).

- 748 112. Airey, D. C., Buchanan, K. L., Szekely, T., Catchpole, C. K. & DeVoogd, T. J. Song,
749 sexual selection, and a song control nucleus (HVC) in the brains of European sedge
750 warblers. *J Neurobiol* **44**, 1–6 (2000).
- 751 113. Brenowitz, E. A., Lent, K. & Kroodsma, D. E. Brain space for learned song in birds
752 develops independently of song learning. *Journal of Neuroscience* **15**, 6281–6286
753 (1995).
- 754 114. Molnár, O., Bajer, K., Török, J. & Herczeg, G. Individual quality and nuptial throat
755 colour in male European green lizards. *J Zool* **287**, 233–239 (2012).
- 756 115. Doucet, S. M. & Montgomerie, R. Multiple sexual ornaments in satin bowerbirds:
757 ultraviolet plumage and bowers signal different aspects of male quality. *Behavioral*
758 *Ecology* **14**, 503–509 (2003).
- 759 116. Peters, A., Kurvers, R. H. J. M., Roberts, M. L. & Delhey, K. No evidence for general
760 condition-dependence of structural plumage colour in blue tits: an experiment. *J Evol*
761 *Biol* **24**, 976–987 (2011).
- 762 117. Roberts, M. L., Ras, E. & Peters, A. Testosterone increases UV reflectance of sexually
763 selected crown plumage in male blue tits. *Behavioral Ecology* **20**, 535–541 (2009).
- 764 118. Peters, A., Delhey, K., Goymann, W. & Kempenaers, B. Age-dependent association
765 between testosterone and crown UV coloration in male blue tits (*Parus caeruleus*).
766 *Behav Ecol Sociobiol* **59**, 666–673 (2006).
- 767 119. Vinkler, M., Schnitzer, J., Munclinger, P. & Albrecht, T. Phytohaemagglutinin skin-
768 swelling test in scarlet rosefinch males: low-quality birds respond more strongly. *Anim*
769 *Behav* **83**, 17–23 (2012).
- 770 120. Merilä, J., Sheldon, B. C. & Lindström, K. Plumage brightness in relation to
771 haematozoan infections in the greenfinch *Carduelis chloris*: bright males are a good bet.
772 *Ecoscience* **6**, 12–18 (1999).
- 773 121. Martínez-Padilla, J., Mougeot, F., Pérez-Rodríguez, L. & Bortolotti, G. R. Nematode
774 parasites reduce carotenoid-based signalling in male red grouse. *Biol Lett* **3**, 161–164
775 (2007).
- 776 122. Chui, C. K. S., McGraw, K. J. & Doucet, S. M. Carotenoid-based plumage coloration in
777 golden-crowned kinglets *Regulus satrapa*: pigment characterization and relationships
778 with migratory timing and condition. *J Avian Biol* **42**, 309–322 (2011).
- 779 123. Biard, C., Saulnier, N., Gaillard, M. & Moreau, J. Carotenoid-based bill colour is an
780 integrative signal of multiple parasite infection in blackbird. *Naturwissenschaften* **97**,
781 987–995 (2010).
- 782 124. Fenoglio, S., Cucco, M., Fracchia, L., Martinotti, M. G. & Malacarne, G. Shield colours
783 of the moorhen are differently related to bacterial presence and health parameters. *Ethol*
784 *Ecol Evol* **16**, 171–180 (2004).

- 785 125. Rémy, A., Grégoire, A., Perret, P. & Doutrelant, C. Mediating male-male interactions:
786 the role of the UV blue crest coloration in blue tits. *Behav Ecol Sociobiol* **64**, 1839–1847
787 (2010).
- 788 126. Korsten, P., Dijkstra, T. H. & Komdeur, J. Is UV signalling involved in male-male
789 territorial conflict in the blue tit (*Cyanistes caeruleus*)? A new experimental approach.
790 *Behaviour* **144**, 447–470 (2007).
- 791 127. Emery, M. A. & Whitten, P. L. Size of sexual swellings reflects ovarian function in
792 chimpanzees (*Pan troglodytes*). *Behav Ecol Sociobiol* **54**, 340–351 (2003).
- 793 128. Peters, A., Delhey, K., Johnsen, A. & Kempenaers, B. The condition-dependent
794 development of carotenoid-based and structural plumage in nestling blue tits: Males and
795 females differ. *American Naturalist* **169**, (2007).
- 796 129. Hadfield, J. D. & Owens, I. P. F. Strong environmental determination of a carotenoid-
797 based plumage trait is not mediated by carotenoid availability. *J Evol Biol* **19**, 1104–
798 1114 (2006).
- 799 130. Dunn, P. O., Garvin, J. C., Whittingham, L. A., Freeman-Gallant, C. R. & Hasselquist,
800 D. Carotenoid and melanin-based ornaments signal similar aspects of male quality in
801 two populations of the common yellowthroat. *Funct Ecol* **24**, 149–158 (2010).
- 802 131. Figuerola, J., Muñoz, E., Gutiérrez, R. & Ferrer, D. Blood parasites, leucocytes and
803 plumage brightness in the ciril bunting, *Emberiza cirilus*. *Funct Ecol* **13**, 594–601 (1999).
- 804 132. Biard, C., Surai, P. F. & Møller, A. P. An analysis of pre- and post-hatching maternal
805 effects mediated by carotenoids in the blue tit. *J Evol Biol* **20**, 326–339 (2007).
- 806 133. Isaksson, C., McLaughlin, P., Monaghan, P. & Andersson, S. Carotenoid pigmentation
807 does not reflect total non-enzymatic antioxidant activity in plasma of adult and nestling
808 great tits, *Parus major*. *Funct Ecol* **21**, 1123–1129 (2007).
- 809 134. Eckert, C. G. & Weatherhead, P. J. Male characteristics, parental quality and the study
810 of mate choice in the red-winged blackbird (*Agelaius phoeniceus*). *Behav Ecol Sociobiol*
811 **20**, 35–42 (1987).
- 812 135. Hõrak, P., Sild, E., Soomets, U., Sepp, T. & Kilk, K. Oxidative stress and information
813 content of black and yellow plumage coloration: an experiment with greenfinches.
814 *Journal of Experimental Biology* **213**, 2225–2233 (2010).
- 815 136. Mougeot, F. *et al.* Oxidative stress and the effect of parasites on a carotenoid-based
816 ornament. *Journal of Experimental Biology* **213**, 400–407 (2010).
- 817 137. Pérez-Rodríguez, L. & Viñuela, J. Carotenoid-based bill and eye ring coloration as
818 honest signals of condition: an experimental test in the red-legged partridge (*Alectoris*
819 *rufa*). *Naturwissenschaften* **95**, 821–830 (2008).
- 820 138. Keyser, A. J. & Hill, G. E. Structurally based plumage coloration is an honest signal of
821 quality in male blue grosbeaks. *Behavioral Ecology* **11**, 202–209 (2000).

- 822 139. Buchanan, K. L. & Catchpole, C. K. Extra-pair paternity in the socially monogamous
823 sedge warbler *Acrocephalus schoenobaenus* as revealed by multilocus DNA
824 fingerprinting. *Ibis* **142**, 12–20 (2000).
- 825 140. Veiga, J. P. Badge size, phenotypic quality, and reproductive success in the house
826 sparrow: a study on honest advertisement. *Evolution* **47**, 1161–1170 (1993).
- 827 141. Kose, M., Mänd, R. & Møller, A. P. Sexual selection for white tail spots in the barn
828 swallow in relation to habitat choice by feather lice. *Anim Behav* **58**, 1201–1205 (1999).
- 829 142. Kose, M. & Møller, A. P. Sexual selection, feather breakage and parasites: the
830 importance of white spots in the tail of the barn swallow (*Hirundo rustica*). *Behav Ecol*
831 *Sociobiol* **45**, 430–436 (1999).
- 832 143. Liker, A. & Barta, Z. Male badge size predicts dominance against females in house
833 sparrows. *Condor* **103**, 151–157 (2001).
- 834 144. Solberg, E. J. & Ringsby, T. H. Does male badge size signal status in small island
835 populations of house sparrows, *Passer domesticus*? *Ethology* **103**, 177–186 (1997).
- 836 145. Møller, A. P. Variation in badge size in male house sparrows *Passer domesticus*:
837 evidence for status signalling. *Anim Behav* **35**, 1637–1644 (1987).
- 838 146. Henderson, L. J., Heidinger, B. J., Evans, N. P. & Arnold, K. E. Ultraviolet crown
839 coloration in female blue tits predicts reproductive success and baseline corticosterone.
840 *Behavioral Ecology* **24**, 1299–1305 (2013).
- 841 147. del Cerro, S. *et al.* Carotenoid-based plumage colouration is associated with blood
842 parasite richness and stress protein levels in blue tits (*Cyanistes caeruleus*). *Oecologia*
843 **162**, 825–835 (2010).
- 844 148. Garamszegi, L. Z. *et al.* The design of complex sexual traits in male barn swallows:
845 associations between signal attributes. *J Evol Biol* **19**, 2052–2066 (2006).
- 846 149. Edler, A. U. & Friedl, T. W. P. Individual quality and carotenoid-based plumage
847 ornaments in male red bishops (*Euplectes orix*): plumage is not all that counts.
848 *Biological Journal of the Linnean Society* **99**, 384–397 (2010).
- 849 150. Maney, D. L., Davis, A. K., Goode, C. T., Reid, A. & Showalter, C. Carotenoid-based
850 plumage coloration predicts leukocyte parameters during the breeding season in
851 northern cardinals (*Cardinalis cardinalis*). *Ethology* **114**, 369–380 (2008).
- 852 151. Dufva, R. & Allander, K. Intraspecific variation in plumage coloration reflects immune
853 response in great tit (*Parus major*) males. *Funct Ecol* **9**, 785 (1995).
- 854 152. Wiehn, J., Korpimäki, E., Bildstein, K. L. & Sorjonen, J. Mate choice and reproductive
855 success in the American kestrel: a role for blood parasites? *Ethology* **103**, 304–317
856 (1997).

- 857 153. Thompson, C. W., Hillgarth, N., Leu, M. & McClure, H. E. High parasite load in house
858 finches (*Carpodacus mexicanus*) is correlated with reduced expression of a sexually
859 selected trait. *Am Nat* **149**, 270–294 (1997).
- 860 154. Seutin, G. Plumage redness in redpoll finches does not reflect hemoparasitic infection.
861 *Oikos* **70**, 280 (1994).
- 862 155. Weatherhead, P. J., Metz, K. J., Bennett, G. F. & Irwin, R. E. Parasite faunas, testosterone
863 and secondary sexual traits in male red-winged blackbirds. *Behav Ecol Sociobiol* **33**,
864 13–23 (1993).
- 865 156. Vergara, P., Fargallo, J. A., Martínez-Padilla, J. & Lemus, J. A. Inter-annual variation
866 and information content of melanin-based coloration in female Eurasian kestrels.
867 *Biological Journal of the Linnean Society* **97**, 781–790 (2009).
- 868 157. Tella, J. L., Forero, M. G., Donazar, J. A. & Hiraldo, F. Is the expressin of male traits in
869 female lesser kestrels related to sexual selection. *Ethology* **103**, 72–81 (1997).
- 870 158. Møller, A. P. Female preference for apparently symmetrical male sexual ornaments in
871 the barn swallow *Hirundo rustica*. *Behav Ecol Sociobiol* **32**, 371–376 (1993).
- 872 159. Brommer, J. E., Ahola, K. & Karstinen, T. The colour of fitness: plumage coloration and
873 lifetime reproductive success in the tawny owl. *Proceedings of the Royal Society B:*
874 *Biological Sciences* **272**, 935–940 (2005).
- 875 160. Horne, T. J. Evolution of females choice in the bank vole. (University of Jyväskylä,
876 1998).
- 877 161. De Lope, F. & Møller, A. P. Female reproductive effort depends on the degree of
878 ornamentation of their mates. *Evolution* **47**, 1152–1160 (1993).
- 879 162. Møller, A. P. Sexual selection in the barn swallow *Hirundo rustica*. III. Female tail
880 ornaments. *Evolution* **47**, 417 (1993).
- 881 163. Welch, A. M., Semlitsch, R. D. & Gerhardt, H. C. Call duration as an indicator of genetic
882 quality in male gray tree frogs. *Science (1979)* **280**, 1928–1930 (1998).
- 883 164. Alatalo, R. V., Kotiaho, J., Mappes, J. & Parri, S. Mate choice for offspring performance:
884 major benefits or minor costs? *Proceedings of the Royal Society B: Biological Sciences*
885 **265**, 2297–2301 (1998).
- 886 165. Hoikkala, A. & Suvanto, L. Male courtship song frequency as an indicator of male
887 mating success in *Drosophila montana*. *J Insect Behav* **12**, 599–609 (1999).
- 888 166. Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire,
889 extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229–
890 232 (1996).
- 891 167. Freeman-Gallant, C. R., Schneider, R. L., Taff, C. C., Dunn, P. O. & Whittingham, L. A.
892 Contrasting patterns of selection on the size and coloration of a female plumage
893 ornament in common yellowthroats. *J Evol Biol* **27**, 982–991 (2014).

- 894 168. Gladbach, A., Gladbach, D. J., Kempenaers, B. & Quillfeldt, P. Female-specific
895 colouration, carotenoids and reproductive investment in a dichromatic species, the
896 upland goose *Chloephaga picta leucoptera*. *Behav Ecol Sociobiol* **64**, 1779–1789
897 (2010).
- 898 169. Quinard, A., Cézilly, F., Motreuil, S., Rossi, J. M. & Biard, C. Reduced sexual
899 dichromatism, mutual ornamentation, and individual quality in the monogamous
900 Zenaida dove *Zenaida aurita*. *J Avian Biol* **48**, 489–501 (2017).
- 901 170. Simons, M. J. P. *et al.* Bill redness is positively associated with reproduction and survival
902 in male and female zebra finches. *PLoS One* **7**, (2012).
- 903 171. Komdeur, J., Oorebeek, M., Van Overveld, T. & Cuthill, I. C. Mutual ornamentation,
904 age, and reproductive performance in the European starling. *Behavioral Ecology* **16**,
905 805–817 (2005).
- 906 172. Daunt, F., Monaghan, P., Wanless, S. & Harris, M. P. Sexual ornament size and breeding
907 performance in female and male European shags *Phalacrocorax aristotelis*. *Ibis* **145**,
908 54–60 (2003).
- 909 173. Morrison, A., Flood, N. J. & Reudink, M. W. Reproductive correlates of plumage
910 coloration of female mountain bluebirds. *J Field Ornithol* **85**, 168–179 (2014).
- 911 174. Mänd, R., Tilgar, V. & Moøller, A. P. Negative relationship between plumage colour and
912 breeding output in female great tits, *Parus major*. *Evol Ecol Res* **7**, 1013–1023 (2005).
- 913 175. Muma, K. E. & Weatherhead, P. J. Male traits expressed in females: direct or indirect
914 sexual selection? *Behav Ecol Sociobiol* **25**, 23–31 (1989).
- 915 176. Beck, M. L., Hopkins, W. A. & Hawley, D. M. Relationships among plumage coloration,
916 blood selenium concentrations and immune responses of adult and nestling tree
917 swallows. *Journal of Experimental Biology* **218**, 3415–3424 (2015).
- 918 177. Potti, J. & Merino, S. Decreased levels of blood trypanosome infection correlate with
919 female expression of a male secondary sexual trait: implications for sexual selection.
920 *Proceedings of the Royal Society B: Biological Sciences* **263**, 1199–1204 (1996).
- 921 178. Bentz, A. B. & Siefferman, L. Age-dependent relationships between coloration and
922 reproduction in a species exhibiting delayed plumage maturation in females. *J Avian*
923 *Biol* **44**, 080–088 (2013).
- 924 179. Osmond, M. M. *et al.* Relationships between carotenoid-based female plumage and age,
925 reproduction, and mate colour in the American redstart (*Setophaga ruticilla*). *Can J Zool*
926 **91**, 589–595 (2013).
- 927 180. Bulluck, L. P. *et al.* Feather carotenoid content is correlated with reproductive success
928 and provisioning rate in female prothonotary warblers. *Auk* **134**, 229–239 (2017).
- 929 181. Doutrelant, C. *et al.* Female coloration indicates female reproductive capacity in blue
930 tits. *J Evol Biol* **21**, 226–233 (2008).

- 931 182. Cuervo, J. J., Møller, A. P. & De Lope, F. Experimental manipulation of tail length in
932 female barn swallows (*Hirundo rustica*) affects their future reproductive success.
933 *Behavioral Ecology* **14**, 451–456 (2003).
- 934 183. Silva, N., Avilés, J. M., Danchin, E. & Parejo, D. Informative content of multiple
935 plumage-coloured traits in female and male European rollers. *Behav Ecol Sociobiol* **62**,
936 1969–1979 (2008).
- 937 184. Viblanc, V. A. *et al.* Mutually honest? Physiological ‘qualities’ signalled by colour
938 ornaments in monomorphic king penguins. *Biological Journal of the Linnean Society*
939 **118**, 200–214 (2016).
- 940 185. van Dijk, R. E., Robles, R., Groothuis, T. G. G., de Vries, B. & Eising, C. M.
941 Reproductive effort of both male and female bar-throated apalis *Apalis thoracica* is
942 predicted by ornamentation of self and mate. *Ibis* **157**, 731–742 (2015).
- 943 186. Lindström, K. M., Hasselquist, D. & Wikelski, M. House sparrows (*Passer domesticus*)
944 adjust their social status position to their physiological costs. *Horm Behav* **48**, 311–320
945 (2005).
- 946 187. Hein, W. K., Westneat, D. F. & Poston, J. P. Sex of opponent influences response to a
947 potential status signal in house sparrows. *Anim Behav* **65**, 1211–1221 (2003).
- 948 188. Riters, L. V., Teague, D. P. & Schroeder, M. B. Social status interacts with badge size
949 and neuroendocrine physiology to influence sexual behavior in male house sparrows
950 (*Passer domesticus*). *Brain Behav Evol* **63**, 141–150 (2004).
- 951 189. Gonzalez, G., Sorci, G., Smith, L. C. & De Lope, F. Social control and physiological
952 cost of cheating in status signalling male house sparrows (*Passer domesticus*). *Ethology*
953 **108**, 289–302 (2002).
- 954 190. Bókony, V., Lendvai, Á. Z. & Liker, A. Multiple cues in status signalling: the role of
955 wingbars in aggressive interactions of male house sparrows. *Ethology* **112**, 947–954
956 (2006).
- 957 191. Delhey, K. & Kempenaers, B. Age differences in blue tit *Parus caeruleus* plumage
958 colour: within-individual changes or colour-biased survival? *J Avian Biol* **37**, 339–348
959 (2006).
- 960 192. Surmacki, A., Stępniewski, J. & Stępniewska, M. Juvenile sexual dimorphism,
961 dichromatism and condition-dependent signaling in a bird species with early pair bonds.
962 *J Ornithol* **156**, 65–73 (2015).
- 963 193. Hidalgo-Garcia, S. The carotenoid-based plumage coloration of adult blue tits *Cyanistes*
964 *caeruleus* correlates with the health status of their brood. *Ibis* **148**, 727–734 (2006).
- 965 194. Mougeot, F., Redpath, S. M. & Leckie, F. Ultra-violet reflectance of male and female
966 red grouse, *Lagopus lagopus scoticus*: sexual ornaments reflect nematode parasite
967 intensity. *J Avian Biol* **36**, 203–209 (2005).

- 968 195. Saino, N. *et al.* Viability is associated with melanin-based coloration in the barn swallow
969 (*Hirundo rustica*). *PLoS One* **8**, e60426 (2013).
- 970 196. Vortman, Y., Lotem, A., Dor, R., Lovette, I. J. & Safran, R. J. The sexual signals of the
971 East-Mediterranean barn swallow: a different swallow tale. *Behavioral Ecology* **22**,
972 1344–1352 (2011).
- 973 197. Faivre, B., Grégoire, A., Préault, M., Cézilly, F. & Sorci, G. Immune activation rapidly
974 mirrored in a secondary sexual trait. *Science (1979)* **300**, 103–103 (2003).
- 975 198. Grunst, A. S., Rotenberry, J. T. & Grunst, M. L. Age-dependent relationships between
976 multiple sexual pigments and condition in males and females. *Behavioral Ecology* **25**,
977 276–287 (2014).
- 978 199. Kraaijeveld, K., Gregurke, J., Hall, C., Komdeur, J. & Mulder, R. A. Mutual
979 ornamentation, sexual selection, and social dominance in the black swan. *Behavioral*
980 *Ecology* **15**, 380–389 (2004).
- 981 200. Tschirren, B., Fitze, P. S. & Richner, H. Proximate mechanisms of variation in the
982 carotenoid-based plumage coloration of nestling great tits (*Parus major* L.). *J Evol Biol*
983 **16**, 91–100 (2003).
- 984 201. Figuerola, J., Domènech, J. & Senar, J. C. Plumage colour is related to ectosymbiont
985 load during moult in the serin, *Serinus serinus*: an experimental study. *Anim Behav* **65**,
986 551–557 (2003).
- 987 202. De Kogel, C. H. & Prijs, H. J. Effects of brood size manipulations on sexual
988 attractiveness of offspring in the zebra finch. *Anim Behav* **51**, 699–708 (1996).
- 989 203. McGraw, K. J. & Hill, G. E. Differential effects of endoparasitism on the expression of
990 carotenoid- and melanin-based ornamental coloration. *Proceedings of the Royal Society*
991 *B: Biological Sciences* **267**, 1525–1531 (2000).
- 992 204. Griffith, S. C., Owens, I. P. F. & Burke, T. Female choice and annual reproductive
993 success favour less-ornamented male house sparrows. *Proceedings of the Royal Society*
994 *B: Biological Sciences* **266**, 765–770 (1999).
- 995 205. McGraw, K. J., Mackillop, E. A., Dale, J. & Hauber, M. E. Different colors reveal
996 different information: how nutritional stress affects the expression of melanin- and
997 structurally based ornamental plumage. *Journal of Experimental Biology* **205**, 3747–
998 3755 (2002).
- 999 206. Fitze, P. S. & Richner, H. Differential effects of a parasite on ornamental structures based
1000 on melanins and carotenoids. *Behavioral Ecology* **13**, 401–407 (2002).
- 1001 207. Johnsen, A., Delhey, K., Schlicht, E., Peters, A. & Kempenaers, B. Male sexual
1002 attractiveness and parental effort in blue tits: a test of the differential allocation
1003 hypothesis. *Anim Behav* **70**, 877–888 (2005).

- 1004 208. Maguire, S. E. & Safran, R. J. Morphological and genetic predictors of parental care in
1005 the North American barn swallow *Hirundo rustica erythrogaster*. *J Avian Biol* **41**, 74–
1006 82 (2010).
- 1007 209. Sundberg, J. & Larsson, C. Male coloration as an indicator of parental quality in the
1008 yellowhammer, *Emberiza citrinella*. *Animal Behaviour* vol. 48 885–892 Preprint at
1009 <https://doi.org/10.1006/anbe.1994.1313> (1994).
- 1010 210. Lozano, G. A. & Lemon, R. E. Male plumage, paternal care and reproductive success in
1011 yellow warblers, *Dendroica petechia*. *Anim Behav* **51**, 265–272 (1996).
- 1012 211. Studd, M. V. & Robertson, R. J. Sexual selection and variation in reproductive strategy
1013 in male yellow warblers (*Dendroica petechia*). *Behav Ecol Sociobiol* **17**, 101–109
1014 (1985).
- 1015 212. Møller, A. P. Sexual selection in the barn swallow (*Hirundo rustica*). IV. Patterns of
1016 fluctuating asymmetry and selection against asymmetry. *Evolution* **48**, 658 (1994).
- 1017 213. Jawor, J. M., Gray, N., Beall, S. M. & Breitwisch, R. Multiple ornaments correlate with
1018 aspects of condition and behaviour in female northern cardinals, *Cardinalis cardinalis*.
1019 *Anim Behav* **67**, 875–882 (2004).
- 1020 214. Linville, S. U., Breitwisch, R. & Schilling, A. J. Plumage brightness as an indicator of
1021 parental care in northern cardinals. *Anim Behav* **55**, 119–127 (1998).
- 1022 215. Hōrak, P. *et al.* How coccidian parasites affect health and appearance of greenfinches.
1023 *Journal of Animal Ecology* **73**, 935–947 (2004).
- 1024 216. Møller, A. P. & Tegelström, H. Extra-pair paternity and tail ornamentation in the barn
1025 swallow *Hirundo rustica*. *Behav Ecol Sociobiol* **41**, 353–360 (1997).
- 1026 217. Møller, A. P. Male ornament size as a reliable cue to enhanced offspring viability in the
1027 barn swallow. *Proc Natl Acad Sci U S A* **91**, 6929–6932 (1994).
- 1028 218. Møller, A. P. Phenotype-dependent arrival time and its consequences in a migratory bird.
1029 *Behav Ecol Sociobiol* **35**, 115–122 (1994).
- 1030 219. Møller, A. P. Symmetrical male sexual ornaments, paternal care, and offspring quality.
1031 *Behavioral Ecology* **5**, 188–194 (1994).
- 1032 220. Burley, N. & Coopersmith, C. B. Bill color preferences of zebra finches. *Ethology* **76**,
1033 133–151 (1987).
- 1034 221. Møller, A. P. Female swallow preference for symmetrical male sexual ornaments. *Nature*
1035 **357**, 238–240 (1992).
- 1036 222. Møller, A. P. Sexual selection in the monogamous barn swallow (*Hirundo rustica*). II.
1037 Mechanisms of sexual selection. *J Evol Biol* **5**, 603–624 (1992).
- 1038 223. Møller, A. P. Male tail length and female mate choice in the monogamous swallow
1039 *Hirundo rustica*. *Anim Behav* **39**, 458–465 (1990).

- 1040 224. Rintamäki, P. T., Alatalo, R. V., Höglund, J. & Lundberg, A. Fluctuating asymmetry and
1041 copulation success in lekking black grouse. *Anim Behav* **54**, 265–269 (1997).
- 1042 225. Collins, S. A., Hubbard, C. & Houtman, A. M. Female mate choice in the zebra finch -
1043 the effect of male beak colour and male song. *Behav Ecol Sociobiol* **35**, 21–25 (1994).
- 1044 226. Møller, A. P. Viability costs of male tail ornaments in a swallow. *Nature* **339**, 132–135
1045 (1989).
- 1046 227. Møller, A. P. Natural and sexual selection on a plumage signal of status and on
1047 morphology in house sparrows, *Passer domesticus*. *J Evol Biol* **2**, 125–140 (1989).
- 1048 228. Norris, K. J. Female choice and the evolution of the conspicuous plumage coloration of
1049 monogamous male great tits. *Behav Ecol Sociobiol* **26**, 129–138 (1990).
- 1050 229. Evans, Matthew R. & Hatchwell, B. J. An experimental study of male adornment in the
1051 scarlet-tufted malachite sunbird: I. The role of pectoral tufts in territorial defence. *Behav*
1052 *Ecol Sociobiol* **29**, 421–427 (1992).
- 1053 230. Saino, N., Bolzern, A. M. & Møller, A. P. Immunocompetence, ornamentation, and
1054 viability of male barn swallows (*Hirundo rustica*). *Proceedings of the National Academy*
1055 *of Sciences* **94**, 549–552 (1997).
- 1056 231. Baeta, R., Faivre, B., Motreuil, S., Gaillard, M. & Moreau, J. Carotenoid trade-off
1057 between parasitic resistance and sexual display: an experimental study in the blackbird
1058 (*Turdus merula*). *Proceedings of the Royal Society B: Biological Sciences* **275**, 427–434
1059 (2008).
- 1060 232. Stirnemann, I., Johnston, G., Rich, B., Robertson, J. & Kleindorfer, S.
1061 Phytohaemagglutinin (PHA) response and bill-hue wavelength increase with carotenoid
1062 supplementation in diamond firetails (*Stagonopleura guttata*). *Emu* **109**, 344–351
1063 (2009).
- 1064 233. Hill, G. E., Hood, W. R. & Huggins, K. A multifactorial test of the effects of carotenoid
1065 access, food intake and parasite load on the production of ornamental feathers and bill
1066 coloration in American goldfinches. *Journal of Experimental Biology* **212**, 1225–1233
1067 (2009).
- 1068 234. Navara, K. J. & Hill, G. E. Dietary carotenoid pigments and immune function in a
1069 songbird with extensive carotenoid-based plumage coloration. *Behavioral Ecology* **14**,
1070 909–916 (2003).
- 1071 235. Peters, A., Delhey, K., Andersson, S., Van Noordwijk, H. & Förchler, M. I. Condition-
1072 dependence of multiple carotenoid-based plumage traits: an experimental study. *Funct*
1073 *Ecol* **22**, 831–839 (2008).
- 1074 236. Karu, U., Saks, L. & Hõrak, P. Carotenoid coloration in greenfinches is individually
1075 consistent irrespective of foraging ability. *Physiological and Biochemical Zoology* **80**,
1076 663–670 (2007).

- 1077 237. Blount, J. D., Metcalfe, N. B., Birkhead, T. R. & Surai, P. F. Carotenoid modulation of
1078 immune function and sexual attractiveness in zebra finches. *Science (1979)* **300**, 125–
1079 127 (2003).
- 1080 238. Ryan, M. J., Perrill, S. A. & Wilczynski, W. Auditory tuning and call frequency predict
1081 population-based mating preferences in the cricket frog, *Acris crepitans*. *Am Nat* **139**,
1082 1370–1383 (1992).
- 1083 239. McGraw, K. J. & Ardia, D. R. Carotenoids, immunocompetence, and the information
1084 content of sexual colors: an experimental test. *Am Nat* **162**, 704–712 (2003).
- 1085 240. Jensen, H. *et al.* Lifetime reproductive success in relation to morphology in the house
1086 sparrow *Passer domesticus*. *Journal of Animal Ecology* **73**, 599–611 (2004).
- 1087 241. Václav, R. & Hoi, H. Different reproductive tactics in house sparrows signalled by badge
1088 size: is there a benefit to being average? *Ethology* **108**, 569–582 (2002).
- 1089 242. Birkhead, A. T. R. & Fletcher, F. Male phenotype and ejaculate quality in the zebra finch
1090 *Taeniopygia guttata*. *Proceedings of the Royal Society B: Biological Sciences* **262**, 329–
1091 334 (1995).
- 1092 243. Potti, J. & Montalvo, S. Male arrival and female mate choice in pied flycatchers
1093 *Ficedula hypoleuca* in central Spain. *Ornis Scandinavica* **22**, 45–54 (1991).
- 1094 244. Evans, J. P. Quantitative genetic evidence that males trade attractiveness for ejaculate
1095 quality in guppies. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3195–
1096 3201 (2010).
- 1097 245. Engen, F. & Folstad, I. Cod courtship song: a song at the expense of dance? *Can J Zool*
1098 **77**, 542–550 (1999).
- 1099 246. Peters, A., Denk, A. G., Delhey, K. & Kempenaers, B. Carotenoid-based bill colour as
1100 an indicator of immunocompetence and sperm performance in male mallards. *J Evol*
1101 *Biol* **17**, 1111–1120 (2004).
- 1102 247. Hasselquist, D. Polygyny in great reed warblers: a long-term study of factors
1103 contributing to male fitness. *Ecology* **79**, 2376–2390 (1998).
- 1104 248. Roulin, A., Dijkstra, C., Riols, C. & Ducrest, A. L. Female- and male-specific signals of
1105 quality in the barn owl. *J Evol Biol* **14**, 255–266 (2001).
- 1106 249. Roulin, A., Riols, C., Dijkstra, C. & Ducrest, A. L. Female plumage spottiness signals
1107 parasite resistance in the barn owl (*Tyto alba*). *Behavioral Ecology* **12**, 103–110 (2001).
- 1108 250. Siefferman, L. & Hill, G. E. Structural and melanin coloration indicate parental effort
1109 and reproductive success in male eastern bluebirds. *Behavioral Ecology* **14**, 855–861
1110 (2003).
- 1111 251. Siefferman, L. & Hill, G. E. Evidence for sexual selection on structural plumage
1112 coloration in female eastern bluebirds (*Sialia sialis*). *Evolution* **59**, 1819–1828 (2005).

- 1113 252. Balenger, S. L., Scott Johnson, L., Brubaker, J. L. & Ostlind, E. Parental effort in relation
1114 to structural plumage coloration in the mountain bluebird (*Sialia currucoides*). *Ethology*
1115 **113**, 838–846 (2007).
- 1116 253. McGregor, P. K., Krebs, J. R. & Perrins, C. M. Song repertoires and lifetime
1117 reproductive success in the great tit (*Parus major*). *Am Nat* **118**, 149–159 (1981).
- 1118 254. Voltura, K. M., Schwagmeyer, P. L. & Mock, D. W. Parental feeding rates in the house
1119 sparrow, *Passer domesticus*: are larger-badged males better fathers? *Ethology* **108**,
1120 1011–1022 (2002).
- 1121 255. Nakagawa, S., Ockendon, N., Gillespie, D. O. S., Hatchwell, B. J. & Burke, T. Does the
1122 badge of status influence parental care and investment in house sparrows? An
1123 experimental test. *Oecologia* **153**, 749–760 (2007).
- 1124 256. Møller, A. P. Parasite load reduces song output in a passerine bird. *Anim Behav* **41**, 723–
1125 730 (1991).
- 1126 257. Alonso-Alvarez, C. *et al.* An experimental test of the dose-dependent effect of
1127 carotenoids and immune activation on sexual signals and antioxidant activity. *Am Nat*
1128 **164**, 651–659 (2004).
- 1129 258. Lambrechts, M. & Dhondt, A. A. Male quality, reproduction, and survival in the great
1130 tit (*Parus major*). *Behav Ecol Sociobiol* **19**, 57–63 (1986).
- 1131 259. Møller, A. P. Viability is positively related to degree of ornamentation in male swallows.
1132 *Proceedings of the Royal Society B: Biological Sciences* **243**, 145–148 (1991).
- 1133 260. Møller, A. P. & Nielsen, J. T. Differential predation cost of a secondary sexual character:
1134 sparrowhawk predation on barn swallows. *Anim Behav* **54**, 1545–1551 (1997).
- 1135 261. Brown, C. R. & Brown, M. B. Natural selection on tail and bill morphology in barn
1136 swallows *Hirundo rustica* during severe weather. *Ibis* **141**, 652–659 (1999).
- 1137 262. Møller, A. P. & de Lope, F. Differential costs of a secondary sexual character: an
1138 experimental test of the handicap principle. *Evolution* **48**, 1676 (1994).
- 1139 263. Buchanan, K. L. & Catchpole, C. K. Song as an indicator of male parental effort in the
1140 sedge warbler. *Proceedings of the Royal Society B: Biological Sciences* **267**, 321–326
1141 (2000).
- 1142 264. Birkhead, T. R., Fletcher, F. & Pellatt, E. J. Sexual selection in the zebra finch
1143 *Taeniopygia guttata*: condition, sex traits and immune capacity. *Behav Ecol Sociobiol*
1144 **44**, 179–191 (1998).
- 1145 265. Adamo, S. A., Kovalko, I., Easy, R. H. & Stoltz, D. A viral aphrodisiac in the cricket
1146 *Gryllus texensis*. *Journal of Experimental Biology* **217**, 1970–1976 (2014).
- 1147 266. Smith, M. J. & Roberts, J. D. Call structure may affect male mating success in the
1148 quacking frog *Crinia georgiana* (Anura: Myobatrachidae). *Behav Ecol Sociobiol* **53**,
1149 221–226 (2003).

- 1150 267. Kingma, S. A. *et al.* Sexual selection and the function of a melanin-based plumage
1151 ornament in polygamous penduline tits *Remiz pendulinus*. *Behav Ecol Sociobiol* **62**,
1152 1277–1288 (2008).
- 1153 268. David Ligon, J. & Zwartjes, P. W. Female red junglefowl choose to mate with multiple
1154 males. *Anim Behav* **49**, 127–135 (1995).
- 1155 269. Ligon, J. D. & Zwartjes, P. W. Ornate plumage of male red junglefowl does not influence
1156 mate choice by females. *Anim Behav* **49**, 117–125 (1995).
- 1157 270. Kempenaers, B., Verheyen, G. R. & Dhondt, A. A. Extrapair paternity in the blue tit
1158 (*Parus caeruleus*): female choice, male characteristics, and offspring quality. *Behavioral*
1159 *Ecology* **8**, 481–492 (1997).
- 1160 271. Leisler, B. Variation in extra-pair paternity in the polygynous great reed warbler
1161 (*Acrocephalus arundinaceus*). *Journal für Ornithologie* **141**, 77 (2000).
- 1162 272. Chappell, M. A., Zuk, M., Johnsen, T. S. & Kwan, T. H. Mate choice and aerobic
1163 capacity in red junglefowl. *Behaviour* **134**, 511–529 (1997).
- 1164 273. Kennedy, C. E. J., Endler, J. A., Poynton, S. L. & McMinn, H. Parasite load predicts
1165 mate choice in guppies. *Behav Ecol Sociobiol* **21**, 291–295 (1987).
- 1166 274. Hall, Z. J., Bertin, M., Bailey, I. E., Meddle, S. L. & Healy, S. D. Neural correlates of
1167 nesting behavior in zebra finches (*Taeniopygia guttata*). *Behavioural Brain Research*
1168 **264**, 26–33 (2014).
- 1169 275. Bernard, D. J., Eens, M. & Ball, G. F. Age- and behavior-related variation in volumes of
1170 song control nuclei in male European starlings. *J Neurobiol* **30**, 329–339 (1996).
- 1171 276. Taylor, M. I., Turner, G. F., Robinson, R. L. & Stauffer, J. R. Sexual selection, parasites
1172 and bower height skew in a bower-building cichlid fish. *Anim Behav* **56**, 379–384
1173 (1998).
- 1174 277. Pruett-jones, S. G., Pruett-jones, M. A. & Jones, H. I. Parasites and sexual selection in
1175 birds of paradise. *Integr Comp Biol* **30**, 287–298 (1990).
- 1176 278. Price, D. K. & Burley, N. T. Constraints on the evolution of attractive traits: selection in
1177 male and female zebra finches. *Am Nat* **144**, 908–934 (1994).
- 1178 279. Sundberg, J. Parasites, plumage coloration and reproductive success in the
1179 yellowhammer, *Emberiza citrinella*. *Oikos* **74**, 331 (1995).
- 1180 280. Bosholn, M., Fecchio, A., Silveira, P., Braga, É. M. & Anciães, M. Effects of avian
1181 malaria on male behaviour and female visitation in lekking blue-crowned manakins. *J*
1182 *Avian Biol* **47**, 457–465 (2016).
- 1183 281. Arak, A. Female mate selection in the natterjack toad: active choice or passive attraction?
1184 *Behav Ecol Sociobiol* **22**, 317–327 (1988).

- 1185 282. Palokangas, P. *et al.* Female kestrels gain reproductive success by choosing brightly
1186 ornamented males. *Anim Behav* **47**, 443–448 (1994).
- 1187 283. Wiehn, J. Plumage characteristics as an indicator of male parental quality in the
1188 American kestrel. *J Avian Biol* **28**, 47 (1997).
- 1189 284. Weatherhead, P. J. & Boag, P. T. Pair and extra-pair mating success relative to male
1190 quality in red-winged blackbirds. *Behav Ecol Sociobiol* **37**, 81–91 (1995).
- 1191 285. Hill, G. E. Female house finches prefer colourful males: sexual selection for a condition-
1192 dependent trait. *Anim Behav* **40**, 563–572 (1990).
- 1193 286. Andersson, S. Female preference for long tails in lekking Jackson's widowbirds:
1194 experimental evidence. *Anim Behav* **43**, 379–388 (1992).
- 1195 287. Møller, A. P. Female choice selects for male sexual tail ornaments in the monogamous
1196 swallow. *Nature* **332**, 640–642 (1988).
- 1197 288. Grant, B. R. & Grant, P. R. Mate choice in Darwin's Finches. *Biological Journal of the*
1198 *Linnean Society* **32**, 247–270 (1987).
- 1199 289. Petrie, M. & Halliday, T. Experimental and natural changes in the peacock's (*Pavo*
1200 *cristatus*) train can affect mating success. *Behav Ecol Sociobiol* **35**, 213–217 (1994).
- 1201 290. Andersson, S. Sexual selection and cues for female choice in leks of Jackson's
1202 widowbird *Euplectes jacksoni*. *Behav Ecol Sociobiol* **25**, 403–410 (1989).
- 1203 291. Smith, H. G. & Montgomerie, R. Sexual selection and the tail ornaments of North
1204 American barn swallows. *Behav Ecol Sociobiol* **28**, 195–201 (1991).
- 1205 292. Grant, B. R. The significance of subadult plumage in Darwin's finches, *Geospiza fortis*.
1206 *Behavioral Ecology* **1**, 161–170 (1990).
- 1207 293. Røskaft, E. & Järvi, T. Male plumage colour and mate choice of female pied flycatchers
1208 *Ficedula hypoleuca*. *Ibis* **125**, 396–400 (1983).
- 1209 294. Zuk, M., Ligon, J. D. & Thornhill, R. Effects of experimental manipulation of male
1210 secondary sex characters on female mate preference in red jungle fowl. *Anim Behav* **44**,
1211 999–1006 (1992).
- 1212 295. Gustafsson, L., Qvarnström, A. & Sheldon, B. C. Trade-offs between life-history traits
1213 and a secondary sexual character in male collared flycatchers. *Nature* **375**, 311–313
1214 (1995).
- 1215 296. Järvi, T., Røskaft, E., Bakken, M. & Zumsteg, B. Evolution of variation in male
1216 secondary sexual characteristics. *Behav Ecol Sociobiol* **20**, 161–169 (1987).
- 1217 297. McLennan, D. A. & Shires, V. L. Correlation between the Level of infection with
1218 *Bunodera inconstans* and *Neoechinorhynchus rutili* and behavioral intensity in female
1219 brook sticklebacks. *J Parasitol* **81**, 675 (1995).

- 1220 298. Tárano, Z. Variation in male advertisement calls in the Neotropical frog *Physalaemus*
1221 *enesefae*. *Copeia* **2001**, 1064–1072 (2001).
- 1222 299. Isaksson, C., Ornborg, J., Prager, M. & Andersson, S. Sex and age differences in
1223 reflectance and biochemistry of carotenoid-based colour variation in the great tit *Parus*
1224 *major*. *Biological Journal of the Linnean Society* **95**, 758–765 (2008).
- 1225 300. Zuk, M. *et al.* The role of male ornaments and courtship behavior in female mate choice
1226 of red jungle fowl. *Am Nat* **136**, 459–473 (1990).
- 1227 301. Houde, A. E. & Torio, A. J. Effect of parasitic infection on male color pattern and female
1228 choice in guppies. *Behavioral Ecology* **3**, 346–351 (1992).
- 1229 302. Greenspan, S. E., Roznik, E. A., Schwarzkopf, L., Alford, R. A. & Pike, D. A. Robust
1230 calling performance in frogs infected by a deadly fungal pathogen. *Ecol Evol* **6**, 5964–
1231 5972 (2016).
- 1232 303. Gibson, R. M. Relationships between blood parasites, mating success and phenotypic
1233 cues in male sage grouse *Centrocercus urophasianus*. *Am Zool* **30**, 271–278 (1990).
- 1234 304. Hill, G. E. Plumage coloration is a sexually selected indicator of male quality. *Nature*
1235 **350**, 337–339 (1991).
- 1236 305. Fiske, P., Kålås, J. A. & Saether, S. A. Correlates of male mating success in the lekking
1237 great snipe (*Gallinago media*): results from a four-year study. *Behavioral Ecology* **5**,
1238 210–218 (1994).
- 1239 306. Clayton, D. H. Mate choice in experimentally parasitized rock doves: lousy males lose.
1240 *Am Zool* **30**, 251–262 (1990).
- 1241 307. Petrie, M., Tim, H. & Carolyn, S. Peahens prefer peacocks with elaborate trains. *Anim*
1242 *Behav* **41**, 323–331 (1991).
- 1243 308. Cherry, M. I. Sexual selection in the raucous toad, *Bufo rangeri*. *Anim Behav* **45**, 359–
1244 373 (1993).
- 1245 309. Senar, J. C., Negro, J. J., Quesada, J., Ruiz, I. & Garrido, J. Two pieces of information
1246 in a single trait? The yellow breast of the great tit (*Parus major*) reflects both pigment
1247 acquisition and body condition. *Behaviour* **145**, 1195–1210 (2008).
- 1248 310. Hőrak, P., Ots, I., Vellau, H., Spottiswoode, C. & Pape Møller, A. Carotenoid-based
1249 plumage coloration reflects hemoparasite infection and local survival in breeding great
1250 tits. *Oecologia* **126**, 166–173 (2001).
- 1251 311. Molnár, O., Bajer, K., Mészáros, B., Török, J. & Herczeg, G. Negative correlation
1252 between nuptial throat colour and blood parasite load in male European green lizards
1253 supports the Hamilton–Zuk hypothesis. *Naturwissenschaften* **100**, 551–558 (2013).
- 1254 312. Safran, R. J. Plumage coloration, not length or symmetry of tail-streamers, is a sexually
1255 selected trait in North American barn swallows. *Behavioral Ecology* **15**, 455–461
1256 (2004).

- 1257 313. Hasegawa, M., Arai, E., Watanabe, M. & Nakamura, M. Colourful males hold high
1258 quality territories but exhibit reduced paternal care in barn swallows. *Behaviour* **151**,
1259 591–612 (2014).
- 1260 314. Harper, D. G. C. Feather mites, pectoral muscle condition, wing length and plumage
1261 coloration of passerines. *Anim Behav* **58**, 553–562 (1999).
- 1262 315. Pick, J. L., Nakagawa, S. & Noble, D. W. A. A. Reproducible, flexible and high-
1263 throughput data extraction from primary literature: the metaDigitise R package. *Methods*
1264 *Ecol Evol* **10**, 426–431 (2019).
- 1265 316. Nakagawa, S. *et al.* Methods for testing publication bias in ecological and evolutionary
1266 meta-analyses. *Methods Ecol Evol* **2021**, 1–18 (2021).
- 1267 317. R Core Team. R: a language and environment for statistical computing. Preprint at
1268 <https://www.r-project.org/> (2024).
- 1269 318. Viechtbauer, W. Conducting meta-analyses in R with the metafor. *J Stat Softw* **36**, 1–48
1270 (2010).
- 1271 319. Nakagawa, S. & Santos, E. S. A. Methodological issues and advances in biological meta-
1272 analysis. *Evol Ecol* **26**, 1253–1274 (2012).
- 1273 320. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and
1274 evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
- 1275 321. Michonneau, F., Brown, J. W. & Winter, D. J. rotl: an R package to interact with the
1276 Open Tree of Life data. *Methods Ecol Evol* **7**, 1476–1481 (2016).
- 1277 322. OpenTreeOfLife *et al.* Open Tree of Life Synthetic Tree. Preprint at
1278 <https://doi.org/10.5281/zenodo.3937741> (2019).
- 1279 323. Norry, F. M., Vilaridi, J. C. & Hasson, E. Genetic and phenotypic correlations among
1280 size-related traits, and heritability variation between body parts in *Drosophila buzzatii*.
1281 *Genetica* **101**, 131–139 (1997).
- 1282 324. Seguin, A. & Forstmeier, W. No band color effects on male courtship rate or body mass
1283 in the zebra finch: four experiments and a meta-analysis. *PLoS One* **7**, e37785 (2012).
- 1284 325. Boake, C. R. B. Genetic consequences of mate choice: a quantitative genetic method for
1285 testing sexual selection theory. *Science (1979)* **227**, 1061–1063 (1985).
- 1286 326. Whittier, T. S. & Kaneshiro, K. Y. Intersexual selection in the Mediterranean fruit fly:
1287 does female choice enhance fitness? *Evolution* **49**, 990–996 (1995).
- 1288 327. Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-Analysis*. (Elsevier, Amsterdam,
1289 1985). doi:10.1016/C2009-0-03396-0.
- 1290 328. Jacobs, P. & Viechtbauer, W. Estimation of the biserial correlation and its sampling
1291 variance for use in meta-analysis. *Res Synth Methods* **8**, 161–180 (2017).

- 1292 329. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. (Lawrence Erlbaum
1293 Associates, Hillsdale, NJ, 1988).
- 1294 330. Gurevitch, J., Morrison, J. A. & Hedges, L. V. The interaction between competition and
1295 predation: a meta-analysis of field experiments. *American Naturalist* **155**, 435–453
1296 (2000).
- 1297 331. Dunlap, W. P., Cortina, J. M., Vaslow, J. B. & Burke, M. J. Meta-analysis of experiments
1298 with matched groups or repeated measures designs. *Psychol Methods* **1**, 170–177 (1996).
- 1299 332. Wan, X., Wang, W., Liu, J. & Tong, T. Estimating the sample mean and standard
1300 deviation from the sample size, median, range and/or interquartile range. *BMC Med Res*
1301 *Methodol* **14**, 1–13 (2014).
- 1302 333. Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance:
1303 a practical guide for biologists. *Biological Reviews* **82**, 591–605 (2007).

1304

1305 **Supplementary tables**

1306 Table S1. Proxies (i.e. variables related to condition-dependency, fitness, and other traits)
1307 considered valid in our study related to bearers of a sexual signal. *N* represents the number of
1308 meta-analytical studies in our dataset including each of these proxies. Underlined proxies are
1309 predicted to have a negative relationship with sexual signal expression, while others are
1310 predicted to have a positive relationship with sexual signal expression.

Proxy	Sub proxy	Examples
Attractiveness	Attractiveness or mating success (<i>N</i> = 15)	Copulation success, harem size, success in mate choice trials, <u>divorce</u> , pairing success or status
	Latency (<i>N</i> = 9)	<u>Latency to arrive at breeding site</u> , <u>latency to nest</u> , <u>latency to pair</u> , <u>latency to mate</u> , <u>latency to breed</u> , <u>latency to lay eggs</u> , <u>latency for eggs to hatch</u> , <u>latency for offspring to fledge</u>
Benefits to sexual signal bearers or to mates	Offspring quality or viability (<i>N</i> = 9)	Antioxidants or hormones in yolk, offspring growth rate, offspring attractiveness, offspring reproductive success, proportion of eggs hatched, fledging success, offspring size
	Parental care (<i>N</i> = 7)	Feeding rate, incubation frequency
	Paternity (<i>N</i> = 10)	Within and extra-pair paternity, <u>cuckoldry occurrence</u>
	Reproductive success	Clutch size, breeding success, number of fledglings, total offspring sired

	(<i>N</i> = 12)	
	Territory (<i>N</i> = 2)	Territory quality or size
Individual condition	Body condition (<i>N</i> = 11)	Carotenoid, protein, or lipid amount in plasma or in feathers, unspecified body condition, feather quality, subcutaneous fat score, residual mass, pectoral score
	External condition (<i>N</i> = 7)	<u>Increase in brood size or reproductive effort</u> , decrease in brood size or reproductive effort, habitat quality, mother's condition, <u>date when reared</u> , diet supplementation, <u>dietary deprivation</u> , <u>nutritional stress</u>
	Immune or antioxidant capacity (<i>N</i> = 10)	Antibody response, <u>glucocorticoids</u> , haematocrit, <u>heterophil-to-lymphocyte ratio</u> , <u>oxidative damage</u> , white blood cells
	Parasite resistance (<i>N</i> = 10)	<u>Abundance of parasites</u> , <u>infection with a pathogen</u> , <u>pathogen richness</u> , parasite removal
	Survival (<i>N</i> = 9)	Days alive, seen or re-captured after a given period
		Age (<i>N</i> = 8)
Other individual traits	Aggression or social dominance (<i>N</i> = 7)	Performed aggression, <u>received aggression</u> , dominance, fights initiated, social rank, nest defence, <u>distance from intruder</u> , territory tenure
	Asymmetry (<i>N</i> = 3)	<u>Fluctuating asymmetry</u>
	Body size (<i>N</i> = 8)	Body (or part of it) mass, length, width, depth, area, or volume
	Brain size (<i>N</i> = 1)	Length, mass, area, or volume of the brain (or part of it)
	Heterozygosity (<i>N</i> = 2)	Multilocus heterozygosity, <u>inbreeding</u>
	Hormones (<i>N</i> = 2)	Androgens, estradiol, fertility phase, ovarian function, progesterone, testosterone
	Individual experience (<i>N</i> = 1)	<u>Direct or indirect experience with individuals of the opposite sex</u> (e.g. mated vs. unmated)
	Sexual signal (<i>N</i> = 4)	Another trait considered a sexual signal
	Traits related to sperm competition (<i>N</i> = 1)	Quantity of seminal fluid, sperm size, sperm viability, testes size

1312 Table S2. Data points reported by meta-analyses that could not be re-extracted ($N = 200$).

1313 Reported IDs refer to the unique identifier for each reported data point across meta-analyses

1314 (see dataset).

Meta-analysis	Reported IDs	Justification
Dougherty 2021	106, 107	Reported data point indicates that both proxy and sexual signal (courtship behaviour) are from the male, but the latter was performed by the female (thus a measure of preference by females)
Dougherty et al. 2023	5028, 5046	Not found in the primary study: data were collected but were not reported
Dougherty et al. 2023	5064	Reported data point indicates that both proxy and sexual signal (courtship behaviour) are from the male, but the latter was performed by the female (thus a measure of preference by females)
Dougherty et al. 2023	5056, 5207	Considered repeated data: "brightness" is the average of other colour measures already reported
Evans et al. 2010	521, 522, 523	Not found in the primary study
Garamszegi & Eens 2004	4783, 4784, 4785, 4786, 4787, 4788, 4789, 4790, 4791	Not found in the primary study: data were collected but were not reported
Gontard-Danek & Moller 1999	648	Not found in the primary study: data were collected but were not reported
Griffith et al. 2006	727, 731	Considered repeated data: reported data point refers to a principal component measure as sexual signal that uses more granular data (hue, brightness, chroma)
Guindre-Parker & Love 2014	6780	Unable to extract this data point: slope without standard error
Hegyi et al. 2015	805, 807	Unable to extract this data point: slope without standard error
Hegyi et al. 2015	810	Not found in the primary study: data were collected but were not reported
Hernandez et al. 2021	4321	Not found in the primary study: data point seems to be related to arrival date instead of attractiveness and it is only approximately reported
Jennions et al. 2001	870	Not found in the primary study: only information given is on offspring survival not on survival of the sexual signal bearer
Mautz et al. 2013	969	Unable to extract this data point: Wald's Chi-Square without p -value

Mautz et al. 2013	994, 1009, 1014	Considered as repeated data: absolute number of live sperm is similar to data already given (number of sperm and percent of live sperm)
Meunier et al. 2011	1287, 1298, 1321	Not found in the primary study
Moller & Alatalo 1999	1423	Not found in the primary study: offspring size reported but not offspring survival
Moller & Jennions 2001	1512	Not found in the primary study: asymmetry not even mentioned
Moller & Jennions 2001	1599	Not found in the primary study: only relative parental care reported, thus not valid for extraction
Moller & Thornhill 1998	6696	Unable to extract this data point: Chi-Square from test with predictor variable with more than one level
Moller & Thornhill 1998	6728	Reported data point appears to refer to territory centrality, which we did not consider a measure of attractiveness
Moller & Thornhill 1998	6697	Not found in the primary study
Moore et al. 2016	1666, 1669	Not found in the primary study: data were collected but were not reported
Moore et al. 2016	1730, 1731	Not found in the primary study: data were collected but were not reported
Nakagawa et al. 2007	1804, 1805, 1806	Not found in the primary study: data appear to be related to the relationship between body index and dominance, not between bib size and body index as reported
Nakagawa et al. 2007	1801	Not found in the primary study: data were collected but were not reported
Nolazco et al. 2022	2805	Considered invalid for extraction: intra-copulation rate as proxy
Nolazco et al. 2022	2813, 2814, 2815, 2816, 2817, 2818, 2819, 2820, 2821	Unable to extract this data point: slope without standard error
Nolazco et al. 2022	1991, 2222, 2503, 2504	Unable to extract this data point: slope without standard error
Nolazco et al. 2022	2091	Not found in the primary study
Nolazco et al. 2022	2337	Not found in the primary study
Nolazco et al. 2022	2287	Not found in the primary study: data were collected but were not reported
Nolazco et al. 2022	1853, 2363	Not found in the primary study: data were collected but were not reported

Parker 2013	6220, 6221	Data point referred to response between cage a and cage b, which were random regarding sexual signal and thus without biological meaning
Parker 2013	5945, 5946	Unable to extract this data point: <i>F</i> -value from test with predictor variable with more than one level
Parker 2013	5460, 5461, 5462	Not found in the primary study
Parker 2013	5573, 5574, 5575, 5576	Data point related to the interaction between testosterone and age on sexual signal, not only age
Parker 2013	5590, 5591, 5592	Data point related to the interaction between age and sexual signal on testosterone, not only sexual signal
Parker 2013	5594, 5595, 5596	Data point related to the interaction between age and sexual signal on corticosterone, not only sexual signal
Parker 2013	6127, 6128	Not enough details on diet and specimens used to be extracted
Parker 2013	6419, 6420, 6422, 6423, 6424, 6440, 6441, 6442, 6458, 6459, 6461, 6462, 6463, 6478, 6479, 6481	Data point related to the interaction between sex and diet on sexual signal, not only diet
Parker 2013	6085	Considered invalid for extraction: moult duration as proxy
Parker et al. 2006	2841, 2842	Unable to extract this data point: <i>U</i> -value without <i>p</i> -value
Parker et al. 2006	2868, 2869, 2870, 2871, 2872, 2874, 2875, 2877, 2878	Not found in the primary study: data were collected but were not reported
Prokop et al. 2012	3096, 3097, 3098, 3099, 3100, 3101, 3102, 3103, 3104, 3105, 3106, 3107	Considered invalid for extraction: genetic correlation
Robinson & Creanza 2019	3227	Not found in the primary study: data were collected but were not reported
Robinson & Creanza 2019	3228	Unable to extract this data point: <i>F</i> -value from test with predictor variable with more than one level
Romano et al. 2017a	3397, 3399, 3401, 3403	Data point related to maternal care and male sexual signals, not female sexual signal
Romano et al. 2017a	3610, 3611, 3615, 3616	Not found in the primary study
Romano et al. 2017a	3480, 3481, 3485, 3488, 3489	Unable to extract this data point: slope without standard error
Romano et al. 2017a	3472, 3474	Unable to extract this data point: <i>F</i> -value from test with predictor variable with more than one level
Romano et al. 2017a	3478	Unable to extract this data point: Chi-Square from test with predictor variable with more than one level
Romano et al. 2017a	3529	Not found in the primary study: asymmetry not even mentioned

Romano et al. 2017a	3525, 3526, 3527, 3528	Unable to extract this data point: slope without standard error
Romano et al. 2017a	3469	Considered as repeated data: same information was extracted from subgroups
Romano et al. 2017a	3494, 3495, 3498, 3499, 3500	Unable to extract this data point: <i>F</i> -value from test with predictor variable with more than one level
Romano et al. 2017a	3519, 3520	Not found in the primary study
Romano et al. 2017a	3638, 3639	Not found in the primary study: data were collected but were not reported
Santos et al. 2011	3832	Not found in the primary study: only territory size and quality were shown but not aggressive interactions
Santos et al. 2011	3895	Not found in the primary study: territory defence was reported but not territory establishment
Simons & Verhulst 2011	3906	Not found in the primary study: only information on attractiveness given is related to diet not sexual signal
Simons et al. 2012	3997	Not found in the primary study: effect of tac only reported in relationship to treatments, no relationship between this measure and sexual signals
Simons et al. 2012	3965	Not found in the primary study: no mention whatsoever of PHA response assay in it
Simons et al. 2012	3920, 3929, 3983, 4058, 4059	Not found in the primary study: experiment involved removing parasites from one group and comparing with a control, all other proxies were related to this setting not to sexual signals
Simons et al. 2012	3956	Not found in the primary study: only plumage colour reported
Soma & Garamszegi 2011	6547	Unable to extract this data point: <i>F</i> -value from test with predictor variable with more than one level
Street et al. 2016	4108, 4114, 4120, 4122, 4124, 4126	Not found in the primary study: data were collected but were not reported
Thornhill & Moller 1998	6666	Unable to extract this data point: Chi-Square from test with predictor variable with more than one level
Thornhill & Moller 1998	6669	Not found in the primary study: data were collected but were not reported
Thornhill & Moller 1998	6667, 6668	Not found in the primary study
Weaver et al. 2018	4526, 4531	Considered as repeated data: same information was extracted from subgroups
Weaver et al. 2018	4529	Considered invalid for extraction: proxy is from both males and females while sexual signal is only from males
White 2020	4555	Not found in the primary study: data were collected but were not reported

White 2020	4685, 4686, 4687, 4688, 4689	Not found in the primary study: species reported in meta-analysis is not even mentioned
White 2020	4690, 4691	Not found in the primary study: no parasite is mentioned
Yasukawa et al. 2010	4750, 4751	The primary study provided the difference in sexual signal between recaptured and non-recaptured individuals, which was considered a measure of survival in this meta-analysis. Yet, we disagree with this interpretation as the authors of the primary study do not suggest this is the case and the time between capture and recapture was short
Yasukawa et al. 2010	4721, 4722, 4723, 4740, 4741	Not found in the primary study: other measures of aggression given, but not dominance rank

1315

1316 Table S3. Slopes from linear regressions between originally reported effect sizes by meta-
1317 analyses and effect sizes that were re-extracted from their sources (but only those that exactly
1318 matched their description). Shaded rows highlight slopes whose 95% confidence interval does
1319 not include 1.

Meta-analysis	Slope	Standard error	95% CI upper bound	95% CI lower bound
Dougherty 2021	0.976	0.101	1.193	0.759
Dougherty et al. 2023	0.956	0.037	1.029	0.883
Fiske et al. 1998	0.254	0.150	0.620	-0.112
Garamszegi & Eens 2004	1.008	0.019	1.047	0.970
Garamszegi & Moller 2004	0.642	0.138	0.981	0.304
Gontard-Danek & Moller 1999	0.828	0.218	1.388	0.268
Griffith et al. 2006	0.696	0.119	0.939	0.453
Hegyi et al. 2015	0.720	0.050	0.839	0.601
Hernandez et al. 2021	0.022	0.132	0.285	-0.241
Jennions et al. 2001	1.026	0.315	1.720	0.332
Mautz et al. 2013	0.619	0.119	0.859	0.380
McLean et al. 2012	0.778	0.055	0.906	0.651
Meunier et al. 2011	0.743	0.248	1.274	0.212
Moller & Jennions 2001	0.311	0.117	0.569	0.053
Moore et al. 2016	0.371	0.311	1.171	-0.429
Nakagawa et al. 2007	0.907	0.043	0.994	0.819
Nolazco et al. 2022	0.755	0.031	0.816	0.693
Parker & Ligon 2003	1.047	0.108	1.311	0.783
Parker 2013	0.460	0.061	0.581	0.339
Parker et al. 2006	0.984	0.011	1.007	0.962
Robinson & Creanza 2019	0.876	0.082	1.043	0.709
Romano et al. 2017a	0.316	0.056	0.428	0.204
Romano et al. 2017b	0.996	0.018	1.036	0.955

Sanchez-Tojar et al. 2018	0.868	0.355	1.736	0.001
Santos et al. 2011	0.800	0.152	1.127	0.472
Simons et al. 2012	0.729	0.087	0.907	0.552
Soma & Garamszegi 2011	0.977	0.041	1.059	0.895
Street et al. 2016	0.123	0.077	0.286	-0.041
Weaver et al. 2018	0.663	0.069	0.800	0.527
White 2020	0.760	0.082	0.929	0.592

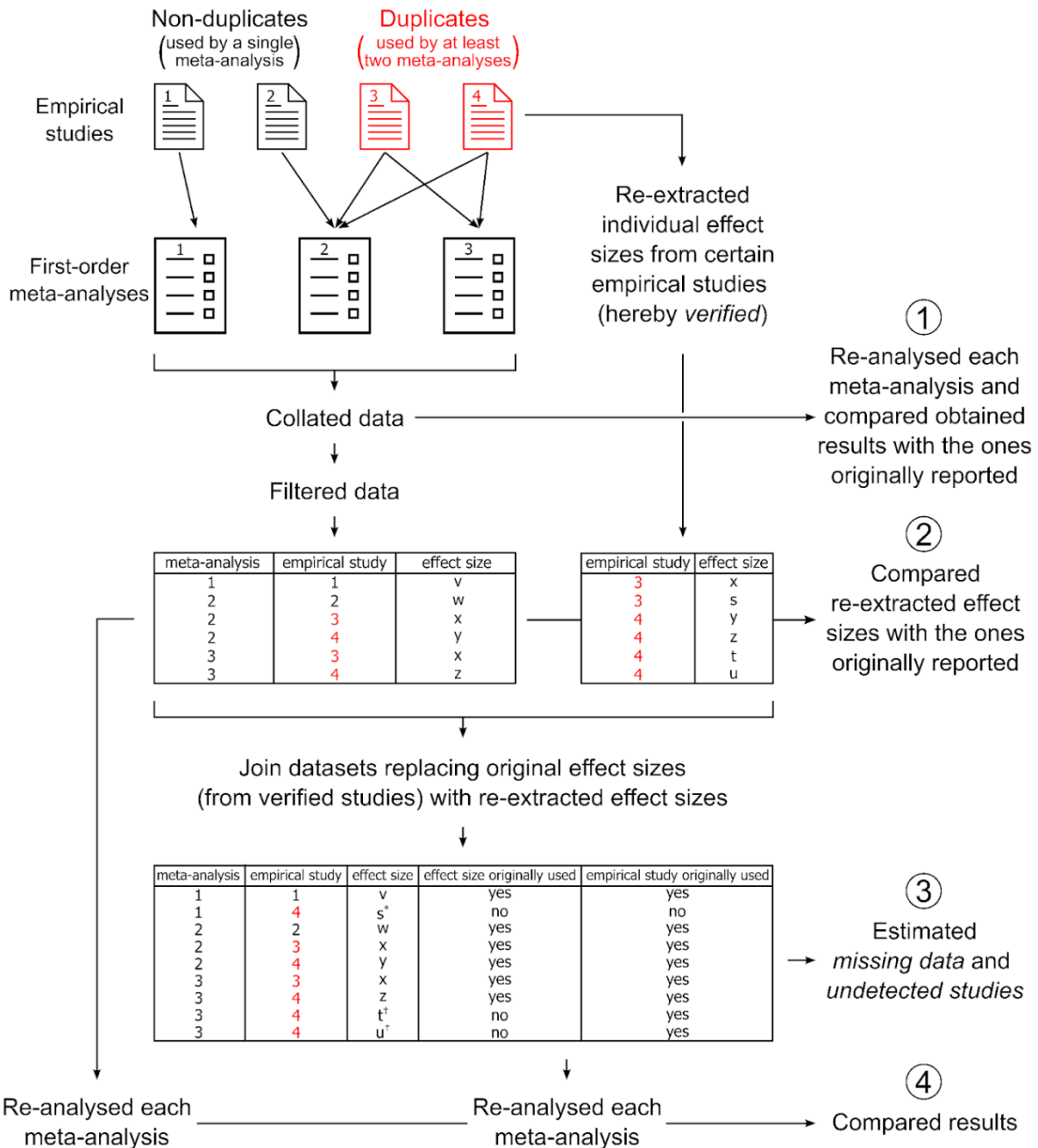
1320

1321 Table S4. Recommendations for meta-analytical research projects.

Item	Recommendation(s)
Data of interest	<ul style="list-style-type: none"> • Detail all criteria and data as much as possible instead of mentioning umbrella terms • Provide data details in a readable format, not only within the dataset
Direction and extraction of effect sizes	<ul style="list-style-type: none"> • Establish a coherent rationale to ascertain direction of effect sizes, fully reporting it in the manuscript • Double-check extracted data (done by a researcher other than the data extractor) • Provide the within-text location of data extracted • Provide equations used to calculate effect sizes, along with assumptions and data transformations utilised

1322

1323



1325

1326 Figure S1. Summarised workflow used in our study, with four main aims. *Examples of effect

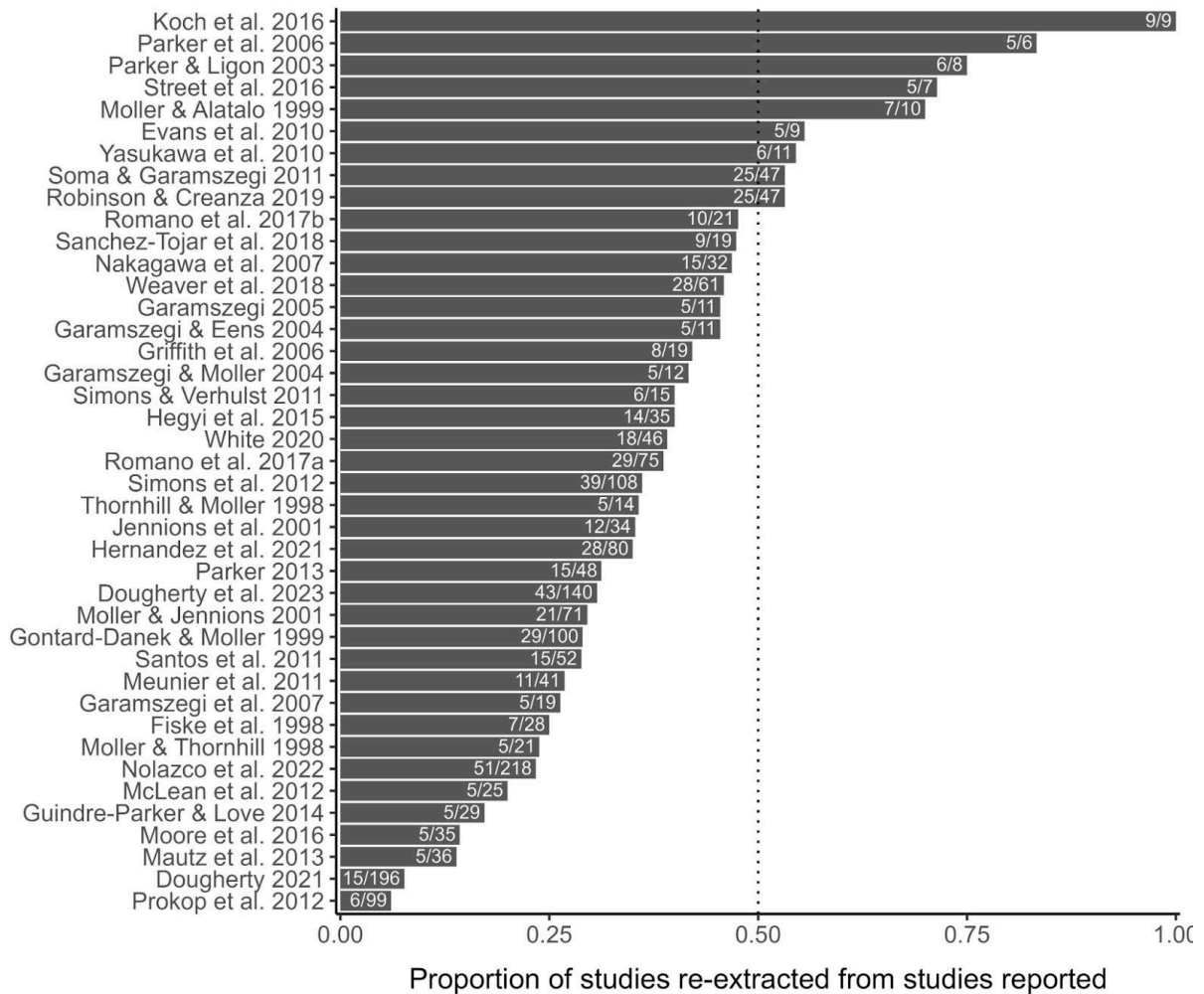
1327 sizes that were not reported in a meta-analysis because the empirical study containing it was

1328 not originally detected (i.e. *undetected studies*). †Examples of effect sizes that were not

1329 originally reported by a meta-analysis even though the empirical study was listed as a source

1330 for other effect sizes (i.e. *missed data*).

1331



1332

1333

1334

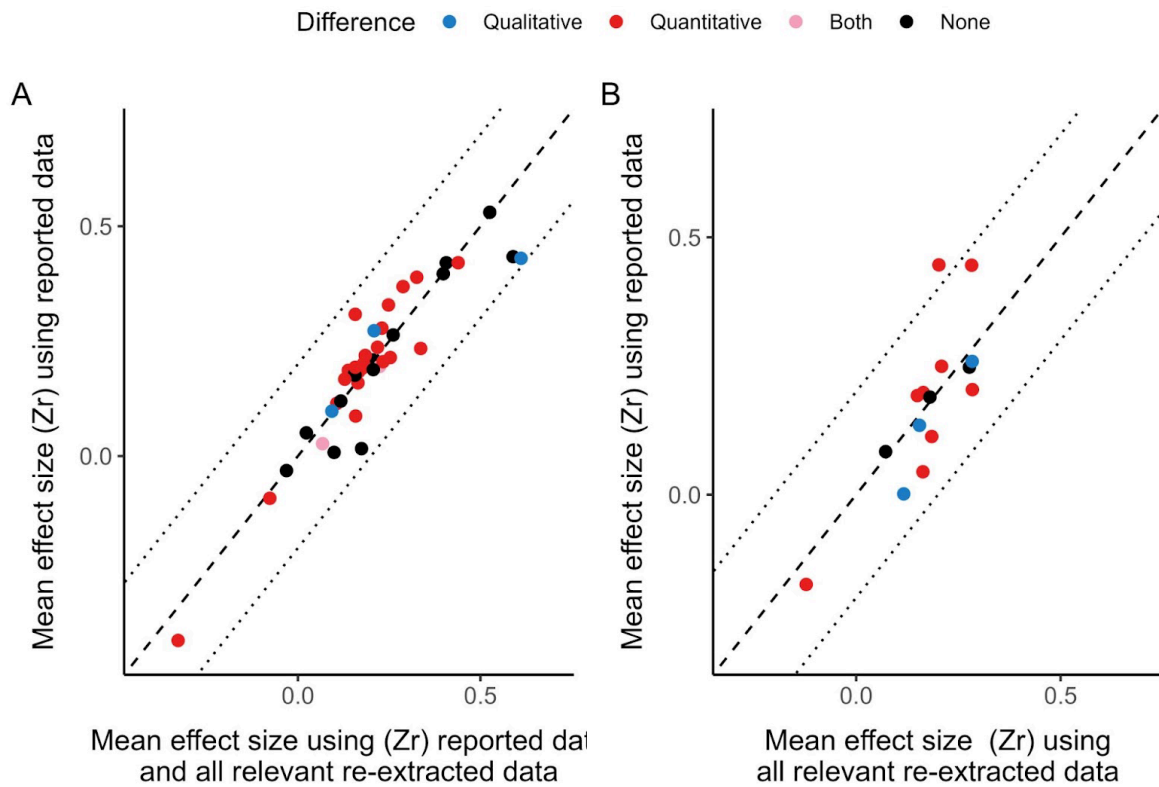
1335

1336

1337

1338

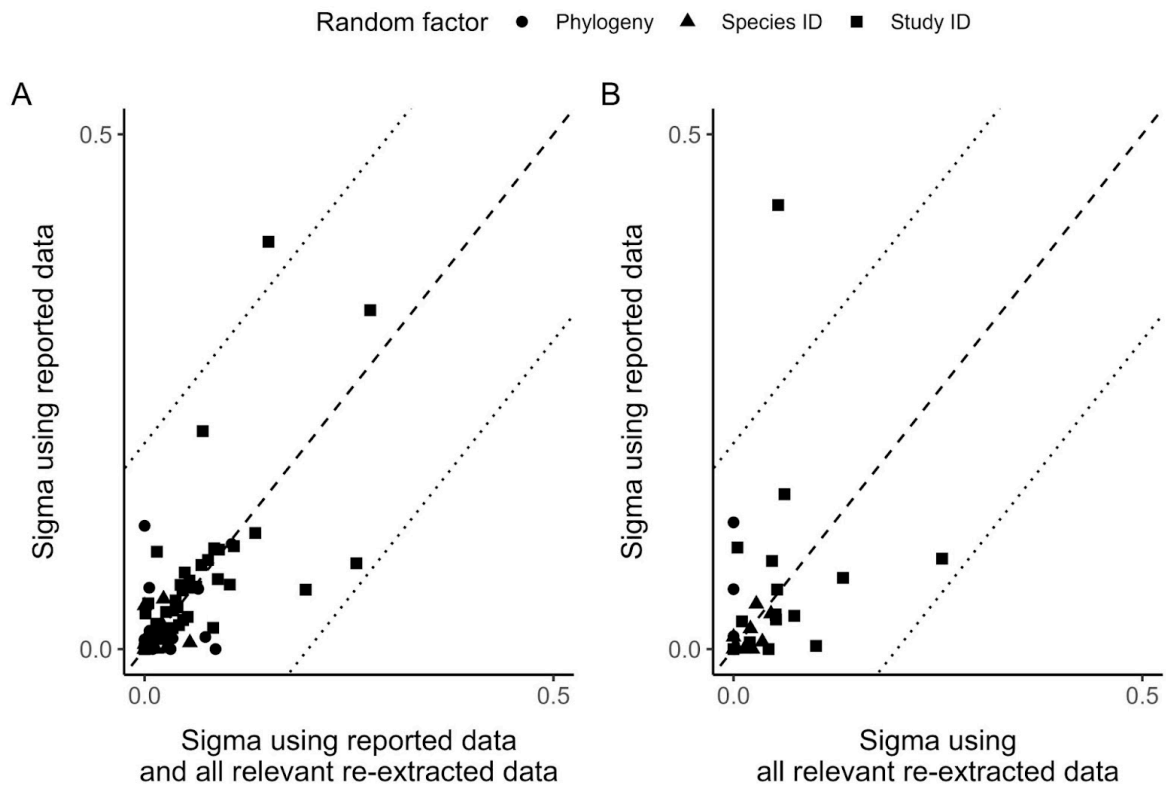
Figure S2. Proportion of empirical studies that we verified (i.e. from which we re-extracted data) from the number of empirical studies reported as sources by each meta-analysis. The number of empirical studies we verified out of the number of empirical studies reported as sources by each meta-analysis is shown within bars. The dotted line highlights when half of the empirical studies were verified.



1339

1340 Figure S3. Comparison between mean effect sizes resulting from analyses using the dataset
 1341 originally reported by meta-analyses and those resulting from analyses incorporating all
 1342 relevant re-extracted data (matched, missed, and undetected). Data from both verified and
 1343 unverified empirical studies are analysed in A, while only data from verified studies are used
 1344 in B. Dashed lines highlight perfect correspondences between variables, while dotted lines
 1345 highlight a difference of 0.2 and -0.2 between variables.

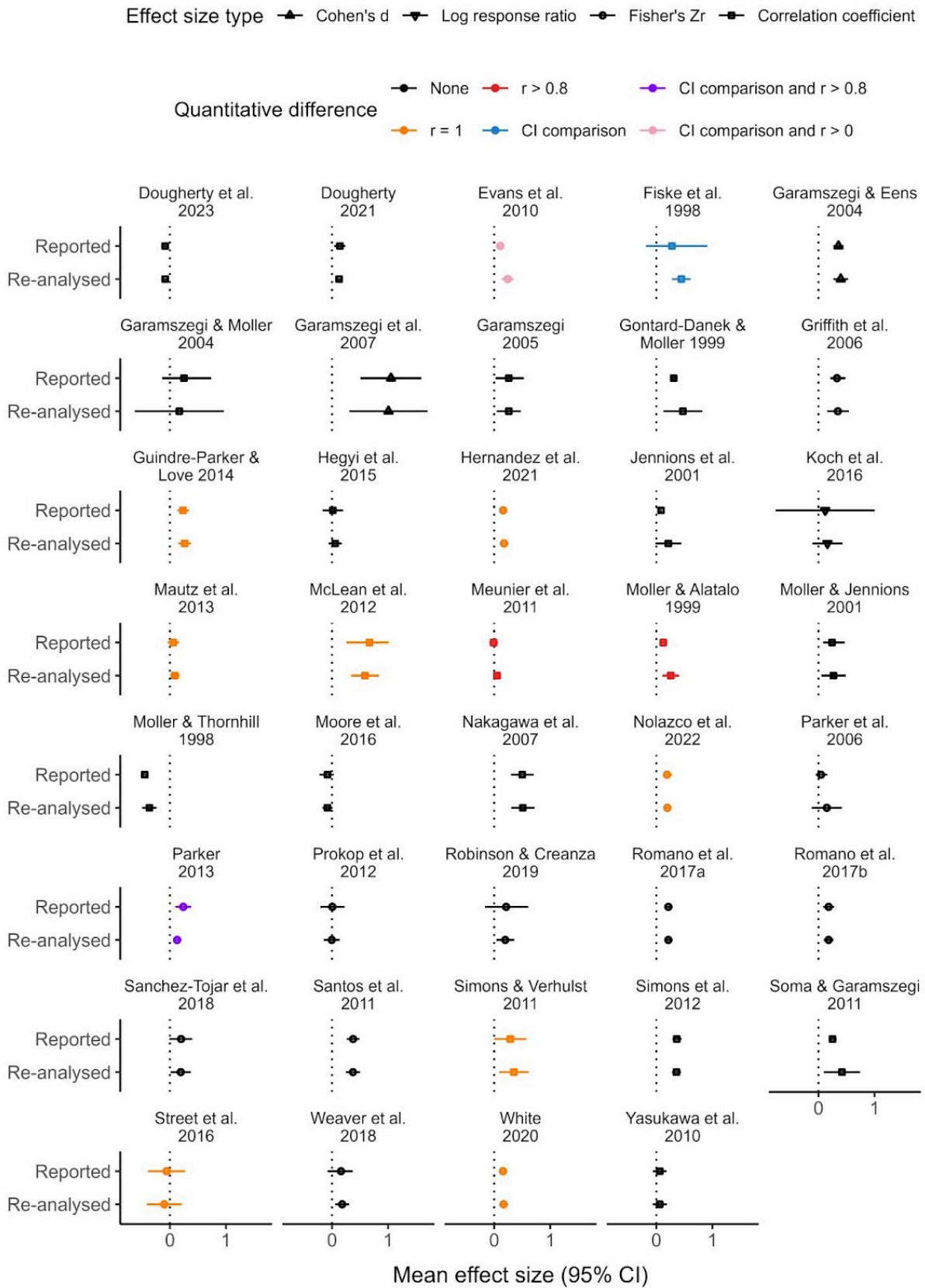
1346



1347

1348 Figure S4. Comparison between sigmas (from each random factor) resulting from analyses
 1349 using the dataset originally reported by meta-analyses and those resulting from analyses
 1350 incorporating all relevant re-extracted data (matched, missed, and undetected). Data from both
 1351 verified and unverified studies are analysed in A, while only data from verified empirical
 1352 studies are used in B. Dashed lines highlight perfect correspondences between variables, while
 1353 dotted lines highlight a difference of 0.2 and -0.2 between variables.

1354

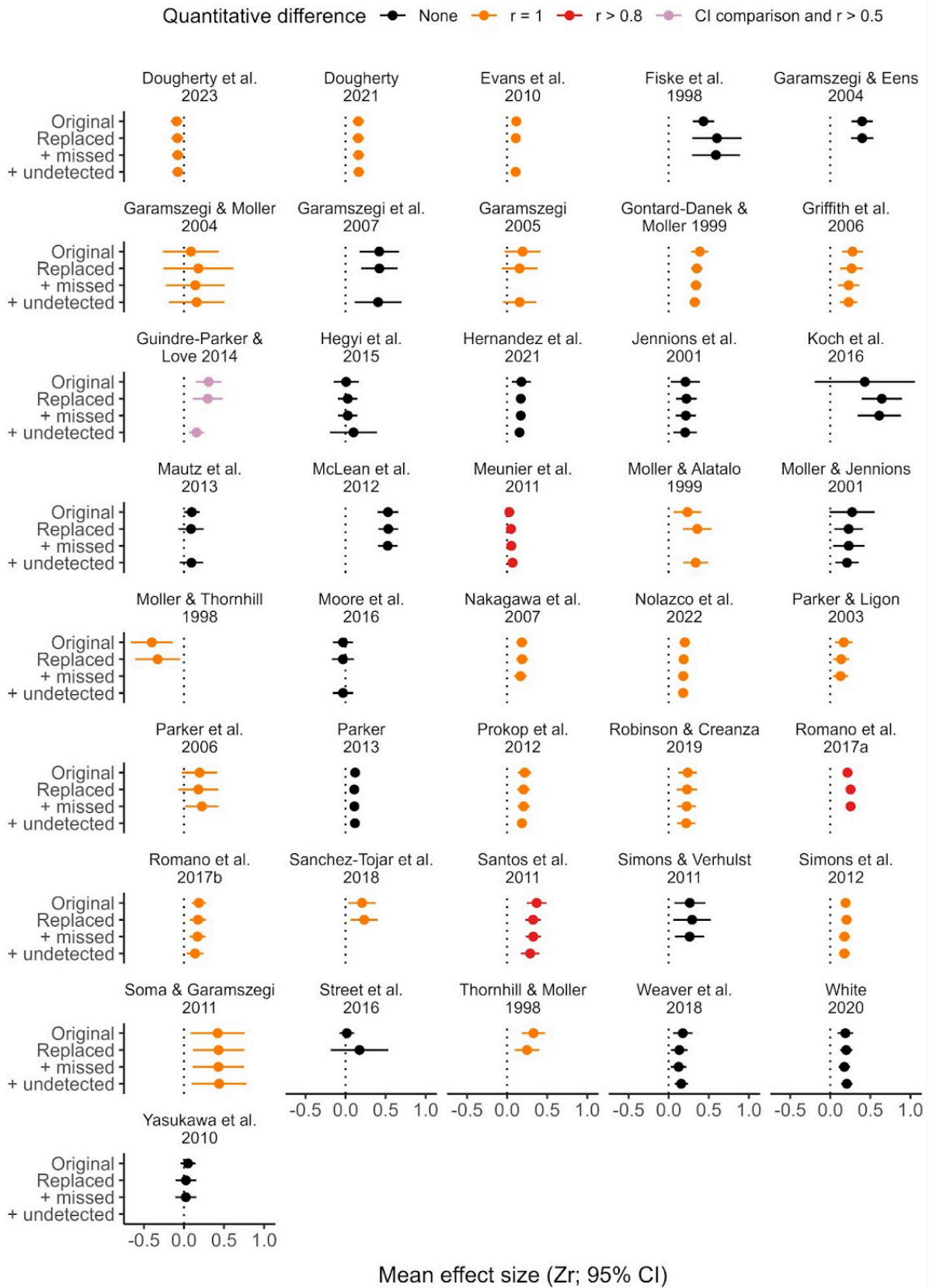


1355

1356 Figure S5. Originally reported and re-analysed mean effect sizes of 39 meta-analyses.

1357 Colourful points indicate quantitative pairwise differences with absolute z-score greater than

1358 1.96 (using different values for r) and/or when the 95% confidence interval of the re-analysed
1359 mean effect size did not include the value of the reported mean effect size (“CI comparison”).
1360



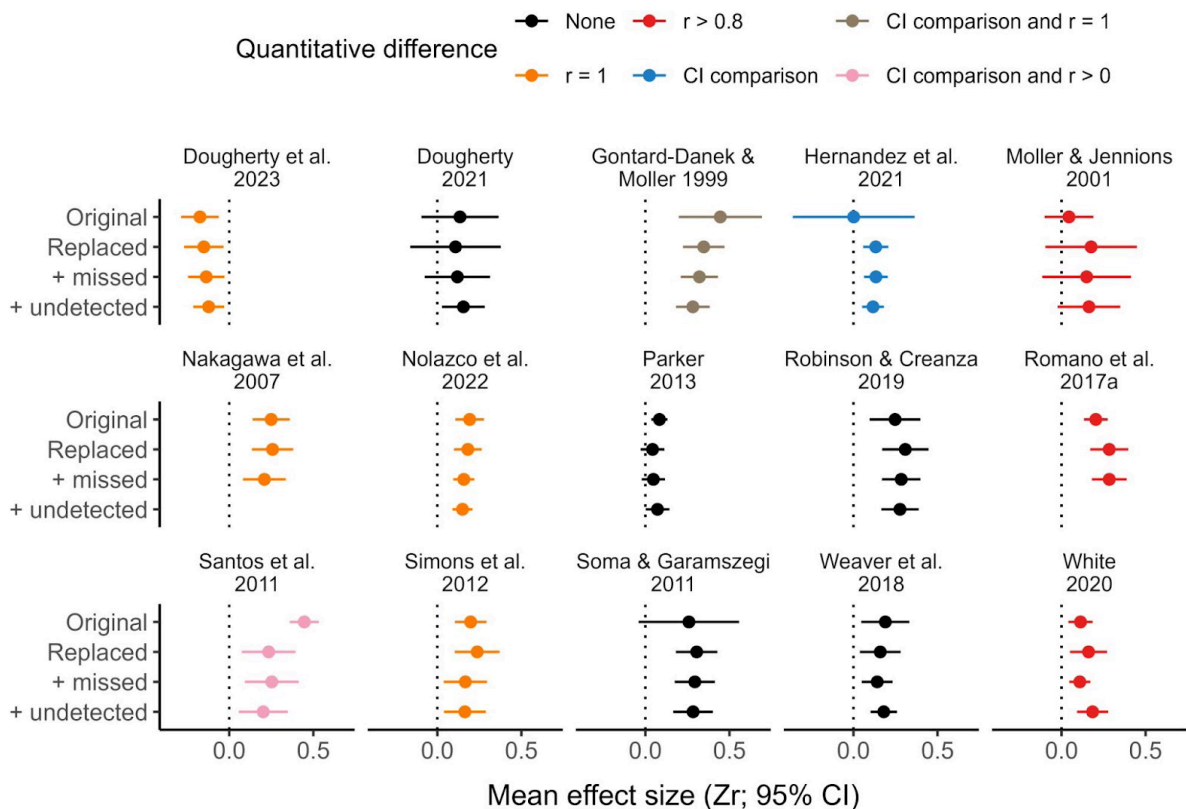
1361

1362 Figure S6. Mean effect sizes from up to four distinct meta-analytical models for each of 41

1363 meta-analyses, using data from all empirical studies (both verified and unverified by us, see

1364 details in-text). Dotted lines highlight zero. Colourful points indicate quantitative pairwise
 1365 differences (between the first and the last result shown within each subplot) with absolute z -
 1366 score greater than 1.96 (using different values for r) and/or when the 95% confidence interval
 1367 of the re-analysed mean effect size did not include the value of the reported mean effect size
 1368 (“CI comparison”).

1369



1370

1371 Figure S7. Mean effect sizes from up to four distinct meta-analytical models for each of 15
 1372 meta-analyses, using data from verified empirical studies (see details in-text). Dotted lines
 1373 highlight zero. Colourful points indicate quantitative pairwise differences (between the first
 1374 and the last result shown within each subplot) with absolute z -score greater than 1.96 (using
 1375 different values for r) and/or when the 95% confidence interval of the re-analysed mean effect
 1376 size did not include the value of the reported mean effect size (“CI comparison”).

1377

1378 **Supplementary information S1. Changes from the pre-registration**

1379 We replaced terms used in our pre-registration regarding our goals (e.g. “precision”,
1380 “completeness”, “detectability”) with other terms (e.g. number of undetected studies) to
1381 improve the clarity of our manuscript. Furthermore, we created different proxy categories (e.g.
1382 “heterozygosity”, “hormones”, “territory”) and reorganised proxies’ master categories to
1383 improve clarity and better show their diversity (see Table S1).

1384

1385 **Supplementary information S2. Data of interest**

1386 First, for the purpose of our study, we interpret ornaments and sexual signals as “non-ordinary
1387 and non-weapon traits potentially used for mate attraction”. This vague definition is intended
1388 to match the rationale observed in the literature, in which conspicuous and/or sexually
1389 dimorphic traits are often subjectively designated as sexual signals (see also the discussion in
1390 ref. ⁷). In practice, this means that we included all traits considered as sexual signals by meta-
1391 analyses’ authors, except for obvious weapons (i.e. antlers, spurs, and horns), ordinary traits
1392 (i.e. body size; body parts without referring to it as a sexual trait, e.g. ref. ³²³), lek related
1393 measures (e.g. lek attendance, distance from lek centre, etc.; see ref. ¹⁷) colour bands put on
1394 animals by researchers (see ref. ³²⁴), and comparisons between “attractive” vs. “unattractive”
1395 based on mate preferences without a direct measurement of a sexual signal (e.g. refs. ^{325,326}).
1396 In addition, we only considered asymmetry as a sexual signal when it was related to a trait that
1397 itself was considered a sexual signal. The only extended phenotypes we considered as valid
1398 sexual signals were measures of structures to attract mates (e.g. bowers, nests). This means that
1399 data related to leks (e.g. lek attendance, distance to centre of the lek) and territory quality were
1400 not considered sexual signals. Second, we are interested in how sexual signals relate to distinct
1401 conditions, fitness proxies, and individual traits (Table S1). However, we excluded effect sizes
1402 related to traits of focal individuals’ mates (e.g. parental care provided by mate), assortative

1403 mating, heritability (i.e. proportion of phenotypic variance explained by genetic variance), and
1404 offspring sex ratio because we deemed these estimates may not be linked to benefits to or reflect
1405 patterns of sexual signal bearers. Third, we excluded effect sizes related to humans to avoid
1406 confounding our cultural influences with biological aspects of non-human animals (see also
1407 ref. ⁷). Fourth, we excluded effect sizes with (exact) zero values from three meta-analyses that
1408 purposefully included non-significant results as such (refs. ^{14,15,39}), unless we were able to
1409 match them with at least one re-extracted data point (see *Re-extractions* section).

1410

1411 **Supplementary information S3. Priority and selection of empirical studies for data re-** 1412 **extraction**

1413 We classified empirical studies (used as sources for data extraction in meta-analyses) according
1414 to the following categories. We first classified all empirical studies listed as sources for at least
1415 three meta-analyses as “high priority” (n = 82). Then, we classified the remaining duplicates
1416 (i.e. empirical studies that were used exactly by two meta-analyses) as: (1) “low priority” when
1417 empirical studies whose all effect sizes were identical (both in their description and in their
1418 value) across meta-analyses (n = 19); (2) “medium priority” when empirical studies contained
1419 effect sizes regarding different traits and measurements (e.g. distinct sexual signals or fitness
1420 proxies) across meta-analyses (n = 79); (3) “high priority” when empirical studies generated
1421 similarly described effect sizes (e.g. same sexual signal and fitness proxy) but with different
1422 (and thus contradicting) values or sample sizes across meta-analyses (n = 136). We primarily
1423 targeted high priority duplicates (n = 218) for effect size re-extraction, although not all meta-
1424 analyses contained these duplicates (e.g. refs. ^{18,50}). We thus established a minimum of five
1425 empirical studies for re-extraction of effect sizes per meta-analysis, adding medium priority
1426 duplicates (n = 12) and non-duplicated empirical studies (n = 19) when necessary to meet this
1427 quota.

1428

1429 **Supplementary information S4. Meta-analyses' inclusion criteria**

1430 Below, we provide excerpts from the meta-analyses in our dataset related to their data of
1431 interest and inclusion criteria. We also provide the issues we detected for each meta-analysis
1432 (if any).

1433

1434 *Dougherty (2021)*

1435 Originally reported

1436 “I focused on sexual signalling behaviour, including long-range attraction signals and short-
1437 range courtship signals. I excluded measures of investment in non-behavioural signals, such as
1438 ornaments or advertising colours, although I did not exclude display behaviours that involved
1439 such ornaments. I excluded intrasexual signals or signals for which a primary intrasexual
1440 function could not be ruled out. However, I acknowledge that all sexual signals probably
1441 function intrasexually to some extent. I also included several lekking species for which displays
1442 probably signal to potential mates and rivals because I consider the primary function of leks to
1443 be mate assessment. I included studies examining both male and female sexual signalling. I
1444 included acoustic, visual, olfactory (pheromone) and tactile display behaviours. I focused on
1445 behavioural traits that reflect the motivation to signal (for example, courtship latency) or
1446 energetic investment in signalling (such as signalling duration, rate and some measures of
1447 intensity). For acoustic signals, I included measures of call loudness, except when related to
1448 body size (because call loudness may be constrained by the size of the sound-producing
1449 organs). I excluded measures of signal complexity because this does not necessarily relate to
1450 overall energetic investment per se. For acoustic signals, I excluded measurements of call
1451 pitch/frequency and fine-scale temporal components of a call. For pheromones, I excluded
1452 measurements of pheromone composition but included measures of time spent releasing

1453 pheromone and pheromone titre if measured outside the body (I excluded measures of
1454 pheromone titre in dissected glands or bodies).

1455 I included studies examining how sexual signalling behaviour related to age. I included
1456 studies in this category if all individuals were sexually mature and age was not confounded
1457 with body size. Importantly, age is often confounded by mated status, especially for wild
1458 individuals and this may influence the motivation to signal independently of age. Therefore, I
1459 only considered studies examining age-related signalling in virgins. This was necessary
1460 because few studies record both age and mated status in a way that allows their independent
1461 effects to be estimated. [...] I included studies examining how sexual signalling behaviour
1462 related to other measures of attractiveness. In these studies, attractiveness was determined in
1463 relation to behavioural or morphological traits which were suggested to signal mate quality
1464 (either genetic quality or current condition) or have been shown to be used in mate choice. I
1465 included studies relating signalling behaviour to: (1) song quality, (2) ornament size, (3)
1466 ornament or body colouration or brightness, (4) morphological asymmetry, (5) inbreeding, (6)
1467 territory or nest quality and (7) social rank. Individuals were assumed to be attractive if they
1468 exhibited high-quality song, large ornaments, bright or intense colouration, were outbred, with
1469 low asymmetry, of a high social rank and had high-quality territories or nests. I included tests
1470 of social rank only when signalling was recorded in the absence of rivals; this is important
1471 because high-ranking individuals may suppress the behaviour of subordinates.

1472 This category included studies relating signalling to body length, weight or some proxy
1473 length measurement (for example, leg length, wing length and pronotum width).

1474 I included studies examining how sexual signalling behaviour related to several
1475 measures of condition: (1) diet or food level, (2) the relationship between body size and weight,
1476 (3) direct measurements of body lipid content or plasma metabolite level and (4) environmental
1477 conditions that could alter physiological stress in the short-term (oxygen, carbon dioxide and

1478 water acidity in aquatic environments). I used several indirect, morphological measures of
1479 condition, although I note that several common measures have been criticised. I assume that
1480 individuals were in poor condition if they were relatively light for a given body size, with low
1481 lipid stores, or had experienced low food levels, poor-quality diets or stressful environments. I
1482 excluded studies examining how signalling behaviour related to physiological markers of
1483 stress, as stress responses are typically short-lived and may have a complex relationship with
1484 condition.

1485 I included studies comparing signalling between virgins and once-mated individuals. I
1486 excluded tests related to the number of matings above one or other forms of mating experience
1487 (that is, the phenotype of previous mating partners).

1488 I included studies examining how sexual signalling behaviour related to parasite load.
1489 This category included studies measuring the presence or number of external (lice, mites and
1490 crustaceans) or internal (acanthocephalans, nematodes, platyhelminthes, alveolates, fungi,
1491 bacteria and viruses) parasites. I included sexually transmitted parasites, even in cases where
1492 host behavioural changes were suggested to be due to parasite manipulation. I excluded studies
1493 relating behaviour to the presence of endosymbionts in insects, as they may have parasitic or
1494 mutualistic effects on their hosts. Finally, I also excluded studies examining the effect of
1495 controlled immune challenges on host behaviour; for example, by introducing sterile pellets or
1496 inactivated pathogens into the host. This is because any consequences for host condition are
1497 indirect in such cases, caused by upregulation of the host immune system and are typically
1498 short-lived.”

1499

1500 Issues

1501 1) The dataset contains sexual signals that are behavioural products (i.e. extended phenotype;
1502 e.g. number of pillars built, domes built, and building rate in different species of crabs). The

1503 manuscript does not mention this or other extended phenotypes. We thus considered data on
1504 extended phenotype as valid sexual signals for matching purposes.

1505 2) The manuscript does not mention brood size manipulation as a valid proxy even though
1506 several data points related to this proxy are in the reported dataset. We thus considered brood
1507 size manipulation as a valid proxy for matching purposes.

1508

1509 *Dougherty et al. (2023)*

1510 Originally reported

1511 “To be included in the analysis, a study had to (a) present data for sexually mature individuals
1512 of a non-human animal species, (b) report within-species variation in a morphological,
1513 behavioural, or extended ornament, (c) report some measure of parasite load for the same host
1514 individuals, and (d) provide sufficient statistical information for an effect size to be calculated.

1515 While parasite-mediated sexual selection hypotheses tend to focus on elaborate male
1516 morphological ornaments, such as plumage or bright skin patches, we expand the scope of our
1517 data set by (a) considering display behaviours and extended ornaments such as the bowers of
1518 bowerbirds because these potentially honestly indicate courter condition or quality and (b)
1519 considering female ornamentation because mating preferences in relation to partner condition
1520 and quality are seen in both sexes.”

1521

1522 Issues

1523 1) Many aspects from acoustic signals (e.g. repertoire size, song switching rate) seemed to have
1524 been excluded without mentioning this in-text, which we interpreted as a hidden criterion.

1525 2) The dataset contains details only for some sexual signals, while others remain unclear (e.g.
1526 "colour" without specifying body part or how it was measured).

1527

1528 *Evans et al. (2010)*

1529 Originally reported

1530 “We conducted a search of peer-reviewed journals to identify studies that examined ventral
1531 plumage reflectance in free-living great tit populations and also contacted known researchers
1532 directly to ask them to contribute data. [...] The data in the meta-analysis included measures
1533 derived from spectrometry, colorimetry, digital photography, and colour tiles, and were based
1534 on measures of live birds and of sampled feathers. As a result of the limited number of studies
1535 represented, we did not test for systematic differences with respect to methodology. We did not
1536 impose restrictions on the colour measures that would be included in our meta-analysis; rather,
1537 we invited researchers to use colour measures of their choice. [...] To examine the overall extent
1538 of sex- and age-dependent colour expression, we compiled data on the mean, standard deviation
1539 and sample size for each sex (males versus females) and, separately, each age class (first-years
1540 versus older birds).”

1541

1542 Issues

1543 No issues detected.

1544

1545 *Fiske et al. (1998)*

1546 Originally reported

1547 “We included studies if they contained correlates (parametric Pearson or nonparametric
1548 Spearman correlation coefficients) between traits and male mating success. Assessing male
1549 mating success across several taxa is not an easy task. In many lekking species (e.g., certain
1550 insects) copulations are hard or even impossible to observe. Two of the studies included
1551 presented their results as correlates of female visitation rates. This is probably legitimate
1552 because this measure is highly intercorrelated with the number of copulations. Still, our

1553 analyses were mainly based on studies that obtained correlations with male copulatory success.
1554 This led to rejection of some ungulate studies because they reported correlations with territory
1555 mating success. [...] Attendance: We included studies that provided information about the time
1556 individual males spent on the lek in relation to male mating success. This variable was
1557 measured both as the proportion of time compared to other males and as absolute time within
1558 a given period. Display frequency: Males vocalisations are probably important traits when
1559 females make their mate choice decision. Therefore we included measurements of display
1560 frequencies (calls per unit time or proportions of time calling) because such measurements can
1561 easily be compared across species. Spectral aspects of calls have also been analysed in several
1562 studies, but we have chosen not to include them in our analyses because we found them
1563 impossible to categorise uniformly.”

1564

1565 Issues

1566 1) All data reported were related to mating success, but authors never provided the specific
1567 measure for each data point or study that represented this proxy. We thus accepted all proxies
1568 of mating success for matching purposes.

1569

1570 *Garamszegi (2005)*

1571 Originally reported

1572 “I performed some simple meta-analyses on the available data to test for general patterns in the
1573 intraspecific association between parasitism and bird song. [...] I used measures of immune
1574 function and parasite prevalence interchangeably, as was done in similar meta-analyses.”

1575

1576 Issues

1577 No issues detected.

1578

1579 *Garamszegi & Eens (2004)*

1580 Originally reported

1581 “We collected published results of studies investigating correlatively the intraspecific
1582 association between male repertoire size and/or song length and the size of HVC and/or RA.
1583 Hence, we did not include studies that investigated seasonal variation in RA and HVC volumes
1584 and song by sampling males in different seasons, or compared groups of birds experiencing
1585 different tutoring regimes or originating from different populations. [...] As some studies
1586 distinguished between the absolute and relative volumes of brain nuclei by taking or not the
1587 covariation with telencephalon into account, we also estimated effect sizes for absolute and
1588 relative volumes of the HVC and RA.”

1589

1590 Issues

1591 No issues detected.

1592

1593 *Garamszegi & Moller (2004)*

1594 Originally reported

1595 “We collected published and, in an effort to control for publication bias, unpublished results of
1596 studies investigating the association between songs and extrapair paternity within species. [...]
1597 We included analyses from t-tests (two-tailed), or other equivalent statistics testing the null
1598 hypothesis that males with more extravagant song display (measured as repertoire size, song
1599 rate, song length, and performance characteristics) have similar paternity in their own nests
1600 than do males with less extravagant song features.”

1601

1602 Issues

1603 1) Manuscript is unclear on which measures of paternity were included, so we considered all
1604 extra- and within-paternity measures as valid for matching purposes.

1605 2) Manuscript is unclear on which song traits were included while the dataset contains various
1606 song traits. We thus considered all possible song traits as valid for matching purposes.

1607

1608 *Garamszegi et al. (2007)*

1609 Originally reported

1610 “[...] we collected published results of studies investigating the relationship between age and
1611 song using cross-sectional approaches. We included analyses based on t-test (two-tailed), or
1612 other equivalent statistics testing the null hypothesis that yearling males have similar repertoire
1613 to older males.”

1614

1615 Issues

1616 No issues detected.

1617

1618 *Gontard-danek and Moller (1999)*

1619 Originally reported

1620 “We searched the literature for correlation coefficients, or other statistics that could be
1621 converted into correlation coefficients, based on the relationship between sexual selection and
1622 the expression of secondary sexual characters. [...] We considered whether effect size differed
1623 in relation to the currency used to estimate success. These were classified as mate preference,
1624 mating success, breeding success, or paternity.”

1625

1626 Issues

1627 1) The manuscript mentions proxies without details ("mate preference, mating success,
1628 breeding success, or paternity"). We thus considered all measurements related to these proxies
1629 (except those related to timing or latency) as valid for matching purposes.

1630 2) The dataset contains one data point related to territory quality even though the manuscript
1631 only mentioned other proxies. We thus included relevant data that used territory quality as
1632 proxies.

1633 3) The manuscript only mentions "secondary sexual characters", which apparently excludes
1634 behavioural sexual signals as these were not present in the dataset. We thus interpreted this as
1635 a hidden criterion and did not consider behavioural sexual signals as valid for matching
1636 purposes.

1637 4) Dataset only contained measurements on timing or latency related to breeding, egg laying,
1638 mating, and pairing. We thus interpreted this an omitted criterion and excluded other similar
1639 measures (e.g. arrival date) for matching purposes.

1640

1641 *Griffith et al. (2016)*

1642 Originally reported

1643 "A number of experimental manipulations have been conducted to test the condition
1644 dependence of carotenoid- and melanin-pigmented ornaments, especially in birds. These
1645 studies manipulated condition-related factors including diet, parasite load and parental effort.
1646 We did not include studies that manipulated the carotenoid content of the diet because the
1647 absence of carotenoids in the diet must lead to the absence of carotenoids in the body, and thus
1648 a response to this treatment does not demonstrate condition dependence of a carotenoid-
1649 pigmented trait.

1650 Pigmented ornaments were assessed for changes in area, brightness, hue, saturation,
1651 chroma, principal component-defined colour variation, or other colour score. We attempted to

1652 locate all such studies from birds for inclusion in meta-analyses. All studies included in the
1653 meta-analyses examined the response of traits that appear ornamental and thus may have
1654 evolved in response to sexual selection.”

1655

1656 Issues

1657 1) The dataset contains a data point related to assessment of a sexual signal before and after an
1658 immune challenge. We thus considered other immune challenges as valid for matching
1659 purposes if they occurred the same way, which does not include the relationship between sexual
1660 signals and quick immune evaluations (e.g. PHA responses, blood tests).

1661

1662 *Guindre-Parker & Love (2014)*

1663 Originally reported

1664 “We surveyed published studies on condition-dependent melanin plumage. We characterised
1665 species as either achromatic (primarily black, brown, grey or white) or chromatic (also
1666 possessing a carotenoid-based red, orange or yellow ornament) to investigate whether the
1667 presence of a carotenoid-based signal can influence the condition-dependence of a melanin
1668 signal. We excluded species with iridescent plumage as it is primarily structurally based, as
1669 well as because it is difficult to determine how birds perceive iridescence without more
1670 complex visual modelling.”

1671

1672 Issues

1673 1) Unclear how authors verified which sexual signals were melanin-based.

1674 2) The manuscript does not specify proxies included (i.e. what condition-dependent entails), so
1675 only proxies present in the dataset were considered to be valid (i.e. hidden criterion for
1676 matching purposes).

1677

1678 *Hegyi et al. (2015)*

1679 Originally reported

1680 “We therefore used correlative studies and took into account alternative explanations arising
1681 from the non-experimental situation when interpreting our results. [...] Our first goal was to
1682 assess the overall evidence for a correlation between ornamentation and the feeding rate of the
1683 ornament bearer or its partner. [...] Effect sizes were converted so that the positive sign
1684 indicated a positive relationship of feeding rate with plumage colour expression. [...] Due to
1685 their negligible number, we had to remove results on incubation feeding of females by males
1686 (five results from three populations of three species).

1687 Our search protocol also encountered nine results from four populations of one species
1688 (the barn swallow) that concerned feather length as an ornament, which were removed from
1689 the dataset. Of the remaining 75 effect sizes, we removed results on white patches due to their
1690 prohibitively small number (n=2 from a single species) and also retained for each population
1691 of a given species only one measure of the same colour type (n=1 point removed; we chose the
1692 point with the more complete information). For each ornament analysed here, the proximate
1693 origin of colour variation (carotenoid, pheomelanin, eumelanin, structural) was clear from the
1694 species-specific literature. [...] Plumage colour had been measured in different ways depending
1695 on the colour category and study approach, and the following changes were regarded as
1696 increased colour expression: for carotenoid colour, increased saturation, chroma or patch size;
1697 for pheomelanin colour, increased saturation, increased chroma, increased patch size or
1698 reduced brightness; for eumelanin colour, increased patch size (no other measures were used
1699 in any relevant study); and for structural colour, more UV-shifted hue or increased saturation.”

1700

1701 Issues

1702 1) The manuscript mentioned that only non-experimental approaches were valid but the dataset
1703 included cases of experimental approaches. Nonetheless, we followed the original rule given
1704 in the manuscript for matching purposes for other studies.

1705

1706 *Hernandez et al. (2021)*

1707 Originally reported

1708 “[...] we evaluated the evidence of an association of colourful female ornaments with the
1709 condition, reproductive performance and male preferences. [...] Three moderators were
1710 considered in the model evaluating the association between female colour and condition:
1711 Condition proxy (residual body mass, immune response -humoral or cellular -, and parasite
1712 load), ornament type (feathers or integuments), and coloration type (carotenoid-dependent,
1713 melanin-based, or structural). In the model evaluating the association between female colour
1714 and reproductive performance, we used the same ornament type and coloration type moderator
1715 variables but replaced the condition proxy with a reproductive performance proxy (laying date,
1716 clutch size, or fledging success). When clutch size and fledging success were assessed in the
1717 same study, only fledging success was considered. In the analyses exploring the relationship
1718 between female colour and male mating preferences, only four effect sizes were available for
1719 melanin-based and structural colours (one and three, respectively), so we fit this model
1720 including effect sizes only from studies evaluating carotenoid-dependent colorations (n = 11).
1721 The ornament type (feathers or integuments) was included as a moderator.

1722 The intensity of coloration was estimated using the number and/or size of coloured
1723 structures/patches, spectrophotometry, digital image analysis (RGB or LBA), and visual rank
1724 scales (colour charts and visual rank score).”

1725

1726 Issues

1727 1) The manuscript mentions immune responses as valid measures for inclusion, but does not
1728 discuss which ones exactly. We only included those which we found matches in the dataset (i.e.
1729 hidden rule), which are phytohemagglutinin response and hematocrit.

1730 2) The manuscript mentions that only fledging success was extracted from studies that assessed
1731 both clutch size and fledging success. However, we detected cases in which both of these
1732 measures were present in the reported dataset from the same study (e.g. ^{183,248}). Nonetheless,
1733 we followed the original rule given in the manuscript for matching purposes for other studies.

1734 3) The manuscript mentions that only laying date, clutch size, and fledgling success were valid
1735 measurements of reproductive success. However, the dataset contained a case in which
1736 breeding success (binary variable) was mislabelled as clutch size. The dataset also contained
1737 cases with hatching date (e.g. ref. ²⁴⁸) and date of first nest (e.g. ref. ²¹³). Nonetheless, we
1738 followed the original rule given in the manuscript for matching purposes for other studies.

1739 4) The manuscript mentions that valid condition measures are "residual body mass, immune
1740 response -humoral or cellular -, and parasite load", but dataset also includes cases in which
1741 body mass is labelled as a condition measure (e.g. refs. ^{171,174,185}). Nonetheless, we followed
1742 the original rule given in the manuscript for matching purposes for other studies.

1743

1744 *Jennions et al. (2001)*

1745 Originally reported

1746 "We calculated the effect size as the Pearson product-moment correlation coefficient (r)
1747 between trait expression and a measure of survival rate. [...] The estimate of survival was based
1748 on a comparison between living and dead individuals or an estimate of adult life span/"days
1749 alive".

1750 Secondary sexual traits were then scored as being morphological or behavioural
1751 characters. We made these comparisons because secondary sexual characters may display a

1752 higher degree of condition dependence than ordinary morphological traits. If so, they should
1753 be associated with more positive effect sizes.”

1754

1755 Issues

1756 1) The manuscript does not contain information on which sexual signals were considered valid
1757 for inclusion. We thus considered all sexual signals as valid for matching purposes.

1758

1759 *Koch et al. (2016)*

1760 Originally reported

1761 “We included only studies (1) reporting the level of carotenoid supplementation as well as the
1762 food source provided; (2) including data on both carotenoid-supplemented and control groups
1763 of individuals; (3) reporting the values of plasma carotenoid levels and/or coloration; (4) not
1764 repeating measures on the same group of birds that were reported in a study already
1765 incorporated into the meta-analysis (a potential source of pseudoreplication); (5) testing adult
1766 male birds rather than nestlings (in which both carotenoid physiology and ornamental function
1767 differ greatly from sexually reproducing adult birds, and the quantity of carotenoids acquired
1768 from egg yolk or parental provisioning is often unknown); and (6) supplementing with only the
1769 carotenoids lutein and/or zeaxanthin, the most prevalent carotenoid pigments in the avian diet.
1770 With the exception of one study supplementing with only lutein, all studies included in our
1771 meta-analysis supplemented primarily with lutein and trace amounts of zeaxanthin (e.g., 20:1
1772 ratio of lutein:zeaxanthin). [...] Because 16 of 19 studies investigated songbird species (order
1773 Passeriformes), we excluded one study of red junglefowl (*Gallus gallus*), one study of mallards
1774 (*Anas platyrhynchos*), and one study of kestrels (*Falco tinnunculus*) to capture the majority of
1775 available data while avoiding comparing data from phylogenetically distant taxa with different
1776 physiologies. We also excluded one study on society finches (*Lonchura striata domestica*)

1777 because this species lacks carotenoid-based ornamentation and so is not subject to the potential
1778 costs of allocating carotenoids as colourants.

1779 In addition, because the colour of feathers is determined only during moult when
1780 carotenoids are actively deposited in growing feathers, we extracted plumage colour effect
1781 sizes only from studies of moulting individuals; we calculated effect sizes from non moulting
1782 birds with plumage ornaments only for the relationship between carotenoid intake and plasma
1783 carotenoid concentration. The colour of a soft part, such as the bill, can change rapidly during
1784 any season, so we could extract both coloration and plasma carotenoid level effect sizes from
1785 studies of these ornaments, regardless of moult status. The means of assessing ornamental
1786 coloration is important to consider in our analysis because colour is generally quantified along
1787 one or more of three main axes: hue, or the shade of the colour (e.g., red, orange, yellow);
1788 chroma, or the intensity of the colour (also called saturation); and brightness, or the lightness/
1789 darkness of the colour. In addition, principal component analysis can be used to create a
1790 composite metric directly from the reflectance spectrum of a colour. Each of these axes of
1791 colour tends to relate to different properties of the coloured ornament itself. For example,
1792 chroma may be a good generalisation of pigment density, while hue may be more representative
1793 of the proportion of red to yellow pigments in a carotenoid-coloured ornament.”

1794

1795 Issues

1796 No issues detected.

1797

1798 *Mautz et al. (2013)*

1799 Originally reported

1800 “We were interested in four assays of ‘ejaculate quality’: sperm number, sperm swimming
1801 speed, sperm size, and sperm viability. We did not include measures of non-sperm

1802 characteristics of ejaculates such as seminal chemicals, even though these might affect female
1803 fertility and/or male competitiveness under sperm competition, because these relationships are
1804 far less clear than those for the four ejaculate traits we do consider. [...] Traits included in each
1805 subcategory include: (A) quantity: spermatocrit, sperm count/number, sperm density; (B) size:
1806 head length, midpiece length, flagellum length, total length, relative midpiece length, head
1807 area, midpiece area; (C) speed: average velocity, average path velocity, curvilinear velocity,
1808 linear velocity, straight-line velocity; (D) viability: absolute live sperm, longevity, per cent live,
1809 percent motile, percent normal, viability.

1810 [...] papers had to address our main study question: is there a relationship between the
1811 expression of a male sexually secondary characteristic (SSC) and an ejaculate characteristic?
1812 Authors of the focal papers defined a range of traits as SSCs, including horn size, plumage
1813 colouration, song rates, courtship rates, social status, condition and body size. As we were
1814 searching for studies related to sexual selection, we relied on the authors of the original paper
1815 to define a trait as a SSC (and, by extension, a trait that has been, or still is, under sexual
1816 selection). We were specifically interested in sexually dimorphic traits and only included traits
1817 that are, generally speaking, not expressed by females (e.g. male courtship colours or
1818 advertisement calls) or differ qualitatively between the sexes (e.g. major differences in horn
1819 shape). We excluded traits such as social dominance, body condition and, most importantly,
1820 body size that are not usually treated as SSCs. [...] We were interested in SSCs rather than
1821 attractiveness per se so we exclude studies that only report on composite measures of
1822 attractiveness (e.g. ‘mating latency’).”

1823

1824 Issues

1825 No issues detected.

1826

1827 *McLean et al. (2012)*

1828 Originally reported

1829 “We carried out an extensive literature search (completed in April, 2010) to find studies that
1830 documented the inter-sexual relationships among the three traits in anuran species: (1) the
1831 correlation between male call frequency and male body size (hereafter, the frequency–size
1832 relationship), (2) female preference for male call frequency (the preference–frequency
1833 relationship) and (3) female preference for male body size (the size–preference relationship).
1834 [...] For the relationships that involved female preferences, we included studies that quantified
1835 the relationship between male mating success and signal or quality traits, assuming differential
1836 male mating success was generally the result of female choice.”

1837

1838 Issues

1839 No issues detected.

1840

1841 *Meunier et al. (2011)*

1842 Originally reported

1843 “For each study, sign and magnitude of the correlation between melanin-based coloration and
1844 laying date, clutch size, brood size and survival were given by the parameter ‘effect size’ r
1845 calculated following standard methodology. We defined a positive effect size when individuals
1846 with larger or darker melanin-based colour traits had lower laying date, had larger clutch size,
1847 larger brood size and higher survival rate than individuals with smaller or paler melanin-based
1848 colour traits. [...] We restricted our meta-analyses to eumelanin-based coloration (i.e. black and
1849 grey coloration) because few studies have yet been published on pheomelanin-based coloration
1850 (i.e. reddish-brown coloration).”

1851

1852 Issues

1853 No issues detected.

1854

1855 *Moller & Alatalo (1999)*

1856 Originally reported

1857 “The aims of the present study were to quantify the viability effects of sexual selection. This
1858 was carried out based on a literature survey of studies of good-genes sexual selection. Although
1859 good-genes effects may be expressed as enhanced growth, fecundity or survival, we have
1860 concentrated our efforts on reviewing the literature on survivorship effects because most
1861 studies have addressed this major fitness component, and because life-time reproductive
1862 success in a diverse array of organisms depends more on longevity than on any other life-
1863 history trait. [...] The variables of interest were classified in the following ways: (i) whether the
1864 target of selection had been identified based on observations or experiments, or whether that
1865 was not the case; (ii) the magnitude of the viability effect, calculated as the correlation
1866 coefficient between a secondary sexual character and viability of the offspring; (iii) the female
1867 mate preference for the male trait estimated from observational or experimental studies,
1868 expressed as the correlation coefficient between the male character and male mating success
1869 [...].”

1870

1871 Issues

1872 1) The manuscript does not contain information on which sexual signals were considered valid
1873 for inclusion. We thus considered all sexual signals as valid for matching purposes.

1874

1875 *Moller & Jennions (2001)*

1876 Originally reported

1877 “We determined the relationship between the expression of male secondary sexual characters
1878 (or other characters associated with male mating success that appear to influence female choice
1879 decisions) and four components of direct fitness for females. [...] First, fertility was determined
1880 as either (1) the proportion of eggs fertilised among females mated to a given male; (2) the
1881 probability that copulation led to offspring production; or (3) the proportion of eggs that
1882 hatched in species without paternal care. Since multiple mating is common among females of
1883 many species and sperm storage may occur, we can be sure that this effect is larger than
1884 estimated here. Second, we determined fecundity as (1) clutch size; (2) the number of eggs laid
1885 over a specified time interval; or (3) litter size in live-bearing species. Third, we determined
1886 male investment in parental care in birds as (1) the proportion of feeding visits to offspring
1887 relative to the total number of feeds by both male and female, or (2) the absolute feeding rate
1888 of the male. Where possible we used the latter measure because differential allocation by
1889 females may lead to an underestimate of the absolute amount a male invests in offspring. Four,
1890 we also included measures of hatching success for fish, amphibians and insects where there is
1891 paternal care but males do not feed offspring. In all these species it is clear that male parental
1892 behaviour is a major factor in the hatching success (e.g. due to fanning of eggs or attacking
1893 predators). We excluded data on fledgling production or the number of young reared to
1894 independence by birds and mammals because we did not feel we could adequately cover the
1895 huge literature that this entails. However, we are unaware of any study relating the expression
1896 of secondary sexual characters to the quality of male parental care and subsequently to
1897 offspring viability (recruitment into the following reproducing population). In addition,
1898 variation in territory quality is likely to play a major role in offspring survival because it will
1899 influence the rate at which offspring are fed. Thus, we would need to partition reproductive
1900 success into effects related to the expression of male secondary sexual characters and effects
1901 due to territory quality per se. We are unaware of any studies that have done so.

1902 We treated eggs in the nest as a sexually selected character in fish. Exclusive male parental care
1903 in insects (and fish) appears to be a sexually rather than naturally selected trait, because females
1904 should use the number of clutches or eggs a male cares for as a reliable signal of paternal intent
1905 and quality.”

1906

1907 Issues

1908 1) The manuscript does not contain information on which sexual signals were considered valid
1909 for inclusion. We thus considered all sexual signals as valid for matching purposes.

1910 2) The manuscript mentions that "excluded data on fledgling production or the number of
1911 young reared to independence by birds and mammals" but we detected cases in the dataset with
1912 number of fledglings as a proxy in bird species (e.g. ref. ¹⁵⁸). Nonetheless, we followed the
1913 original rule given in the manuscript for matching purposes for other studies.

1914 3) Despite the criteria listed in the manuscript, occurrence of a second clutch as a reproductive
1915 success measurement was detected in a few cases of the dataset reported. Nonetheless, we
1916 followed the original rule given in the manuscript for matching purposes for other studies.

1917

1918 *Moore et al. (2016)*

1919 Originally reported

1920 “We only included those studies in which the following criteria were met: 1) subjects were
1921 adults; 2) subject sex was specified; 3) physiological indices of stress were measured [...] Four
1922 categories of stress measurement were reported: baseline GCs, peak or total GCs produced in
1923 response to a stressor, experimental elevation of GCs, and long-term stress. Baseline GCs were
1924 typically measured within 3–5 min of capture. Experimental elevation of GCs up to 4 times
1925 above baseline was achieved via subcutaneous implants containing GCs. Long-term stress was
1926 assessed in 3 ways: GCs deposited in feathers, faeces or hair; the ratio of heterophils to

1927 lymphocytes (a white blood cell count that correlates with baseline GCs); and the expression
1928 of heat shock proteins (highly conserved proteins that are elevated under stress). Both
1929 heterophil-to-lymphocyte ratio and heat shock proteins are widely used as proxies of recent
1930 and long-term stress in the ecological literature.

1931 The effect sizes we obtained considered a wide range of secondary sexual traits, which
1932 we sorted into 4 categories: coloration, vocalisation, morphological traits, and opposite-sex
1933 preferences. The coloration category included examples in birds, mammals, and reptiles. The
1934 amount of coloration was measured in several different ways, including brightness, hue,
1935 saturation, proportion of structure (e.g., eye ring) that is pigmented, ultraviolet reflectance, and
1936 colour reflectance. The vocalisation category included singing in birds and calling in
1937 amphibians and a mammal species (rock hyrax *Procavia capensis*). The parameters measured
1938 varied according to the nature of vocalisation in each species and included song rate,
1939 complexity, and repertoire size in birds; the latency to call, call duration, call rate, and vocal
1940 effort in amphibians, and whether calling/singing was observed or not (rock hyrax,
1941 amphibians). Effect sizes included in the morphological trait category all considered bird
1942 species and assessed the size of secondary sexual characters, such as comb or tail length. We
1943 also included in this category effect sizes considering the size of a coloured structure (but not
1944 the coloration itself), such as epaulet size in the red-winged black- bird (*Agelaius phoeniceus*).
1945 Finally, while not a secondary sexual trait per se, opposite-sex preference was included as an
1946 indirect measure of the level of sexual signalling, with the assumption being that attractiveness
1947 to the opposite sex is a function of investment in secondary sexual traits. We rely on author
1948 judgments regarding whether each trait is a secondary sexual trait or not.”

1949

1950 Issues

1951 No issues detected.

1952

1953 *Nakagawa et al. (2007)*

1954 Originally reported

1955 “We conducted an extensive search of the literature on the relationship between life-history
1956 traits and bib size in house sparrows. [...] Our criteria resulted in 6 life-history traits that could
1957 be investigated: fighting ability, parental ability, age, body condition (i.e., standardised weight),
1958 cuckoldry (the rate or absence/presence of lost paternity due to extra pair paternity [EPP]), and
1959 reproductive success. In the analysis of parental behaviour, we used studies investigating both
1960 incubation and feeding behaviour. A recent study showed incubation time to be a significant
1961 predictor of subsequent nestling provisioning in a population, so that treating incubation and
1962 provisioning together may be justifiable.”

1963

1964 Issues

1965 1) The manuscript mentions that reproductive success is a proxy without any details but their
1966 dataset only contains the number of fledglings (even though other measures were available,
1967 e.g. number of eggs, number of hatchlings). Thus, this was considered a hidden criterion for
1968 matching purposes.

1969

1970 *Nolazco et al. (2022)*

1971 Originally reported

1972 “Condition parameters fell into six categories: (1) body condition: mainly measurements of
1973 body mass adjusted by structural body size and others associated with the physical condition
1974 of individuals; (2) body size: structural size (measurements of tarsus, wing, beak, keel, and tail
1975 alone or in combination) and mass; (3) immunity: indicators of constitutive immunity, immune
1976 challenges and responses; (4) stress: indicators of baseline physiological stress, stress

1977 challenges, and capacity to cope with oxidative stress; (5) environment: climatic conditions
1978 and resources; and (6) parasites: incidence and abundance of parasites. [...] Fitness parameters
1979 included not only estimates of reproductive success, survival and offspring quality, but also
1980 parental quality (because parental investment has been hypothesised to vary as a function of
1981 ornamentation) and timing of breeding. Hence these factors were classified into five categories:
1982 (1) reproductive success: mating success and offspring production; (2) offspring quality or
1983 condition: measurements of egg quality, offspring body condition, immunity, parasites, and
1984 other indicators of physical condition; (3) parental quality: provisioning during incubation, and
1985 offspring feeding and defence; (4) timing of breeding: measured directly or as arrival time to
1986 breeding grounds; and (5) survival.

1987 We applied a phylogenetically controlled bivariate meta-analytic approach to quantify
1988 the strength and direction of associations between ornaments and condition or fitness, in bird
1989 species in which both sexes were ornamented. Only morphological ornaments that were
1990 visually recognisable and identified as such by the authors of the original studies were included,
1991 excluding traits that did not appear decorative such as body size and weapons. In all cases,
1992 ornaments were similar in structure and location between sexes. We note that in most cases
1993 ornamental function of these traits is assumed and experimental evidence is not available. Thus,
1994 we broadly defined ornaments as any phenotypic traits that look like decorations rather than
1995 having an apparent naturally selected function. [...] We excluded publications if only male traits
1996 were investigated (i.e., no conspecific female traits were studied in this or another publication).
1997 For any female-only studies, we used the cited reference list to identify the corresponding
1998 information on conspecific males, adding 25 additional publications. Our criteria for including
1999 these studies was that the data originated from the same populations. [...] We classified
2000 ornamental traits into six categories: (1) carotenoid-based colouration: yellow, orange, or red
2001 coloured ornaments; (2) melanin-based colouration: black, grey (eumelanin), or brown

2002 (pheomelanin); (3) structural-based colouration: iridescent and non-iridescent; (4)
2003 unpigmented: white patches; (5) morphological: morphology of ornamental appendages (e.g.,
2004 comb, wattle, tail, plumes), and (6) others: to cases in which the operational variables were a
2005 combination of two or more ornament categories or rare pigments (e.g., spheniscin in
2006 penguins).”

2007

2008 Issues

2009 No issues detected.

2010

2011 *Parker (2013)*

2012 Originally reported

2013 “I attempted to locate all papers published in English which assess the function or
2014 characteristics of plumage colour in blue tits. [...] No clear expectations about seasonal changes
2015 in colour emerge from sexual selection hypotheses, and so I did not assess tests for these
2016 patterns. [...] there is no clear prediction about the hue (wavelength of peak reflectance) of
2017 white plumage because it is characterised by relatively uniform reflectance across the spectrum
2018 visible to birds. Carotenoid-pigmented avian plumage requires dietary carotenoids regardless
2019 of any role in sexual selection and so I did not consider the effect of manipulation of dietary
2020 carotenoid intake on the colour of the carotenoid-pigmented yellow breast to be a test of a
2021 sexual selection hypothesis. I also excluded the one observational study that used a qualitative
2022 colour score rather than a quantitative colour measurement. [...] Age: Differences in plumage
2023 colour between second-year (yearling) and after-second-year individuals; Aggression directed:
2024 Aggression predicted by the plumage colour of the aggressor; Aggression received: Aggression
2025 predicted by the plumage colour of the victim, including the relationship between plumage
2026 colour manipulation and resulting dominance rank; Aggression dominance: Relationship

2027 between plumage colour and dominance rank; Mate choice – assortative: Correlations between
2028 plumage colour of one member of a mated pair and plumage colour or other morphological
2029 trait of the other member of mated pair; Mate choice – differential investment: Parental
2030 plumage colour related to the mate’s investment in reproduction; Mate choice – EPP: Male
2031 plumage colour and rates of extra-pair paternity (EPP) determined by molecular paternity
2032 analysis; Mate choice – WPP: Male plumage colour and rates of within-pair paternity (WPP)
2033 determined by molecular paternity analysis; Mate choice – trials: Captive mate-choice trials;
2034 Quality: Plumage colour and measures or manipulations of individual quality or measures of
2035 individual’s offspring quality; Sex: Differences in plumage colour between males and females;
2036 Sex ratio: Male plumage colour and sex ratio of offspring.”

2037

2038 Issues

2039 No issues detected.

2040

2041 *Parker & Ligon (2003)*

2042 Originally reported

2043 “We were interested in testing two basic hypotheses: (a) female mating decisions correlate with
2044 male comb morphology, and (b) female mating decisions correlate with male feather
2045 morphology.”

2046

2047 Issues

2048 1) The dataset was unclear regarding sexual signals (e.g. whether colour was hue, chroma,
2049 brightness, etc).

2050

2051 *Parker et al. (2006)*

2052 Originally reported

2053 “We located each published study comparing blue tit song with potential indices of individual
2054 quality or condition.”

2055

2056 Issues

2057 No issues detected.

2058

2059 *Prokop et al. (2012)*

2060 Originally reported

2061 “We collected studies reporting on either (1) genetic correlations between male sexual traits
2062 and other fitness-related characters, estimated using quantitative genetic methods (animal
2063 model, full-sibling/half-sibling designs), (2) correlations between sire sexual traits and
2064 offspring fitness-related traits, or (3) heritability of male sexual characters. [...] We also
2065 included studies where measures of male mating success (e.g., comparing males that did and
2066 did not achieve copulations in mating trials) were used instead of specific sexual traits,
2067 provided that (1) the success could be attributed to female choice rather than/apart from male–
2068 male competition and (2) different individual females were used to determine mating success
2069 of any given male and to produce his progeny scored for fitness traits. [...] Included only when
2070 paternity had been experimentally controlled or genetically confirmed—or the frequency of
2071 extra-pair offspring was known to be <15% in the population studied (<20% if the sample size
2072 was at least 200), as such levels of extra-pair paternity should not bias the estimates of genetic
2073 parameters. [...]

2074 In (1) and (2), we included data from species with multiple sexually selected traits,
2075 where pairs of such traits were correlated with each other. In all cases, we only took into
2076 account male sexual traits known or supposed to be targets of female choice, therefore

2077 excluding data on traits used only for intrasexual competition for mates. We also included
2078 studies where measures of male mating success (e.g., comparing males that did and did not
2079 achieve copulations in mating trials) were used instead of specific sexual traits, provided that
2080 (1) the success could be attributed to female choice rather than/apart from male– male
2081 competition and (2) different individual females were used to determine mating success of any
2082 given male and to produce his progeny scored for fitness traits. [...] We classified sexually
2083 selected traits as display (purely ornamental) and competitive (size/dominance related); a third
2084 category, “other,” consisted of traits that could not be assigned to either of the two (such as
2085 nuptial gift, mating success that could not be specifically attributed to either display or
2086 dominance, or a trait increasing postcopulatory success).”

2087

2088 Issues

2089 No issues detected.

2090

2091 *Robinson & Creanza (2019)*

2092 Originally reported

2093 “We obtained field studies that examined the link between individual song elaboration (number
2094 of songs or syllables) and reproductive success (reproductive output or mating success). From
2095 the 48 field studies that remained, we compiled 134 relevant measurements. These commonly
2096 used measurements of reproductive success were categorised as follows: Number of females:
2097 how many social mates a male attracts, where males who attract more females are assumed to
2098 be more successful. Latency to pairing date or laying date: these two measures are traditionally
2099 used as a measure of reproductive success, because attractive males should pair first, and those
2100 who produce offspring sooner have a better chance of parenting a larger brood or more than
2101 one brood of offspring in a single breeding season. Chicks born earlier in the season also tend

2102 to be more viable. Extrapair paternity: this is often considered a metric of reproductive success
2103 because males that sire offspring in extra pair matings are assumed to be more attractive to
2104 females. However, it has been suggested to be an unreliable metric. Clutch size or number of
2105 offspring/recruits: these three measures are affected by both male and female genetic quality;
2106 however, it has been shown that females exposed to more elaborate songs can respond by
2107 producing larger clutches, so male song quality can also potentially affect this metric. The
2108 number of offspring or number of recruits (offspring that return to the parental territory) is
2109 related to the genetic fitness of males and females, but also to parental investment.
2110 In the studies that remained, individual song elaboration was measured by either song
2111 repertoire size (unique number of songs per individual) or syllable repertoire size (unique
2112 number of syllables per individual). We included studies that measured the association between
2113 reproductive success and either of these song elaboration metrics, because syllable repertoire
2114 size and song repertoire size are correlated between species and are likely also correlated within
2115 species.”

2116

2117 Issues

2118 1) We detected one data point mislabeled as number of recruits when it should be residual
2119 recruitment (i.e. controlled for number of fledglings, a measure of viability rather than
2120 reproductive success). Nonetheless, we followed the original rule given in the manuscript for
2121 matching purposes for other studies.

2122

2123 *Romano et al. (2017a)*

2124 Originally reported

2125 “In our data set we only included the statistical relationships where plumage ornaments were
2126 hypothesised to determine fitness-related traits, independently of how the test was designed.

2127 Thus, we included the results of analyses where an ornament was included as a predictor of a
2128 fitness trait, as well as analyses testing for a difference in ornament expression between groups
2129 of individuals with different fitness (e.g. tests comparing tail length of mated versus unmated
2130 individuals). Importantly, we did not include tests of condition-dependence of the expression
2131 of ornaments (e.g. tests of variation in ornament expression according to parasite load,
2132 physiological parameters or environmental conditions) nor tests of the potential costs of
2133 ornaments, because these were not the focus of the study. [...] Briefly, data on reproduction
2134 were analysed by comparing relationships between each plumage ornament and both male and
2135 female reproductive output by separately considering the following fitness proxies, which refer
2136 to different phases of the breeding cycle: (i) mating success, which comprised the probability
2137 of obtaining a social mate, and the time elapsed between arrival date to the breeding site and
2138 reproduction; (ii) mating date; (iii) success in paternity, as gauged by gaining extra-pair and
2139 within-pair offspring, as well as by successfully engaging in extra-pair copulations; (iv) laying
2140 date; (v) breeding success, including clutch size, brood size, and fledging success for any
2141 breeding attempt; and (vi) overall reproductive success, considering the number of broods, all
2142 eggs produced, and total number of offspring sired (but not when total number of eggs/offspring
2143 was weighted by the number of broods) during the entire breeding season. We note that data
2144 on clutch and brood size were pooled because in the barn swallow both the number of
2145 unhatched eggs and mortality rate during the nestling period are normally very low. In practice,
2146 the number of eggs laid and the number of nestlings fledged are highly correlated. Data on
2147 parental care were divided into two categories, corresponding to different phases of the
2148 breeding period: (i) incubation period, including the duration of incubation and the (absolute
2149 or relative) time spent by females or males in incubating eggs; and (ii) care provisioning of
2150 nestlings, including feeding rate, number of prey brought to the nest, and duration of the
2151 nestling period. Importantly, we distinguished between parental care provided by the mother

2152 and the social father of the nestlings. We then identified two categories of offspring quality: (i)
2153 offspring size, including skeletal size (e.g. tarsus length) and body mass measurements of
2154 nestlings; and (ii) offspring physiology, accounting for immune function and other
2155 physiological variables. Finally, we also recorded effect sizes concerning the associations
2156 between arrival date or survival/mortality and plumage ornaments. Importantly, data regarding
2157 laying date, incubation, breeding success, care provisioning and offspring quality were
2158 categorised according to the breeding attempt to which they referred, because barn swallows
2159 often lay more than one clutch per breeding season, and the intensity of sexual selection may
2160 vary during the breeding season. We thus considered separately the results concerning first or
2161 second broods (no data were available for subsequent broods, as very few females lay more
2162 than two clutches in a breeding season). Data on paternity may refer to the first brood only or
2163 to both broods pooled. Because the fitness proxies could be ordered chronologically, we could
2164 test whether the intensity of sexual selection varied among different phases of the breeding
2165 cycle. We note that ‘breeding success’, indicating clutch and brood size, was placed before
2166 ‘care provisioning’ because it mainly reflects parental decisions on initial clutch size rather
2167 than subsequent adjustments of brood size (see above). Analysis of variation in the effect size
2168 among different phases of the breeding cycle was limited to first broods because of the small
2169 number of effect sizes (and breeding stages) concerning second broods. In addition, because
2170 only high-quality breeding individuals (i.e. the more ornamented ones) usually lay a second
2171 clutch, during second broods smaller inter-individual variability in ornament expression
2172 compared to the first broods is expected. This difference in ornament variation between first
2173 and second broods may therefore reduce mean effect during the breeding season.

2174 In our analyses we considered the following sexually dimorphic characters: tail length,
2175 tail asymmetry, size of white spots on tail, ventral plumage colour, throat patch colour and
2176 size.”

2177

2178 Issues

2179 No issues detected.

2180

2181 *Romano et al. (2017b)*

2182 Originally reported

2183 “We included information about both annual survival (e.g. probability to survive from one
2184 breeding season to the next) and survival during the breeding season (e.g. survival from
2185 predators and severe weather). We emphasise that in our model species, both breeding dispersal
2186 and emigration from one breeding site to another are extremely rare. Considering that all the
2187 studies included in the analyses have been performed during the breeding season, the effects
2188 of dispersal and emigration on individual survival are therefore virtually absent.

2189 We considered the published studies where survival was tested in relation to the expression of
2190 six sexually dimorphic characters which have been suggested to be under sexual selection in at
2191 least one of the barn swallow subspecies: tail length, tail asymmetry, size of white spots on the
2192 tail, ventral plumage colour, throat plumage colour, and throat colour patch size.”

2193

2194 Issues

2195 No issues detected.

2196

2197 *Sanchez-Tojar et al. (2018)*

2198 Originally reported

2199 “We only included articles in which dominance was directly inferred from agonistic dyadic
2200 interactions over resources such as food, water, sand baths or roosting sites.

2201 When the presence of multiple estimates was due to the use of different methods to estimate
2202 bib size and dominance rank on the same data, we chose a single estimate per group of birds
2203 or study based on the order of preference [...]” [exposed/visible bib > hidden bib > total bib].

2204

2205 Issues

2206 No issues detected.

2207

2208 *Santos et al. (2011)*

2209 Originally reported

2210 “The study must have: (1) investigated the relationship between dominance and plumage traits
2211 in one sex, or statistically accounted for sex effects on the dominance interactions (such sex
2212 effects are well known, and must be accounted for so that the relationship between dominance
2213 and plumage is not confounded); (2) reported whether the plumage characteristic of interest
2214 was manipulated or not; (3) reported whether interactions observed to assess dominance were
2215 territorial or over food resources; and (4) fully reported test statistics and their associated
2216 degrees of freedom (or sample size) for both significant and nonsignificant results. As a result
2217 of applying our first inclusion criterion, our data consist only of the relationship between
2218 dominance among males and male plumage (i.e. no data were available on the relationship
2219 between dominance and plumage among females).”

2220

2221 Issues

2222 No issues detected.

2223

2224 *Simons & Verhulst (2011)*

2225 Originally reported

2226 “To test whether the correlation between attractiveness and bill color can be attributed to a
2227 correlation between bill color and song rate, our approach was to quantify the association
2228 association between bill color and song rate using meta-analysis on a different set of studies
2229 and compare the strength of this correlation with the correlation between the color of a male’s
2230 bill and his attractiveness.

2231 The statistical approach between studies differed, with some reporting the preference for the
2232 reddest male and others reporting the relationship between the difference in redness and the
2233 resulting female preference. The second approach includes both the effect of the difference
2234 between males in redness together with the overall preference for the reddest male. We
2235 recommend reporting both in future research to ease comparison between studies. For the
2236 purpose of this review, we included both approaches because the rejection of either approach
2237 would have resulted in a substantial loss of studies. We preferred the statistic of the preference
2238 for the reddest males if both approaches were available.”

2239

2240 Issues

2241 1) The manuscript specified measures of body condition as size, size-corrected mass, or growth
2242 rate, but the dataset also included others like subcutaneous fat score. Nonetheless, we followed
2243 the original rule given in the manuscript for matching purposes for other studies.

2244

2245 *Simons et al. (2012)*

2246 Originally reported

2247 “We summarised five phenotypic relationships: circulating carotenoid levels with trait redness,
2248 immune function and oxidative stress state; and trait redness with immune function and
2249 oxidative stress state. The relationships with trait redness represent signalling value, i.e. the
2250 information that can be obtained by a choosing individual regarding the physiological state of

2251 the signaler. The relationships with carotenoid levels represent the hypothesised mechanisms
2252 maintaining signal honesty. [...] Exclusion criteria: [...] iii) An immune challenge or oxidative
2253 stress challenge was given after which carotenoid levels or sexual coloration were assessed.
2254 Our focus here is whether carotenoid levels or carotenoid-dependent coloration predict
2255 oxidative stress parameters or immune response. The question of whether challenges reduce
2256 carotenoid levels or redness of sexual coloration is relevant, and this mechanism may in part
2257 or fully underlie between-individual variation in sexual coloration. However, the effects of
2258 experimentally induced immune or oxidative stress cannot be directly scaled to natural
2259 variation or direct manipulation of carotenoid levels and may involve different trade-offs and
2260 hence we excluded such studies. iv) When carotenoid supplementation was applied
2261 experimentally, but data on natural variation in circulating carotenoid levels or coloration were
2262 also available, we used the latter because this is the variation that a choosing potential mate is
2263 confronted with. [...] We included [...] whether the effect size was subject to experimental
2264 variation, caused by treatments other than carotenoid supplementation, which potentially
2265 increased variation in the traits of interest. To avoid such effects we selected pre-experimental
2266 (including carotenoid supplementation studies) values or results of analyses of the control
2267 group only, when possible. [...] In our analyses we considered the measures of the immune
2268 system of which we found four or more independent studies. These measures were as follows:
2269 PHA response, antibody production against experimentally induced antigens, parasite load, and
2270 white blood cell counts.”

2271

2272 Issues

2273 1) The manuscript reports that only circulating carotenoids in plasma were included as a
2274 measure of body condition and the dataset reports the same (e.g. "carotenoid level"). However,
2275 we noticed that studies with distinct carotenoid diets appear to have been equated as carotenoid

2276 level. Nonetheless, we followed the original rule given in the manuscript for matching
2277 purposes for other studies.

2278 2) The manuscript focuses on redness as a sexual signal and mentions chroma and hue as valid
2279 measurements. However, we noticed that the extent of a colourful trait was also used in some
2280 occasions and that traits could be of other colours than red (e.g. yellow). Nonetheless, we
2281 followed the original rule given in the manuscript for matching purposes for other studies.

2282

2283 *Soma & Garamszegi (2011)*

2284 Originally reported

2285 “We used the following criteria for inclusion of studies: 1) results on the relationships between
2286 song complexity and reproductive success were based on observational data in wild birds
2287 without experimental manipulations; 2) song complexity was measured by either song or
2288 syllable repertoire size (i.e., the number of unique song or syllable types in the full repertoire
2289 or in a given amount of vocal sample), versatility, or principal component score of multiple
2290 song variables that reflect song complexity; 3) the measure of reproductive success was based
2291 on either the number of females mated (including both social and extrapair mating), the number
2292 of offspring sired (including the number of eggs, fledglings, and recruits), or the timing of the
2293 onset of reproduction.”

2294

2295 Issues

2296 No issues detected.

2297

2298 *Street et al. (2016)*

2299 Originally reported

2300 “To fulfil the inclusion criteria, articles had to contain either (1) data on changes in swelling
2301 size across days relative to an independent estimate of the day of ovulation based on ovarian
2302 hormone levels (from blood, urine or faecal samples) or laparoscopy, or (2) data on swelling
2303 size and any measures of individual female quality (e.g. body condition or social rank). [...] Only those measures that were reported in at least four studies were analysed, i.e. age, social
2304 rank and body condition.
2305

2306 [...] data from species with small swellings (defined as swellings that involve the
2307 vulval and clitoral areas only) were excluded (i.e. *Hylobates lar*; *Gorilla gorilla*), as were data
2308 from species with subcaudal swellings (i.e. *Macaca fasciata*). The relationship between swelling
2309 coloration and ovulation was not investigated, as few studies with suitable data were available.”
2310

2311 Issues

2312 No issues detected.
2313

2314 *Thornhill & Moller (1998)*

2315 Originally reported

2316 “Because we investigated the relative effect size for symmetry versus character size for the
2317 same trait, the only studies included from Moller and Thornhill (1998) were those that had one
2318 or more estimates that allowed calculation of an effect size for the relationship between
2319 asymmetry and character size, and mating success or a mate preference, respectively.”
2320

2321 Issues

2322 No issues detected.
2323

2324 *Weaver et al. (2018)*

2325 Originally reported

2326 “We quantitatively synthesised published results on the relationships between individual
2327 quality and plumage coloration of passerines produced via dietary versus converted
2328 carotenoids. [...] studies must have investigated at least one of the following proxies of
2329 individual quality: (1) nutritional condition, (2) immune function or oxidative capacity, (3)
2330 parasite resistance, or (4) reproductive or parental quality.

2331 [...] we focused our study on a single trait, feather coloration, to avoid confounding
2332 biological factors such as blood flow, carotenoid esterification, and differences in requisite
2333 enzymes that are relevant to bare-part coloration, but not feather coloration. [...] Because we
2334 were interested in the signal content of carotenoid-based plumage, we focused on studies that
2335 quantified feather colour using standardised colour metrics of natural (i.e., unsupplemented)
2336 adult bird colour levels. Therefore, we excluded studies from our meta-analysis for any of the
2337 following reasons: only coloration of non-feathered structures was measured (e.g., wattles,
2338 legs, beaks); a non-passerine species was studied; only plasma concentrations of carotenoids
2339 were measured; or only nestling or juvenile coloration was studied. We did not include
2340 measures of feather brightness as it is sensitive to factors unrelated to pigmentation. [...]
2341 Common metrics used to quantify this variation in feather reflectance include comparisons to
2342 standard colour charts (e.g., Munsel), calculations of hue, chroma and brightness or composite
2343 metrics such as principal components (PCA) from spectrophotometer data or digital
2344 photographs. Hue describes the unique spectral colour (e.g., “red”, “orange”, “yellow”) and
2345 chroma describes the saturation or spectral purity of the colour display relative to total
2346 reflectance across the visible range of the electromagnetic spectrum. [...] We did not include
2347 measures of brightness in our analyses because it is strongly influenced by the physical
2348 structure of the feather which may be altered by abrasion and wear and is difficult to interpret
2349 for carotenoid content.”

2350

2351 Issues

2352 1) The manuscript states that only colour metrics are valid sexual signals but dataset reveals
2353 that also included size of colourful plumage (e.g. patch size). Nonetheless, we followed the
2354 original rule given in the manuscript for matching purposes for other studies.

2355 2) The manuscript states that only adult sexual signals were used but the dataset includes data
2356 from juveniles from many empirical studies. Nonetheless, we followed the original rule given
2357 in the manuscript for matching purposes for other studies.

2358 3) The dataset includes offspring size but states it as "offspring quality" even though this is
2359 omitted in the manuscript. Nonetheless, we followed the original rule given in the manuscript
2360 for matching purposes for other studies.

2361

2362 *White (2020)*

2363 Originally reported

2364 "I included all experimental and observational studies that quantified the relationship between
2365 intersexual structural colour signal expression (via the measurement of hue, saturation or
2366 brightness, or a composite thereof) and any one of age, body condition (size, size-corrected
2367 mass or growth rate), immune function (oxidative damage, PHA response, circulating CORT
2368 or testosterone) or parasite resistance as a measure of individual quality."

2369

2370 Issues

2371 1) The manuscript states that only structurally-based sexual signals were included but some
2372 carotenoid-based ornaments were included in the dataset. Nonetheless, we followed the
2373 original rule given in the manuscript for matching purposes for other studies.

2374 2) Data on distinct carotenoid diets were apparently used as condition proxy, even though the
2375 manuscript lists only other proxies for "condition" ("size, size-corrected mass or growth rate").
2376 Nonetheless, we followed the original rule given in the manuscript for matching purposes for
2377 other studies.

2378 3) The manuscript mentions that the condition proxy could be "size, size-corrected mass or
2379 growth rate", but only body mass was used as a body size measurement, which we interpreted
2380 as a hidden criterion.

2381 4) Unclear how colours of sexual signals were assessed to be structurally-based. Description
2382 of sexual signals in the dataset only specify colour measurement (e.g. hue) but not body part.

2383

2384 *Yasukawa et al. (2010)*

2385 Originally reported

2386 “We performed a meta-analysis to determine whether epaulet expression affects male-male
2387 competition, female choice, or reproductive success. This analysis provides a quantitative way
2388 to interpret the wide range of results from a variety of observational and experimental studies
2389 of Red-winged Blackbird epaulets as sexually selected adaptations.”

2390

2391 Issues

2392 1) Specific included measurements for aggression are unclear in the manuscript and dataset
2393 given by authors contains only vague descriptions of aggression (number of hits, close
2394 approaches, approaches, dominance rank) but more were available in empirical studies from
2395 where data were extracted. We interpreted that only those reported were considered valid, and
2396 thus as a hidden criterion for matching purposes.

2397

2398 **Supplementary information S5. Examples of matching attempts**

2399 The easiest attempts to find the correct matches between datasets occurred in cases in which a
2400 given empirical study (source for data extraction) contained only one data point with the same
2401 description (i.e. same sexual signal and proxy) in the originally reported dataset as well as in
2402 the re-extracted dataset. For instance, ref. ⁴⁷ reported a data point from ref. ²⁶⁴ on the
2403 relationship between song rate and bill redness in zebra finches. In our re-extracted dataset,
2404 only one data point existed for this empirical study on this same relationship, so there was no
2405 doubt that it was a match to the data point originally reported. Unfortunately, most cases were
2406 not as simple. For example, ref. ⁴⁶ reported a single data point from ref. ¹²⁶ on the relationship
2407 between plumage colour and territory defence in Eurasian blue tits. Yet, our re-extracted dataset
2408 contained 38 data points for the relationship between plumage colour and aggression or social
2409 dominance. Luckily, only two of these mentioned territorial defence, with one of them with the
2410 same sample size as the one reported. If the proxy in the dataset of ref. ⁴⁶ was slightly more
2411 vague (aggression instead of territory defence) or if the sample size reported was different from
2412 the any of our re-extracted data points, then we would have to label all re-extracted data points
2413 with the same extraction ID, as if that single reported data point was an average of all possible
2414 data points given in the empirical study. Indeed, this is what happened for complicated cases
2415 like ref. ²²⁰, used by ref. ⁴⁷, who reported three data points, all on the relationship between bill
2416 redness and attractiveness in zebra finches. Our re-extracted dataset contained nine data points
2417 with similar description, and no other information was available to clarify this conundrum. We
2418 thus matched the three originally reported data points with all nine re-extracted data points. We
2419 classified matches one-to-one as exact matches, while other matches (one-to-multiple,
2420 multiple-to-one, multiple-to-multiple) as non-exact.

2421

2422 **Supplementary information S6. Calculation and conversion of effect sizes**

2423 We used Fisher's Zr in most of our analyses. Following ref. ³²⁷, we mainly used Pearsons'
2424 correlation coefficient (r) to calculate Zr as:

$$2425 \quad Zr = 0.5 \ln \left(\frac{1+r}{1-r} \right)$$

2426 with its sampling variance (se^2) as:

$$2427 \quad se^2 = \frac{1}{n-3}$$

2428 where n is the number of individuals tested (i.e. total sample size). If needed, we
2429 transformed Zr to r as:

$$2430 \quad r = \frac{e^{2Zr} - 1}{e^{2Zr} + 1}$$

2431 Data reported by empirical studies were available in various formats. If r itself could
2432 not be obtained, we calculated it from other data. For instance, if only the coefficient of
2433 determination was available (R^2), we calculated r as:

$$2434 \quad r = \sqrt{R^2}$$

2435 Other types of correlations were directly transformed to r following ref. ³, such as
2436 Spearman's rank correlation (ρ , only needed if $n < 90$, otherwise equates to r), as:

$$2437 \quad r = 2 \sin \left(\frac{\pi \rho}{6} \right)$$

2438 and Kendall's correlation (τ), as:

$$2439 \quad r = \sin \left(\frac{\pi \tau}{2} \right)$$

2440 Goodman and Kruskal's gamma (γ) is similar to Kendall's correlation, so we used the
2441 equation above with γ instead of τ to obtain r when necessary.

2442 By contrast, other types of data needed further steps to obtain r . For example, following
2443 ref. ³²⁸, t from independent samples t -tests were transformed to point-biserial correlations (r_{pb})
2444 as:

$$r_{pb} = \frac{t}{\sqrt{t^2 + n - 2}}$$

2445

2446 Similarly, following ref. ³, F -values were transformed to r_{pb} as:

$$r_{pb} = \frac{\sqrt{F}}{\sqrt{F + n - 2}}$$

2447

2448 Following ref. ³²⁸, we assume that r equals to the biserial correlation coefficient (r_b),

2449 which can be calculated from r_{pb} as:

$$r = r_b = \frac{\sqrt{pq}}{f(z_p)} r_{pb}$$

2450

2451 where $p = \frac{n_1}{n}$, $q = \frac{n_2}{n}$, n_1 is the sample size of the first group (e.g. treatment), n_2 is the
 2452 sample size of the second group (e.g. control), and $f(z_p)$ denotes the density of the standard
 2453 normal distribution at value z_p , which is the point for which $P(Z > z_p) = p$, with Z denoting a
 2454 random variable following a standard normal distribution. We assumed $p = q = 0.5$ in all
 2455 instances in which we used this equation.

2456 Other statistics had to be first transformed to t , such as (1) regression slopes (β) as:

$$t = \frac{\beta}{se},$$

2457

2458 (2) Chi-Squares (χ^2 ; but only those with one degree of freedom, i.e. comparisons
 2459 between two groups) as:

$$t = \sqrt{F} = \sqrt{\chi^2},$$

2460

2461 and (3) p -values using the qt function (t distribution) in R 4.4.0³¹⁷.

2462 Differences between groups were also reported in different ways, and most of them could not
 2463 be directly used to calculate r . Instead, we calculated Cohen's d from these data before
 2464 converting them to r . Following ref. ⁴, we obtained r from Cohen's d as:

$$r = \frac{d}{\sqrt{d^2 + a}}$$

2465

2466

2467 where a is:

$$2468 \quad a = \frac{(n_1 + n_2)^2}{n_1 n_2}$$

2469 When only total sample size was given, we assumed $n_2 = n_1 = n_2$. Following ref. ³²⁹,
2470 descriptive data were first used to calculate Cohen's d as:

$$2471 \quad d = \frac{m_2 - m_1}{S_{pooled}}$$

2472 where m_2 and m_1 are group means (e.g. treatment and control), while S_{pooled} was
2473 calculated as:

$$2474 \quad S_{pooled} = \sqrt{\frac{(n_2 - 1)s_2^2 + (n_1 - 1)s_1^2}{n_2 + n_1 - 2}}$$

2475 where s^2 is the variance of each group. However, if data from two points in time (e.g.
2476 pre- and post-experiment) were given for each of the two groups, following ref. ³³⁰, we
2477 calculated Cohen's d as:

$$2478 \quad d = \frac{(m_{2,post} - m_{2,pre}) - (m_{1,post} - m_{1,pre})}{S_{pooled*}}$$

2479 while $S_{pooled*}$ was calculated as:

$$2480 \quad S_{pooled*} = \sqrt{\frac{(n_{2,post} - 1)s_{2,post}^2 + (n_{2,pre} - 1)s_{2,pre}^2 + (n_{1,post} - 1)s_{1,post}^2 + (n_{1,pre} - 1)s_{1,pre}^2}{n_{2,post} + n_{2,pre} + n_{1,post} + n_{1,pre} - 4}}$$

2481 If only the difference between groups was given instead (as a single value), we
2482 calculated Cohen's d as:

$$2483 \quad d = \frac{\bar{m}}{S_{pooled}}$$

2484 where m is the difference between groups. If only results from paired-samples t -tests (t_{paired})
2485 were provided, following ref. ³³¹, we obtained Cohen's d from them as:

$$2486 \quad d = t_{paired} \sqrt{\frac{2(1-r_{12})}{n}}$$

2487 where r_{12} is the correlation coefficient between two groups (set as 0.5 in all instances). If only
2488 medians and interquartile range for each group were given (as in boxplots), we used equations
2489 14 and 15, along with table 2, from ref. ³³².

2490 Furthermore, data could be reported in the form of frequency of events or proportions.
2491 If the former, following ref. ³³³, we first calculated r as:

$$2492 \quad r = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

2493 where A, B, C, and D are frequencies from a two-by-two contingency table. If
2494 proportions for two groups were given and they could not be transformed into frequencies then,
2495 as in ref. ⁵⁸, we calculated Cohen's d as:

$$2496 \quad d = \frac{\text{logit}(p_2) - \text{logit}(p_1)}{\pi} \sqrt{3}$$

2497 where p_2 and p_1 are the proportions reported for each group. Lastly, if only a single proportion
2498 was given (e.g. proportion of time spent with one prospective mate over another), as in ref. ⁵⁸,
2499 we calculated Cohen's d as:

$$2500 \quad d = \frac{\text{logit}(p) - \text{logit}(1 - p)}{\pi} \sqrt{3}$$

2501 where p is the proportion reported.

2502 We could not compute effect sizes when the only data reported were (1) measures
2503 mentioned above without all necessary components to calculate effect sizes (e.g. means without
2504 standard errors), (2) genetic correlations, (3) Wald Chi-Square, (4) selection gradients, (5)
2505 Mann-Whitney U -Tests, and (6) F -values from ANOVAs with predictor variables with more
2506 than two levels.

2507

2508 **Supplementary information S7. Direction of effect sizes**

2509 We had to invert the direction of some effect sizes originally reported by two meta-analyses
2510 because authors reported these effect sizes in the opposite direction than the one they were

2511 analysed: (1) ref. ³⁰ with effect sizes on the relationship between frequency of acoustic signals
2512 and body size or mating success, and (2) ref. ⁴² with effect sizes related to latency to court.
2513 Moreover, we obtained the absolute value for all effect sizes from ref. ⁵³ in which the sexual
2514 signal involved colour hue, as the author stated doing so for their analysis.

2515 When re-extracting data from empirical studies, we followed the rationale that
2516 individual condition and fitness are reflected by greater sexual signal expression, except for the
2517 following sexual signals: asymmetry, inter-pop or intercall interval, latency to signal, song
2518 frequency in amphibians (but not in other animals), and other measures (e.g. composite
2519 measures, brightness, hue) in specific cases (e.g. lower hue for blue traits). Likewise, the
2520 direction rationale we used for various proxies (i.e. how they reflected individual condition or
2521 fitness) are shown in Table S1. The direction of data points re-extracted is also reported in detail
2522 in our dataset.

2523 Despite how we proceeded when re-extracting data, we adjusted the direction of re-
2524 extracted effect sizes to comply with the rationale stated in original meta-analyses. Therefore,
2525 we re-inverted re-extracted effect sizes for the following cases because they were in the
2526 opposite direction than the one we assigned to them during re-extraction:

- 2527 1. All effect sizes that could be included by ref. ¹⁴, as positive values there represent a
2528 positive relationship between sexual signal expression and parasite load.
- 2529 2. All effect sizes that could be included by ref. ³⁴, as positive values there represent a
2530 positive relationship between asymmetry of sexual signals and attractiveness.
- 2531 3. All effect sizes that could be included by ref. ³⁵, as positive values there represent a
2532 positive relationship between sexual signal expression and stress.
- 2533 4. Effect sizes related to age that could be included by ref. ¹⁵, as positive values there
2534 represent a negative relationship between sexual signal expression and age.

- 2535 5. Effect sizes related to parasite load, heterophil-to-lymphocyte ratio, and oxidative
2536 damage that could be included by ref. ⁴⁸, as positive values there represent a positive
2537 relationship between sexual signal expression and both parasite load and stress.
- 2538 6. Effect sizes related to asymmetry of sexual signals that could be included by ref. ⁵¹, as
2539 positive values there represent a positive relationship between asymmetry of sexual
2540 signals and attractiveness.
- 2541 7. Effect sizes related to parental care that could be included by ref. ⁴³, as positive values
2542 there represent a negative relationship between sexual signals and parental care.
- 2543 8. Effect sizes related to aggression received that could be included by ref. ³⁹, as positive
2544 values there represent a positive relationship between sexual signals and aggression
2545 received.

2546

2547 The direction of some re-extracted effect sizes was unclear because the authors of the
2548 empirical study did not clarify the expected direction for a given sexual signal, proxy, or
2549 estimate they reported (e.g. unclear whether colour PC1 positively or negatively related to
2550 signal conspicuity, extravagance, or attractiveness in ref. ¹⁸³). We excluded re-extracted effect
2551 sizes with an unclear proxy direction that were not matched to any reported effect sizes
2552 (unusual proxies, e.g. resting metabolic rate). For all other cases with unclear direction,
2553 whenever the re-extracted effect size was similar to their matched reported effect size(s) in
2554 magnitude (absolute difference less than 0.02) but in opposite directions, we changed the
2555 direction of that re-extracted effect size to be the same as the reported one(s) which it matched
2556 to minimise differences when information was ambiguous.

2557 Regarding data on the relationship between condition-dependency (see Table S1 for
2558 specific proxies) and sexual signals, we expected that individuals that express sexual signals
2559 do so because they can pay the costs of developing and maintaining that trait, indicating their

2560 good condition. However, this rationale is only valid for correlational studies. When studies
2561 manipulated sexual signals of individuals, we expected that individuals whose sexual signal
2562 was enlarged or made more intense would experience increased costs and thus ultimately lower
2563 survival.