Meta-analysis challenges a textbook example of status signalling: evidence for publication bias

[1] Alfredo Sánchez-Tójar[1,2]*, Shinichi Nakagawa[3], Moisès Sánchez-Fortún[1,4], Dominic A. Martin[2], Sukanya Ramani[1,5], Antje Girndt[1,2], Veronika Bókony[6], Bart Kempenaers[7], András Liker[8], David F. Westneat[9], Terry Burke[4], Julia Schroeder[1,2]

Authors' affiliations:

1. Evolutionary Biology Group, Max Planck Institute for Ornithology, Seewiesen, Germany

2. Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, United Kingdom

3. School of Biological, Earth and Environmental Sciences, University of New South Wales, Sidney, Australia

4. Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

5. Department of Animal Behaviour, University of Bielefeld, Germany.

6. Lendület Evolutionary Ecology Research Group, Plant Protection Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

7. Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany

8. MTA-PE Evolutionary Ecology Research Group, University of Pannonia, Veszprém, Hungary.

9. Department of Biology and Center for Ecology, Evolution, and Behaviour, University of Kentucky, Lexington, USA.

23   Short title: **Challenging a textbook example of status signalling**

24   *Corresponding author: Alfredo Sánchez-Tójar, Max Planck Institute for Ornithology,

25   Eberhard-Gwinner-Strasse, House 11, 82319, Seewiesen, Germany. Email address:

26   alfredo.tojar@gmail.com. Phone number: 00498157932445. Fax number:

27   00498157932344

28   Type of article: Meta-research article

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

Abstract

The status signalling hypothesis aims to explain conspecific variation in ornamentation by suggesting that some ornaments signal dominance status. Here, we use multilevel meta-analytic models to challenge the textbook example of this hypothesis, the black bib of house sparrows (*Passer domesticus*). We conducted a systematic review, and obtained raw data from published and unpublished studies to test whether dominance rank is positively associated with bib size across studies. Contrary to previous studies, our meta-analysis did not support this prediction. Furthermore, we found several biases in the literature that further question the support available for the status signalling hypothesis. First, the overall effect size of unpublished studies was zero, compared to the medium effect size detected in published studies. Second, the effect sizes of published studies decreased over time, and recently published effects were, on average, no longer distinguishable from zero. We discuss several explanations including pleiotropic, population- and context-dependent effects. Our findings call for reconsidering this established textbook example in evolutionary and behavioural ecology, raise important concerns about the validity of the current scientific publishing culture, and should stimulate renewed interest in understanding within-species variation in ornamental traits.

Introduction

Plumage ornamentation is a very striking example of colour and pattern diversity in the animal kingdom, and has attracted a lot of research [1]. Most studies have focused on sexual selection as the key mechanism to explain this diversity in ornamentation [2,3]. The status signalling hypothesis explains conspecific variation in ornaments by suggesting that these traits signal individual dominance status or fighting ability [4]. Aggressive contests are costly in terms of energy use, and risk of injuries and predation [5–9]. These costs could be reduced if individuals can predict the outcome of such contests beforehand using so-called "badges of status" – individuals could base the decision of whether to avoid or engage in aggressive interactions based on the signal's message [4].

Patches of ornamentation have been suggested to function as badges of status in a wide range of taxa, including insects [10], reptiles [11] and birds [12]. The status signalling hypothesis was originally proposed to explain variation in the size of mountain sheep horns [13,14], but the hypothesis has become increasingly important in the study of variability in plumage ornamentation in birds [4,12]. Among the many bird species studied [15], the house sparrow (*Passer domesticus*) has become the classic textbook example of status signalling [2,12,16,17]. The house sparrow is a sexually dimorphic passerine, in which the main difference between the sexes is a prominent black patch on the male's throat and chest (hereafter "*bib*"). Many studies have suggested that bib size serves as a badge of status, but most studies are based on limited sample sizes, and have used inconsistent methodologies for measuring bib and dominance status [15,18].

89   Meta-analysis is a powerful tool to quantitatively test the overall (across-study) effect

90   size (i.e. the "*meta-analytic mean*") for a specific hypothesis. Meta-analyses are

91   therefore able to provide more robust conclusions than single studies, and are

92   increasingly used in evolutionary ecology [19–22]. Traditional meta-analyses

93   combine summary data across different studies, where design and methodology are

94   study-specific (e.g. effect sizes among studies are typically adjusted for different

95   fixed effects). These differences among studies are expected to increase

96   heterogeneity, and therefore, the uncertainty of the meta-analytic mean [23]. Meta-

97   analysis of primary or raw data is a specific type of meta-analysis where studies can

98   be analysed in a consistent manner [23]. This type of meta-analysis allows

99   methodology to be standardized so that comparable effect sizes can be obtained

100   across studies and is, therefore, considered the gold standard in disciplines such as

101   medicine [24]. Unfortunately, meta-analysis of primary data is still rarely used in

102   evolutionary ecology (but see [25–27]), perhaps due to the difficulty of obtaining the

103   primary data of previously published studies [28].

104   An important feature of any meta-analysis is to identify the existence of bias in the

105   literature [20,29]. For example, publication bias occurs whenever particular effect

106   sizes (e.g. larger ones) are more likely found in the literature than others (e.g.

107   smaller ones). This tends to be the case when statistical significance and/or direction

108   of effect sizes determines whether results were submitted or accepted for publication

109   [29]. Thus, publication bias can strongly affect the estimation of the meta-analytic

110   mean, and distort the interpretation of the hypothesis [30]. Several methods have

111   been developed to identify this and other biases [20,29]; however, such methods are

112   imperfect, dependent on the number of effect sizes available, and therefore should

113   be considered as types of sensitivity analysis [20,31].

5

114   Here we meta-analytically assessed the textbook example of the status signalling

115   hypothesis in the house sparrow. Specifically, we combined summary and primary

116   data from published and unpublished studies to test the prediction that dominance

117   rank is positively associated with bib size across studies. We found that the meta-

118   analytic mean was small, uncertain and overlapped zero. Hence, our results

119   challenge the status signalling function of the male house sparrow's bib. Also, we

120   identified several biases in the published literature that call for substantial changes in

121   scientific publication culture. Finally, we discuss potential biological explanations for

122   our results, and provide advice for future studies testing the status signalling

123   hypothesis.

Results

124

125   Overall, we obtained the primary data for seven of 13 (54%) published studies, and

126   we provided data for six additional unpublished studies (Table 1, S1 and S2

127   Appendix).

128

129   **Table 1**. Studies used in the meta-analyses and meta-regressions testing the across-study

130   relationship between dominance rank and bib size in male house sparrows.

Study ID	Reference	Population ID	Primary data?	Number of groups[1]	Total number of males[2]	Comments
1	Ritchison 1985	Kentucky (captivity)	No	3	35	
2	Møller 1987	Denmark (wild)	Yes	3	37	
3	Andersson & Åhlund 1991	Sweden (captivity)	No	10	20	Estimate originally reported as statistically non-significant.

6

4	Solberg & Ringsby 1997	Norway (captivity)	Yes	5	44	
5	Liker & Barta 2001	Hungary (captivity)	Yes	1	10	
6	Gonzalez *et al.* 2002	Spain (captivity)	No	8	41	
7	Hein *et al.* 2003	Kentucky (wild)	Yes	4	39	
8	Riters *et al.* 2004	Wisconsin (captivity)	No	4	20	
9	Lindström *et al.* 2005	New Jersey (captivity)	No	4	28	Author shared processed data, but group ID was unavailable, so data were not re-analysed.
10	Bókony *et al.* 2006	Hungary (captivity)	Yes	2	19	
11	Buchanan *et al.* 2010	Scotland (captivity)	No	14 / 5	56 / 20	Groups were tested twice. Post-breeding estimates originally reported as statistically non-significant.
12	Dolnik & Hoi 2010	Austria (captivity)	No	4 / 4	31 / 31	Groups were tested twice. Pre-infection estimates originally reported as statistically non-significant.
13	Rojas Mora *et al.* 2016	Switzerland (captivity)	Yes	14	56	
14	Lendvai et al.	Hungary (captivity)	Yes[3]	4	46	Unpublished data part of: [45,46]
15	Tóth et al.	Hungary (captivity)	Yes[3]	3	35	Unpublished data part of: [46,47]

16	Bókony et al.	Hungary (captivity)	Yes[3]	4	26	Unpublished data part of: [46,48]
17	Sánchez-Tójar et al.	Germany (captivity)	Yes[3]	4	95	Unpublished study conducted in 2014.
18	Sánchez-Tójar et al.	Lundy Island (wild)	Yes[3]	7	172	Unpublished study conducted from 2013-2016.
19	Westneat	Kentucky (captivity)	Yes[3]	10	40	Unpublished study conducted in 2005.

131 [1] for primary data = yes, groups of birds containing less than 4 individuals were not included (see

132 Materials and Methods).[2] Note: since most studies analysed more than one group of birds, the total

133 number of males is different from group size in most cases (see below). [3] Information for the

134 unpublished datasets is available in S2 Appendix.

135

Dominance hierarchies

137 Mean sampling effort was 36 interactions/individual (SD = 24), which highlights that,

138 overall, dominance hierarchies were inferred reliably across groups [49]. The mean

139 Elo-rating repeatability was 0.92 (SD = 0.07) and the mean triangle transitivity was

140 0.63 (SD = 0.28). Thus, the dominance hierarchies observed across groups of house

141 sparrows were of medium both steepness and transitivity.

Meta-analytic mean

143 Our meta-analyses revealed a small overall effect size with large 95% credible

144 intervals that overlapped zero (Table 2; Fig 1). Additionally, the overall heterogeneity

145 ($I^2_{overall}$) was moderate (53%; Table 2). Thus, our results suggested that generally,

146 bib size does not predict dominance status in male house sparrows.

147

148 **Table 2**. Results of the multilevel meta-analyses on the relationship between dominance rank and bib

149 size in male house sparrows. Additionally, the results of the Egger's regressions are shown.

150 Estimates are presented as standardized effect sizes using Fisher's transformation (*Zr*). Both meta 1

151 and meta 2 include published and unpublished estimates, with meta 2 including two non-reported

152 estimates assumed to be zero (see section "Meta-analyses").

Meta-analysis	*k*	Meta-analytic mean [95% CrI]	$I^2_{population\ ID}$ [95% CrI] (%)	$I^2_{study\ ID}$ [95% CrI] (%)	$I^2_{overall}$ [95% CrI] (%)	Egger's regression [95% CrI]
meta 1	85	0.23 [-0.01,0.45]	16 [0,48]	21 [0,51]	53 [33,73]	-0.13 [-0.59,0.27]
meta 2	87	0.20 [-0.01,0.40]	15 [0,46]	20 [0,49]	53 [34,74]	-0.12 [-0.55,0.28]

153 *k* = number of estimates; CrI = credible intervals; I^2 = heterogeneity.
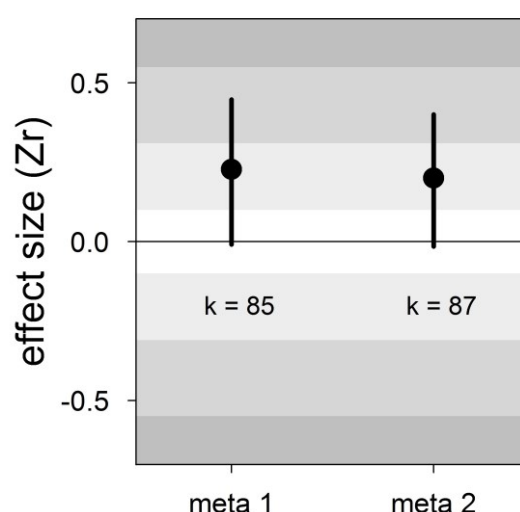

154

155 **Figure 1**. Forest plot showing the across-study effect size for the relationship between dominance

156 rank and bib size in male house sparrows. Both meta 1 and meta 2 include published and

157 unpublished estimates, with meta 2 including two non-reported estimates assumed to be zero (see

158 section "Meta-analyses"). We show posterior means and 95% credible intervals from multilevel meta-

159 analyses. Estimates are presented as standardized effect sizes using Fisher's transformation (*Zr*).

160 Light, medium and dark grey show small, medium and large effect sizes, respectively [50]. *k* is the

161 number of estimates.

9

162   **Moderators of the relationship between dominance rank and bib size**

163   None of the three biological moderators studied (season, group composition and

164   type of interactions) explained differences among studies (Table 3). Sampling effort

165   (i.e. the ratio of interactions to individuals recorded) was not an important moderator

166   either (Table 3).

167   **Table 3**. Results of the multilevel meta-regressions testing the effect of several moderators on the

168   relationship between dominance rank and bib size in male house sparrows. Estimates are presented

169   as standardized effect sizes using Fisher's transformation (Zr).

Meta-regression	Estimates	Mean [95% CrI]
meta 1	intercept	0.17 [-0.11,0.46]
(k = 85)	season	-0.11 [-0.41,0.21]
	group composition	0.14 [-0.34,0.59]
	type of interactions	0.33 [-0.17,0.91]
	$R^2_{marginal}$ =	23 [2,48]
meta 2	intercept	0.15 [-0.10,0.45]
(k = 87)	season	-0.08 [-0.42,0.22]
	group composition	0.12 [-0.32,0.62]
	type of interactions	0.27 [-0.17,0.85]
	$R^2_{marginal}$ =	20 [0,45]
sampling effort	intercept	0.24 [-0.15,0.55]
(k = 61)	sampling effort	0.11 [-0.49,0.74]
	sampling effort^2	-0.14 [-0.77,0.43]
	$R^2_{marginal}$ =	8 [0,24]

171   k = number of estimates; CrI = credible intervals; $R^2_{marginal}$ = percentage of variance explained by the

172   moderators. The factors season (non-breeding: 0, breeding: 1), group composition (mixed-sex: 0,

173   male-only: 1), and type of interactions (all: 0, aggressive-only: 1) were mean-centred, and the

174   covariates "sampling effort" and its squared term were z-transformed.

10

Detection of publication bias

There was no clear asymmetry in the funnel plots (Fig 2). Also, Egger's regression

tests did not show evidence of funnel plot asymmetry in any of the meta-analyses

(Table 2 and S4 Table). However, published effect sizes were larger than

unpublished ones, and the latter were not different from zero (Table 4; Fig 3).

Additionally, we found evidence for a time-lag bias in the published literature as

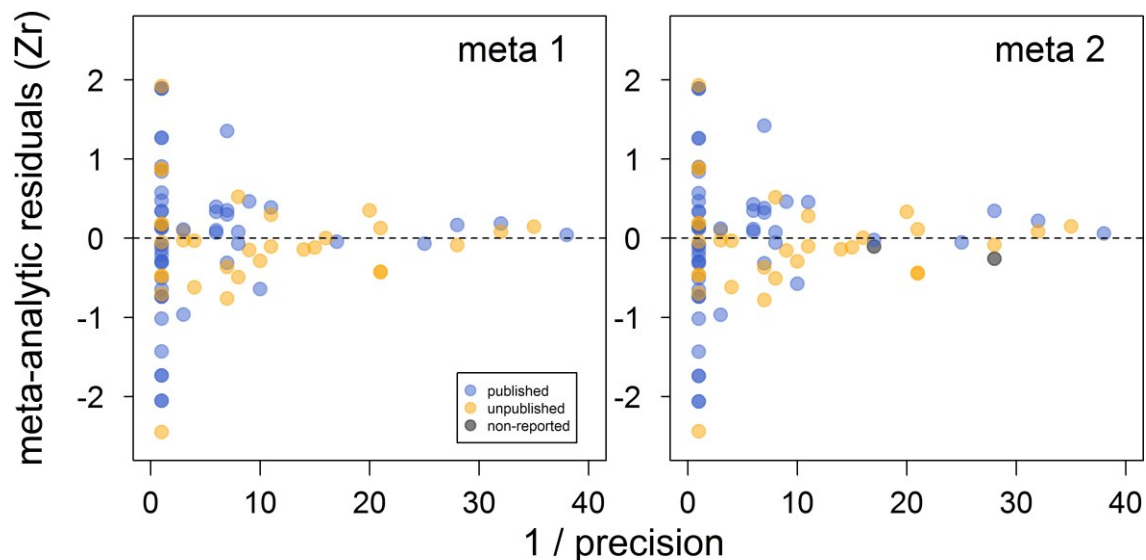effect sizes decreased over time (Table 4; Fig 4).

Figure 2. Funnel plots of the meta-analytic residuals against their precision for the meta-analyses

used to test the across-study relationship between dominance rank and bib size in male house

sparrows. Both meta 1 and meta 2 include published (blue) and unpublished (orange) estimates, with

meta 2 including two additional non-reported estimates (grey; see section "Meta-analyses").

Estimates are presented as standardized effect sizes using Fisher's transformation (Zr). Precision =

square root of the inverse of the variance.

11

192 **Table 4**. Results of the multilevel meta-regressions testing for time-lag and publication bias in the

193 literature on status signalling in male house sparrows. Estimates are presented as standardized effect

194 sizes using Fisher's transformation (Zr). Credible intervals not overlapping zero are highlighted in

195 bold.

Meta-regression	Estimates	Mean [95% CrI]
time-lag bias	intercept	**0.26 [0.03,0.57]**
(k = 53)	year of publication	**-0.21 [-0.41,-0.01]**
	$R^2_{marginal}$ =	29 [0,66]
published *vs.*	intercept	-0.09 [-0.37,0.18]
unpublished (k = 85)	published[a]	**0.50 [0.19,0.81]**
	$R^2_{marginal}$ =	38 [0,68]

196 k = number of estimates; CrI = credible intervals; $R^2_{marginal}$ = percentage of variance explained by the

197 moderators; [a] relative to unpublished. Year of publication was z-transformed.
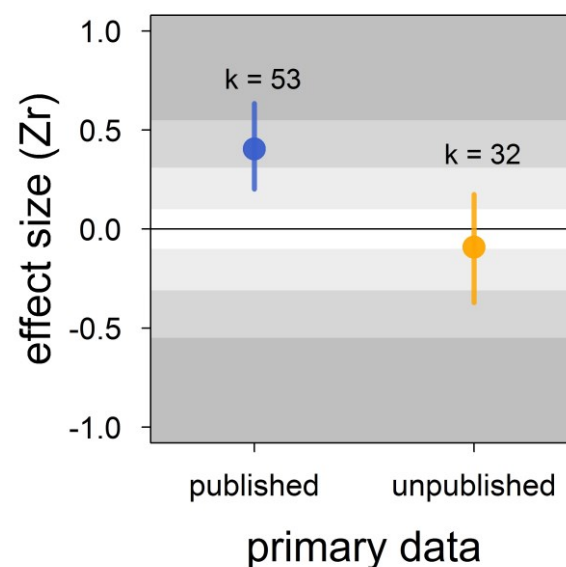

199 **Figure 3**. Published effect sizes for the status signalling hypothesis in male house sparrows are larger

200 than unpublished ones. We show posterior means and 95% credible intervals from a multilevel meta-

201 regression. Estimates are presented as standardized effect sizes using Fisher's transformation (Zr).

202 Light, medium and dark grey show small, medium and large effects sizes, respectively [50]. k is the
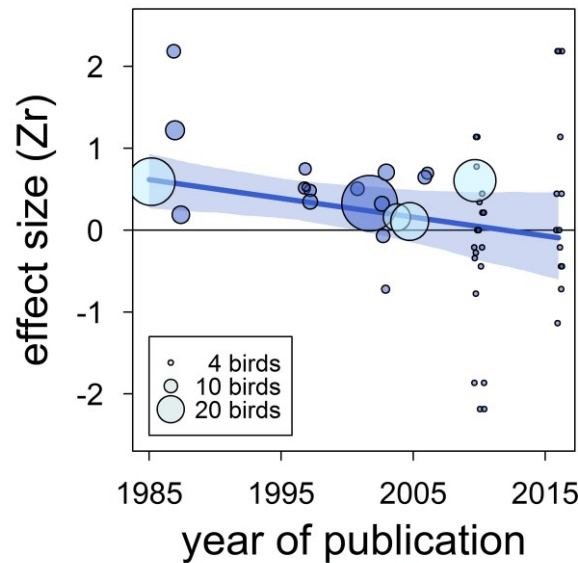
203 number of estimates.

12

204

Figure 4. The overall published effect size for the status signalling hypothesis in male house sparrows has decreased over time since first described (k = 53 estimates from 12 publications). The solid blue line represents the model estimate, and the shading shows the 95% credible intervals of a multilevel meta-regression based on published studies (see section "Detection of publication bias"). Estimates are presented as standardized effect sizes using Fisher's transformation (Zr). Circle area represents the size of the group of birds tested to obtain each estimate, where light blue denotes estimates for which group size is inflated due to birds from different groups being pooled, as opposed to dark blue where group size is accurate.

Discussion

The male house sparrow's bib is not the strong across-study predictor of dominance status once believed. In contrast to the medium-to-large effect found in the previous meta-analysis [18], our updated meta-analytic mean was uncertain and overlapped zero. Thus, the male house sparrows' bib should not be unambiguously considered or called a badge of status. Furthermore, we found evidence for the existence of bias in the published literature that further undermines the validity of the available support

13

221 for the status signalling hypothesis. First, the meta-analytic mean of unpublished

222 studies was zero, compared to the medium effect size detected in published studies.

223 Second, we found evidence for a time-lag bias. The effect size estimated in

224 published studies has been decreasing over time, and recently published effects

225 were on average no longer distinguishable from zero. Our findings call for

226 reconsidering this textbook example in evolutionary and behavioural ecology, raise

227 important concerns about the validity of the current scientific publishing culture, and

228 should stimulate renewed attention to hypotheses explaining within-species variation

229 in ornamentation.

230 The status signalling hypothesis [4] has been extensively tested to try and explain

231 intraspecific trait variation (e.g. reptiles: [11]; insects: [10]; humans: [51]), particularly

232 plumage variation [15]. Soon after the first empirical tests on birds, the black bib of

233 house sparrows became a textbook example of the status signalling hypothesis

234 [2,12,16,17], an idea that was later confirmed meta-analytically [18]. However,

235 Nakagawa *et al.*'s [18] meta-analytic mean was over-estimated because only 9 low-

236 powered studies were available (more in [52]). Here we updated that meta-analysis

237 with newly published and unpublished data. Our results showed that the overall

238 effect size is much smaller and much more uncertain than previously thought. The

239 status signalling hypothesis is thus no longer a compelling explanation for the

240 evolution of bib size across populations of house sparrows.

241 Similar contradicting conclusions have been reported for other model species. An

242 exhaustive review and meta-analysis on plumage coloration of blue tits (*Cyanistes*

243 *caeruleus*) revealed that, after dozens of publications studying the function of

244 plumage ornamentation in this species, the only robust conclusion is that females'

245 plumage differs from that of males [53]. Another example is the long-believed effect

14

246   of leg bands of particular colours on the perceived attractiveness of male zebra

247   finches (*Taeniopygia guttata*), which has been also experimentally and meta-

248   analytically refuted [54,55]. Finally, the existence of a badge of status in a non-bird

249   model species, the paper wasp (*Polistes dominulus*; [10]) has also been challenged

250   multiple times (e.g. [56–58]), generating doubts about its generality. Our findings

251   corroborate studies showing that abundant replication is needed before any strong or

252   general conclusion can be drawn [59], and highlight the existence of important

253   impediments to scientific progress in ecology.

254   Indeed, our results showed that the published literature on status signalling in house

255   sparrows is likely a biased subsample. The main evidence for this is that the mean

256   effect size of unpublished studies was essentially zero and clearly different from the

257   mean effect size based of published studies, which was of medium size.

258   Furthermore, this moderator (i.e. unpublished vs. published) explained a large

259   percentage of the model's variance. In some of our own unpublished datasets, the

260   relationship between dominance rank and bib size was never formally tested

261   (Westneat & Bókony, *personal communication*), suggesting that these unpublished

262   datasets are not strictly speaking examples of the "file drawer problem" (*sensu* [60]).

263   Egger's regressions failed to detect any funnel plot asymmetry, even in the meta-

264   analyses based on published effect sizes only (S4 Table). However, because

265   unpublished data indeed existed (i.e. those obtained for this study), the detection

266   failure was likely the consequence of the limited number of available estimates [61].

267   An additional common bias in the published literature is the time-lag bias [62]. We

268   detected evidence for such bias because the correlation between dominance rank

269   and bib size in published studies has decreased over time. Year of publication

270   explained a large percentage of variance, and accounting for year of publication

15

271 resulted in a strong reduction of the mean effect size across published studies (Table

272 4 *vs*. S4 Table). Time-lag bias has been detected in other ecological studies [63,64],

273 including a meta-analysis on status signalling across bird species [15]. In the latter

274 study, a positive overall (across-species) effect size persisted regardless of the time-

275 lag bias, and no strong evidence for other types of biases was found [15]. However,

276 Santos *et al*. [15] did not attempt to analyse unpublished data, so additional evidence

277 is needed to determine the effect that unpublished data have on the overall validity of

278 the status signalling hypothesis across bird species. If effect sizes based on

279 unpublished data for other species were of similar magnitude than those obtained for

280 house sparrows, the validity of the status signalling hypothesis across species would

281 need reconsideration. The existence of publication bias in ecology has long been

282 recognized [65–67]. Publication bias leads to false conclusions if not accounted for

283 [30], and is, thus, a serious impediment to scientific progress.

284 In addition to estimating the overall effect size for a hypothesis, meta-analyses can

285 also assess differences among estimates or heterogeneity [68,69]. Understanding

286 the sources of heterogeneity is an important step towards the correct interpretation

287 of a meta-analytic mean, and it can be done using meta-regressions [20]. Here, we

288 found that the percentage of variance that was not attributable to sampling error (i.e.

289 heterogeneity) was moderate. This value is below what has been quantified in

290 ecological and evolutionary meta-analyses [21], and indicates that we accounted for

291 large differences among estimates. Our meta-regressions based on biological

292 moderators explained 20-23% of the variance (Table 3). However, none of the

293 biological moderators that we tested influenced the overall effect size, but this might

294 be because of limited sample sizes.

16

295 The badge of status idea is more subtle than typically portrayed [70]. While the

296 evolution of badges of status in New and Old World sparrows has been related to

297 sociality during the non-breeding season [71], additional factors need to be involved

298 if the signal is to function in reducing aggression but retaining honesty. A recent

299 study on black-crested titmice (*Baeolophus atricristatus*) suggested that badges of

300 status might function only in high-competition scenarios [72]. However, time of the

301 year was not a strong predictor in our models, though most studies were conducted

302 in captivity, where food is normally provided *ad libitum*. Badges of status are

303 expected to function both within and between sexes [4,12]. Indeed, we found little

304 evidence that the status signalling function of bib size differed between male-only

305 and mixed-sex flocks. Interestingly, when competing for resources, possessing a

306 badge of status would be beneficial for both males and females. However, males but

307 not female house sparrows have a bib. This sexual dimorphism suggests that the

308 bib's function is likely more important when competing for resources other than

309 essential, *a priori* non-sex-specific, resources such as food, water, sand baths and

310 roosting sites. [73,74] reported that female house sparrows preferentially choose

311 males with large bibs (but see [75]), and bib size has been positively correlated with

312 sexual behaviour [76,77], which suggests that the bib may play a role in mate choice.

313 Furthermore, the original status signalling hypothesis posits that the main benefit of

314 using badges of status would be to avoid fights, which should be particularly

315 important when interacting with non-familiar individuals [4,12]. Although we did not

316 have data to test whether unfamiliarity between contestants is an important pre-

317 requisite for the status signalling hypothesis, we found no change in mean effect size

318 when only obviously aggressive interactions were studied. In practice, testing

319 whether the bib is important in mediating aggression between unfamiliar individuals

17

320 is difficult because the certainty of estimates of individual dominance increases over

321 time as more contests are recorded, but so does familiarity between individuals.

322 Our analyses have several potential limitations. First, although the number of studies

323 included in this meta-analysis is more than double that of the previous meta-analysis

324 [18], it is still limited. Also, it is likely (see above) that additional unpublished data are

325 stored in "file drawers" *sensu* [60]. Second, most tests included in this study were still

326 low-powered in terms of group size (median = 6 individuals/estimate, range = 4-41),

327 and the sample size is inflated because some of the published studies pooled

328 individuals from different groups (Fig 4). Third, although our results showed little

329 evidence of an effect of sampling effort on the overall effect size, the quality of the

330 data on dominance and bib size may still be a potential factor explaining differences

331 between studies. Additionally, data from studies not specifically designed to test the

332 status signalling hypothesis (such as some of the unpublished data) may be less

333 prone to confirmation bias [78], although other unknown factors might affect the

334 adequacy of such data for testing the hypothesis. Fourth, experimental effect sizes

335 will normally be larger, because effects of confounding factors can be reduced [67].

336 However, our systematic review only identified two studies where the status

337 signalling hypothesis was tested experimentally in house sparrows [37,79],

338 preventing us from estimating the meta-analytic mean for experimental studies.

339 Note, however, that the results of those experiments were inconclusive, and could be

340 partially explained by the phenomenon known as regression to the mean [78].

341 There are some additional explanations for the small and uncertain effect detected

342 by our meta-analyses. First, different populations might be under different selective

343 pressures regarding status signalling. Indeed, the population-specific heterogeneity

344 ($I^2_{\text{population ID}}$) estimated in our meta-analyses was 15-16%, suggesting that

18

345   population-dependent effects might exist. Second, although none of the moderators

346   had a strong influence on the overall effect size, the study-specific heterogeneity

347   estimated in our meta-analyses ($I^2_{study\ ID}$ = 20-21%) suggests that the uncertainty

348   observed could still be explained by the status signal being context-dependent.

349   Although context-dependence is often invoked to explain variation between studies,

350   there is little evidence for it. Last, most studies testing the status signalling

351   hypothesis in house sparrows are observational (Table 1), and the only two

352   experimental studies conducted so far were inconclusive [37,79]. Thus, it cannot be

353   ruled out that the weak correlation observed between dominance status and bib size

354   is driven by a third, unknown variable. In this respect, it has been proposed that the

355   association between melanin-based coloration (such as the bib; e.g. [80,81]) and

356   aggression is due to pleiotropic effects of the genes involved in regulating the

357   synthesis of melanin (reviewed by [82]). Furthermore, bib size has been shown to

358   correlate with testosterone, a hormone often involved in aggressive behaviour ([83];

359   but see [84]). Future studies should shift the focus towards understanding the

360   function of bib size in wild populations and increase considerably the number of birds

361   studied per group. The latter is essential because the statistical power of published

362   tests of the status signalling hypothesis in house sparrows is alarmingly low (power =

363   8.5%, S4 Appendix) and lower than the average in behavioural ecology [85].

364   In conclusion, our results challenge an established textbook example of the "badge

365   of status" hypothesis to explain variation in ornament size. In house sparrows, we

366   find no evidence that bib size consistently acts as a badge of status across studies

367   and populations, and thus, this can no longer be considered a textbook example of

368   the status signalling hypothesis. Furthermore, our analyses highlight the existence of

369   publication and time-lag biases in the published literature, further undermining the

370 validity of past conclusions. Bias against the publication of small ("non-significant")

371 effects hinders scientific progress. We thus join the call for a change in incentives

372 and scientific culture in ecology and evolution [78,86–88].

Materials and Methods

Systematic review

375 We used several approaches to maximize the identification of relevant studies. First,

376 we included all studies reported in a previous meta-analysis that tested the

377 relationship between dominance rank and bib size in house sparrows [18]. Second,

378 we conducted a keyword search on the Web of Science, PubMed and Scopus from

379 2006 to June 2017 to find studies published after [18], using the combination of

380 keywords ["bib/badge", "sparrow", "dominance/status/fighting"]. Third, we screened

381 all studies on house sparrows used in a meta-analysis that tested the relationship

382 between dominance and plumage ornamentation across species [15] to identify

383 additional studies that we may have missed in our keyword search. We screened

384 titles and abstracts of all articles and removed the irrelevant articles before

385 examining the full texts. We followed the preferred reporting items for systematic

386 reviews and meta-analyses (PRISMA: [89]; see S1 Appendix). We only included

387 articles in which dominance was directly inferred from agonistic dyadic interactions

388 over resources such as food, water, sand baths or roosting sites (S1 Table).

Summary data extraction

390 Some studies had more than one effect size estimate per group of birds studied.

391 When the presence of multiple estimates was due to the use of different statistical

392 analyses on the same data, we chose a single estimate based on the following order

393 of preference: (1) direct reports of effect size per group of birds studied (e.g.

20

394   correlation coefficient), (2) inferential statistics (e.g. t, F and χ^2 statistics) from

395   analyses where group ID was accounted for and no other fixed effects were

396   included, (3) direct reports of effect size where individuals from different groups

397   where pooled together, (4) inferential statistics from models including other fixed

398   effects. When the presence of multiple estimates was due to the use of different

399   methods to estimate bib size and dominance rank on the same data, we chose a

400   single estimate per group of birds or study based on the order of preference shown

401   in S2 Appendix.

Primary data acquisition

403   We requested primary data (i.e. agonistic dyadic interactions and bib size measures)

404   of all relevant studies identified by our systematic review. Additionally, we asked

405   authors to share, if available, any unpublished data that could be used to test the

406   relationship between dominance rank and bib size in house sparrows. We emailed

407   the corresponding author, but if no reply was received, we tried contacting all the

408   other authors listed. One study [33] provided all primary data in the original

409   publication and, therefore, its author was not contacted. Last, we included our own

410   unpublished data (S2 Appendix).

411   Most studies recorded data from more than one group of birds (Table 1). For each

412   primary dataset obtained, we inferred the dominance hierarchy of each group of

413   birds from the observed agonistic dyadic interactions (wins and losses) between

414   individuals using the randomized Elo-rating method, which estimates dominance

415   hierarchies more precisely than other methods [49]. We then used the provided

416   measures of individual bib size (e.g. area outlined from pictures) or, if possible,

417   calculated bib area from length and width measures following [33]. Subsequently, we

21

418 estimated the Spearman's rho rank correlation (ρ) between individual rank and bib

419 size for each group of birds. For one study [42], we received the already inferred

420 dominance hierarchies for each group of birds, which we then correlated with bib

421 size to obtain ρ.

Effect size coding

423 Regardless of their source (primary or summary data), we transformed all estimates

424 (e.g. ρ, F statistics, etc) into Pearson's correlation coefficients (r), and then into

425 standardized effect sizes using Fisher's transformation (Zr) for between-study

426 comparison. We used the equations from [90] and [91]. Since log(0) is undefined, r

427 values equal to 1.00 and -1.00 were transformed to 0.975 and -0.975, respectively,

428 before calculating Zr. Zr values of 0.100, 0.310 and 0.549 were considered small,

429 medium and large effect sizes, respectively (equivalent benchmarks from [50]).

430 When not reported directly, the number of individuals (n) was estimated from the

431 degrees of freedom. The variance in Zr was calculated as: $V_{Zr} = 1/(n\text{-}3)$. Estimates

432 (k) based on less than four individuals were discarded (k = 33 estimates discarded).

Meta-analyses

434 We ran two multilevel meta-analyses to test whether dominance rank and bib size

435 were positively correlated across studies. The first meta-analysis, "*meta 1*", included

436 published and unpublished (re-)analysed effect sizes (i.e. effect sizes estimated from

437 the studies we obtained primary data from), plus the remaining published effect sizes

438 obtained from summary data (i.e. effect sizes for which primary data were

439 unavailable).

440 Second, three studies reported "statistically non-significant" results without showing

441 either the magnitude or the direction of the estimates (Table 1). Receipt of primary

442   data allowed us to recover some but not all the originally non-reported estimates.

443   Two "non-significant" estimates were still missing. To test the robustness of the

444   results to those two cases of selective reporting, we ran an additional meta-analysis

445   (see [92] for a similar approach). This second meta-analysis, "*meta 2*", was like meta

446   1 but included the two non-significant non-reported estimates, which were assumed

447   to be zero. Note that non-significant estimates can be either negative or positive, and

448   thus, assuming that they were zero may have either underestimated or

449   overestimated them, something we cannot know from non-reported estimates. Meta-

450   analyses based on published studies only are shown in S3 Appendix.

451   We investigated inconsistency across studies by estimating the heterogeneity (I^2)

452   from our meta-analyses following [20]. I^2 values around 25, 50 and 75% are

453   considered as low, moderate and high levels of heterogeneity, respectively [69].

Meta-regressions

455   We tested if season, group composition and/or the type of interactions recorded had

456   an effect on the meta-analytic mean. For that, we ran two multilevel meta-

457   regressions that included the following moderators (hereafter "*biological*

458   *moderators*"): (1) "*season*", referring to whether the study was conducted during the

459   non-breeding (September-February) or the breeding season (March-August); (2)

460   "*group composition*", referring to whether birds were kept in male-only or in mixed-

461   sex groups; and, (3) "*type of interactions*", referring to whether the dyadic

462   interactions recorded were only aggressive (e.g. threats and pecks), or also included

463   interactions that were not obviously aggressive (e.g. displacements). Because only

464   three of 19 studies were conducted in the wild (k = 12 estimates; Table 1), we did not

465   include a moderator testing for captive *versus* wild environments. The three

466   biological moderators were mean-centred following [93] to aid interpretation.

467   The ratio of agonistic dyadic interactions recorded to the total number of interacting

468   individuals observed (hereafter "*sampling effort*") is a measure of sampling effort that

469   correlates positively and logarithmically with the ability to infer the latent dominance

470   hierarchy [49]. The higher this ratio, the more precisely the latent hierarchy can be

471   inferred [49]. For the subset of studies for which the primary data of the agonistic

472   dyadic interactions were available (12 out of 19 studies; Table 1), we ran a multilevel

473   meta-regression including sampling effort and its squared term as *z*-transformed

474   moderators [93]. The squared term was included because of the observed

475   logarithmic relationship between sampling effort and the method's performance [49].

476   This meta-regression tested whether sampling effort had an effect on the meta-

477   analytic mean.

478   For all meta-regressions, we estimated the percentage of variance explained by the

479   moderators ($R^2_{marginal}$) following [94].

480   **Random effects**

481   All meta-analyses and meta-regressions included the two random effects "*population*

482   *ID*" and "*study ID*". Population ID was related to the geographical location of the

483   population of birds studied. We used Google maps to estimate the distance over

484   land (i.e. avoiding large water bodies) between populations, and assumed the same

485   population ID when the distance was below 50 km (13 populations; Table 1). Study

486   ID encompassed those estimates obtained within each specific study (19 studies).

487   Two studies tested the prediction twice for the same groups of birds (Table 1) and,

488   within each population, some individuals may have been sampled more than once.

24

489   However, we could not include "group ID" and/or "individual ID" as additional random

490   effects due to either limited sample size or because the relevant data were not

491   available.

Detection of publication bias

493   For the meta-analyses, we assessed publication bias using two methods that are

494   based on the assumption that funnel plots should be symmetrical. First, we visually

495   inspected asymmetry in funnel plots of meta-analytic residuals against the inverse of

496   their precision (defined as the square root of the inverse of V_{zr}). Funnel plots based

497   on meta-analytic residuals (the sum of effect-size-level effects and sampling-

498   variance effects) are more appropriate than those based on effect sizes when

499   multilevel models are used [20]. Second, we ran Egger's regressions using the meta-

500   analytic residuals as the response variable, and the precision (see above) as the

501   moderator [20]. If the intercept of such a regression does not overlap zero, estimates

502   from the opposite direction to the meta-analytic mean might be missing and hence

503   we consider this evidence of publication bias [20]. Further, we tested whether

504   published estimates differed from unpublished estimates. For that, we ran a

505   multilevel meta-regression that included population ID and study ID as random

506   effects, and "*unpublished*" (two levels: yes (0), no (1)) as a moderator. This meta-

507   regression was based on meta 1 (i.e. it did not include the two non-reported

508   estimates). We did not use the trim-and-fill method [95,96] because this method has

509   been advised against when significant heterogeneity is present [29,97], as it was the

510   case in our meta-analyses (see below).

511   Finally, we analysed temporal trends in effect sizes that could indicate "time-lag

512   bias". Time-lag bias is common in the literature [63,64], and occurs when the effect

25

513   sizes of a specific hypothesis are negatively correlated with publication date (i.e.

514   effect sizes decrease over time; [62]). We ran a multilevel meta-regression based on

515   published effect sizes only, where "*year of publication*" was included as a *z*-

516   transformed moderator [20].

517   All analyses were run in R v. 3.4.0 [98]. We inferred individual dominance ranks from

518   agonistic dyadic interactions using the randomized Elo-rating method from the R

519   package "aniDom" v. 0.1.3 [49,99]. Additionally, we described the dominance

520   hierarchies observed in the groups of house sparrows for which primary data was

521   available. For that we estimated the uncertainty of the dominance hierarchies using

522   the R package "aniDom" v. 0.1.3 [49,99] and the triangle transitivity [100] using the R

523   package "compete" 3.1.0 [101]. We used the R package "MCMCglmm" v. 2.24 [102]

524   to run the multilevel meta-analytic (meta-regression) models [103]. For each meta-

525   analysis and meta-regression, we ran three independent MCMC chains for 2 million

526   iterations (thinning = 1800, burn-in = 200,000) using inverse-Gamma priors (V = 1,

527   nu = 0.002). Model chains were checked for convergence and mixing using the

528   Gelman-Rubin statistic. The auto-correlation within the chains was < 0.1 in all cases.

529   For each meta-analysis and meta-regression, we chose the model with the lowest

530   DIC value to extract the posterior mean and its 95% highest posterior density

531   intervals (hereafter 95% credible interval).

532

Acknowledgments

Data Accessibility

548 We provide all of the R code and data used for our analyses [104].

References

550 1. Hill GE. A red bird in a brown bag: the function and evolution of colorful

551   plumage in the House Finch. New York: Oxford University Press; 2002.

552 2. Andersson M. Sexual selection. New Jersey: Princeton University Press; 1994.

553 3. Dale J, Dey CJ, Delhey K, Kempenaers B, Valcu M. The effects of life history

554   and sexual selection on male and female plumage colouration. Nature.

555   2015;527: 367–370. doi:10.1038/nature15509

556 4. Rohwer S. The social significance of avian winter plumage variability.

557   Evolution. 1975;29: 593–610. doi:10.2307/2407071

558 5. Prenter J, Elwood RW, Taylor PW. Self-assessment by males during

559   energetically costly contests over precopula females in amphipods. Anim

560   Behav. 2006;72: 861–868. doi:10.1016/j.anbehav.2006.01.023

561   6.   Kelly CD, Godin J-GJ. Predation risk reduces male-male sexual competition in

562   the Trinidadian guppy (*Poecilia reticulata*). Behav Ecol Sociobiol. 2001;51: 95–

563   100. doi:10.1007/s002650100410

564   7.   Sneddon LU, Huntingford FA, Taylor AC. Impact of an ecological factor on the

565   costs of resource acquisition: fighting and metabolic physiology of crabs. Funct

566   Ecol. 1998;12: 808–815. doi:10.1046/j.1365-2435.1998.00249.x

567   8.   Neat FC, Taylor AC, Huntingford FA. Proximate costs of fighting in male cichlid

568   fish: the role of injuries and energy metabolism. Anim Behav. 1998;55: 875–

569   882. doi:10.1006/anbe.1997.0668

570   9.   Jakobsson S, Brick O, Kullberg C. Escalated fighting behaviour incurs

571   increased predation risk. Anim Behav. 1995;49: 235–239. doi:10.1016/0003-

572   3472(95)80172-3

573   10.   Tibbetts EA, Dale J. A socially enforced signal of quality in a paper wasp.

574   Nature. 2004;432: 218–222. doi:10.1038/nature02949

575   11.   Whiting MJ, Nagy KA, Bateman PW. Evolution and maintenance of social

576   status-signalling badges. In: Fox SF, McCoy K, Baird TA, editors. Lizard Social

577   Behavior. Baltimore: Johns Hopkins University Press; 2003. pp. 47–82.

578   12.   Senar JC. Color displays as intrasexual signals of aggression and dominance.

579   In: Hill GE, McGraw KJ, editors. Bird coloration: Function and evolution.

580   London: Harvard University Press; 2006. pp. 87–136.

581   13.   Beninde J. Naturgeschichte des Rothirshes. Monographie Wildsiiugetiere IV.

582   Leipzig; 1937.

14. Geist V. The evolutionary significance of mountain sheep horns. Evolution. 1966;20: 558–566. doi:10.2307/2406590

15. Santos ESA, Scheck D, Nakagawa S. Dominance and plumage traits: meta-analysis and metaregression analysis. Anim Behav. 2011;82: 3–19. doi:10.1016/j.anbehav.2011.03.022

16. Searcy WA, Nowicki S. The evolution of animal communication. Reliability and deception in signaling systems. New Jersey: Princeton University Press; 2005.

17. Davies NB, Krebs JR, West SA. An Introduction to Behavioural Ecology. 4th ed. Oxford: Wiley-Blackwell; 2012.

18. Nakagawa S, Ockendon N, Gillespie DOS, Hatchwell BJ, Burke T. Assessing the function of house sparrows' bib size using a flexible meta-analysis method. Behav Ecol. 2007;18: 831–840. doi:10.1093/beheco/arm050

19. Nakagawa S, Poulin R. Meta-analytic insights into evolutionary ecology: an introduction and synthesis. Evol Ecol. 2012;26: 1085–1099. doi:10.1007/s10682-012-9593-z

20. Nakagawa S, Santos ESA. Methodological issues and advances in biological meta-analysis. Evol Ecol. 2012;26: 1253–1274. doi:10.1007/s10682-012-9555-5

21. Senior AM, Grueber CE, Kamiya T, Lagisz M, O'Dwyer K, Santos ES, et al. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. Ecology. 2016;97: 3293–3299. doi:10.1002/ecy.1591

22. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature. 2018;555: 175–182.

606   doi:10.1038/nature25753

607   23.   Mengersen K, Gurevitch J, Schmid CH. Meta-analysis of primary data. In:

608   Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in

609   Ecology & Evolution. Princenton: Princeton University Press; 2013. pp. 300–

610   312.

611   24.   Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson

612   SG. Meta-analysis of individual patient data from randomized trials: a review of

613   methods used in practice. Clin Trials. 2005;2: 209–217.

614   doi:10.1191/1740774505cn087oa

615   25.   Barrowman NJ, Myers RA, Hilborn R, Kehler DG, Field CA. The variability

616   among populations of coho salmon in the maximum reproductive rate and

617   depensation. Ecol Appl. 2003;13: 784–793. doi:10.1890/1051-

618   0761(2003)013[0784:TVAPOC]2.0.CO;2

619   26.   Richards TA, Bass D. Molecular screening of free-living microbial eukaryotes:

620   diversity and distribution using a meta-analysis. Curr Opin Microbiol. 2005;8:

621   240–252. doi:10.1016/j.mib.2005.04.010

622   27.   Krasnov BR, Vinarski M V., Korallo-Vinarskaya NP, Mouillot D, Poulin R.

623   Inferring associations among parasitic gamasid mites from census data.

624   Oecologia. 2009;160: 175–185. doi:10.1007/s00442-009-1278-0

625   28.   Schmid CH, Landa M, Jafar TH, Giatras I, Karim T, Reddy M, et al.

626   Constructing a database of individual clinical trials for longitudinal analysis.

627   Control Clin Trials. 2003;24: 324–340. doi:10.1016/S0197-2456(02)00319-7

628   29.   Jennions MD, Lortie C, Rosenberg M, Rothstein H. Publication and related

629   biases. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-

630   analysis in Ecology & Evolution. Princenton: Princeton University Press; 2013.

631   pp. 207–236.

632 30. Rothstein H, Sutton A, Borenstein M. Publication bias in meta-analysis:

633   prevention, assessment and adjustments. Chichester: Wiley; 2005.

634 31. Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-

635   analysis: ten appraisal questions for biologists. BMC Biol. 2017;15: 18.

636   doi:10.1186/s12915-017-0357-7

637 32. Ritchison G. Plumage variability and social status in captive male house

638   sparrows. Kentucky Warbler. 1985;61: 39–42.

639 33. Møller A. Variation in badge size in male house sparrows *Passer domesticus*:

640   evidence for status signaling. Anim Behav. 1987;35: 1637–1644.

641   doi:10.1016/S0003-3472(87)80056-8

642 34. Andersson S, Åhlund M. Hunger affects dominance among strangers in house

643   sparrows. Anim Behav. 1991;41: 895–897. doi:10.1016/S0003-

644   3472(05)80356-2

645 35. Solberg EJ, Ringsby TH. Does male badge size signal status in small island

646   populations of house sparrows, *Passer domesticus*? Ethology. 1997;103: 177–

647   186. doi:10.1111/j.1439-0310.1997.tb00114.x

648 36. Liker A, Barta Z. Male badge size predicts dominance against females in

649   house sparrows. Condor. 2001;103: 151–157. doi:10.1650/0010-

650   5422(2001)103[0151:MBSPDA]2.0.CO;2

651 37. Gonzalez G, Sorci G, Smith LC, de Lope F. Social control and physiological

652 cose of cheating in status signalling male house sparrows (*Passer*

653 *domesticus*). Ethology. 2002;108: 289–302. doi:10.1046/j.1439-

654 0310.2002.00779.x

655 38. Hein WK, Westneat DF, Poston JP. Sex of opponent influences response to a

656 potential status signal in house sparrows. Anim Behav. 2003;65: 1211–1221.

657 doi:http://dx.doi.org/10.1006/anbe.2003.2132

658 39. Riters L V., Teague DP, Schroeder MB. Social status interacts with badge size

659 and neuroendocrine physiology to influence sexual behavior in male house

660 sparrows (*Passer domesticus*). Brain Behav Evol. 2004;63: 141–150.

661 doi:10.1159/000076240

662 40. Lindström KM, Hasselquist D, Wikelski M. House sparrows (*Passer*

663 *domesticus*) adjust their social status position to their physiological costs.

664 Horm Behav. 2005;48: 311–320. doi:10.1016/j.yhbeh.2005.04.002

665 41. Bókony V, Lendvai ÁZ, Liker A. Multiple cues in status signalling: the role of

666 wingbars in aggressive interactions of male house sparrows. Ethology.

667 2006;112: 947–954. doi:10.1111/j.1439-0310.2006.01246.x

668 42. Buchanan KL, Evans MR, Roberts ML, Rowe L, Goldsmith AR. Does

669 testosterone determine dominance in the house sparrow *Passer domesticus*?

670 An experimental test. J Avian Biol. 2010;41: 445–451. doi:10.1111/j.1600-

671 048X.2010.04929.x

672 43. Dolnik O V., Hoi H. Honest signalling, dominance hierarchies and body

673 condition in House Sparrows *Passer domesticus* (Aves: *Passeriformes*) during

674 acute coccidiosis. Biol J Linn Soc. 2010;99: 718–726. doi:10.1111/j.1095-

675 8312.2010.01370.x

676   44.   Rojas Mora A, Meniri M, Glauser G, Vallat A, Helfenstein F. Badge size

677   reflects sperm oxidative status within social groups in the house sparrow

678   *Passer domesticus*. Front Ecol Evol. 2016;4: 67. doi:10.3389/fevo.2016.00067

679   45.   Lendvai ÁZ, Barta Z, Liker A, Bókony V. The effect of energy reserves on

680   social foraging: hungry sparrows scrounge more. Proc R Soc B. 2004;271:

681   2467–2472. doi:10.1098/rspb.2004.2887

682   46.   Bókony V, Seress G, Nagy S, Lendvai ÁZ, Liker A. Multiple indices of body

683   condition reveal no negative effect of urbanization in adult house sparrows.

684   Landsc Urban Plan. 2012;104: 75–84. doi:10.1016/j.landurbplan.2011.10.006

685   47.   Tóth Z, Bókony V, Lendvai ÁZ, Szabó K, Pénzes Z, Liker A. Kinship and

686   aggression: do house sparrows spare their relatives? Behav Ecol Sociobiol.

687   2009;63: 1189–1196. doi:10.1007/s00265-009-0768-8

688   48.   Bókony V, Kulcsár A, Liker A. Does urbanization select for weak competitors in

689   house sparrows? Oikos. 2010;119: 437–444. doi:10.1111/j.1600-

690   0706.2009.17848.x

691   49.   Sánchez-Tójar A, Schroeder J, Farine DR. A practical guide for inferring

692   reliable dominance hierarchies and estimating their uncertainty. J Anim Ecol.

693   2017; doi:10.1111/1365-2656.12776

694   50.   Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New

695   Jersey: Taylor & Francis Inc; 1988.

696   51.   Dixson BJ, Vasey PL. Beards augment perceptions of men's age, social

697   status, and aggressiveness, but not attractiveness. Behav Ecol. 2012;23: 481–

698   490. doi:10.1093/beheco/arr214

699   52.   Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al.

700   Power failure: why small sample size undermines the reliability of

701   neuroscience. Nat Rev Neurosci. 2013;14: 365–376. doi:10.1038/nrn3475

702   53.   Parker TH. What do we really know about the signalling role of plumage colour

703   in blue tits? A case study of impediments to progress in evolutionary biology.

704   Biol Rev. 2013;88: 511–536. doi:10.1111/brv.12013

705   54.   Seguin A, Forstmeier W. No band color effects on male courtship rate or body

706   mass in the zebra finch: four experiments and a meta-analysis. PLoS One.

707   2012;7: e37785. doi:10.1371/journal.pone.0037785

708   55.   Wang D, Forstmeier W, Ihle M, Khadraoui M, Jerónimo S, Martin K, et al.

709   Irreproducible text-book knowledge: the effects of color bands on zebra finch

710   fitness. Evolution. 2018; doi:10.5061/dryad.cc145b6

711   56.   Cervo R, Dapporto L, Beani L, Strassmann J., Turillazzi S. On status badges

712   and quality signals in the paper wasp *Polistes dominulus*: body size, facial

713   colour patterns and hierarchical rank. Proc R Soc B. 2008;275: 1189–1196.

714   doi:10.1098/rspb.2007.1779

715   57.   Green JP, Field J. Interpopulation variation in status signalling in the paper

716   wasp *Polistes dominulus*. Anim Behav. 2011;81: 205–209.

717   doi:10.1016/j.anbehav.2010.10.002

718   58.   Green JP, Leadbeater E, Carruthers JM, Rosser NS, Lucas ER, Field J.

719   Clypeal patterning in the paper wasp *Polistes dominulus*: no evidence of

720   adaptive value in the wild. Behav Ecol. 2013;24: 623–633.

721   doi:10.1093/beheco/ars226

722   59.   Aarts AA, Anderson JE, Anderson CJ, Attridge PR, Attwood A, Axt J, et al.

723   Estimating the reproducibility of psychological science. Science. 2015;349:

724   aac4716. doi:10.1126/science.aac4716

725   60.   Rosenthal R. The file drawer problem and tolerance for null results. Psychol

726   Bull. 1979;86: 638–641. doi:10.1037/0033-2909.86.3.638

727   61.   Sterne J, Egger M. Regression methods to detect publication and other bias in

728   meta-analysis. In: Rothstein H, Sutton A, Borenstein M, editors. Publication

729   Bias in Meta-Analysis2. Chichester: John Wiley; 2005. pp. 99–110.

730   62.   Trikalinos T, Ioannidis JPA. Assessing the evolution of effect sizes over time.

731   In: Rothstein H, Sutton A, Borenstein M, editors. Publication Bias in Meta-

732   Analysis. Chichester: John Wiley; 2005. pp. 241–259.

733   63.   Poulin R. Manipulation of host behaviour by parasites: a weakening paradigm?

734   Proc R Soc B. 2000;267: 787–792. doi:10.1098/rspb.2000.1072

735   64.   Jennions MD, Møller AP. Relationships fade with time: a meta-analysis of

736   temporal trends in publication in ecology and evolution. Proc R Soc B.

737   2002;269: 43–48. doi:10.1098/rspb.2001.1832

738   65.   Cassey P, Ewen JG, Blackburn TM, Moller AP. A survey of publication bias

739   within evolutionary ecology. Proc R Soc B. 2004;271: S451–S454.

740   doi:10.1098/rsbl.2004.0218

741   66.   Jennions MD, Møller AP. Publication bias in ecology and evolution: an

742   empirical assessment using the "trim and fill" method. Biol Rev. 2002;77: 211–

743   222. doi:10.1017/S1464793101005875

744   67.   Palmer AR. Quasireplication and the contract of error: Lessons from Sex

745   Ratios, Heritabilities and Fluctuating Asymmetry. Annu Rev Ecol Syst.

746   2000;31: 441–480. doi:10.1146/annurev.ecolsys.31.1.441

747 68. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis.

748   Stat Med. 2002;21: 1539–1558. doi:10.1002/sim.1186

749 69. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in

750   meta-analyses. BMJ. 2003;327: 557–560. doi:10.1136/bmj.327.7414.557

751 70. Diep SK, Westneat DF. The integration of function and ontogeny in the

752   evolution of status signals. Behaviour. 2013;150: 1015–1044.

753   doi:10.1163/1568539X-00003066

754 71. Tibbetts EA, Safran RJ. Co-evolution of plumage characteristics and winter

755   sociality in New and Old World sparrows. J Evol Biol. 2009;22: 2376–2386.

756   doi:10.1111/j.1420-9101.2009.01861.x

757 72. Queller PS, Murphy TG. Seasonal variation in the utility of a status signaling

758   system: plumage ornament predicts foraging success only during periods of

759   high competition. PLoS One. 2017;12: e0185584.

760   doi:10.1371/journal.pone.0185584

761 73. Møller AP. Badge size in the house sparrow *Passer domesticus* - Effects of

762   intra- and intersexual selection. Behav Ecol Sociobiol. 1988;22: 373–378.

763   doi:10.1007/BF00295107

764 74. Møller AP. Natural and sexual selection on a plumage signal of status and on

765   morphology in house sparrows, *Passer domesticus*. J Evol Biol. 1989;2: 125–

766   140. doi:10.1046/j.1420-9101.1989.2020125.x

767 75. Kimball RT. Female choice for male morphological traits in house sparrows,

768   *Passer domesticus*. Ethology. 1996;102: 639–648. doi:10.1111/j.1439-

769   0310.1996.tb01155.x

770 76. Veiga JP. Mate replacement is costly to males in the multibrooded house

771   sparrow: an experimental study. Auk. 1996;113: 664–671.

772   doi:10.2307/4088987

773 77. Møller AP. Sexual behavior is related to badge size in the house sparrow

774   *Passer domesticus*. Behav Ecol Sociobiol. 1990;27: 23–29.

775   doi:10.1007/BF00183309

776 78. Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely

777   false-positive findings – a practical guide. Biol Rev. 2017;92: 1941–1968.

778   doi:10.1111/brv.12315

779 79. Diep S. The role of social interactions on the development and honesty of a

780   signal of status [Internet]. University of Kentucky. 2012. Available:

781   http://uknowledge.uky.edu/biology_etds/9/

782 80. Galván I, Wakamatsu K, Camarero PR, Mateo R, Alonso-Alvarez C. Low-

783   quality birds do not display high-quality signals: the cysteine-pheomelanin

784   mechanism of honesty. Evolution. 2015;69: 26–38. doi:10.1111/evo.12549

785 81. Galván I, Alonso-Alvarez C. Individual quality via sensitivity to cysteine

786   availability in a melanin-based honest signaling system. J Exp Biol. 2017;220:

787   2825–2833. doi:10.1242/jeb.160333

788 82. Ducrest AL, Keller L, Roulin A. Pleiotropy in the melanocortin system,

789   coloration and behavioural syndromes. Trends Ecol Evol. 2008;23: 502–510.

790   doi:10.1016/j.tree.2008.06.001

83. Gonzalez G, Sorci G, Smith LC, de Lope F. Testosterone and sexual signalling in male house sparrows (*Passer domesticus*). Behav Ecol Sociobiol. 2001;50: 557–562. doi:10.1007/s002650100399

84. Laucht S, Kempenaers B, Dale J. Bill color, not badge size, indicates testosterone-related information in house sparrows. Behav Ecol Sociobiol. 2010;64: 1461–1471. doi:10.1007/s00265-010-0961-9

85. Jennions M, Møller A. A survey of the statistical power of research in behavioral ecology and animal behavior. Behav Ecol. 2003;14: 438–445. doi:10.1093/beheco/14.3.438

86. Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, Chee YE, et al. Transparency in ecology and evolution: real problems, real solutions. Trends Ecol Evol. 2016;31: 711–719. doi:10.1016/j.tree.2016.07.002

87. Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. BMC Biol. 2015;13: 88. doi:10.1186/s12915-015-0196-3

88. Ihle M, Winney IS, Krystalli A, Croucher M. Striving for transparent and credible research: practical guidelines for behavioral ecologists. Behav Ecol. 2017;28: 348–354. doi:10.1093/beheco/arx003

89. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group T. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6: e1000097. doi:10.1371/journal.pmed.1000097

90. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev. 2007;82: 591–605.

814   doi:10.1111/j.1469-185X.2007.00027.x

815   91.   Lajeunesse MJ. Recovering Missing or Partial Data from Studies: A Survey of

816   Conversions and Imputations for Meta-analysis. In: Koricheva J, Gurevitch J,

817   Mengersen K, editors. Handbook of Meta-analysis in Ecology & Evolution.

818   Princenton: Princeton University Press; 2013. pp. 195–206.

819   92.   Booksmythe I, Mautz B, Davis J, Nakagawa S, Jennions MD. Facultative

820   adjustment of the offspring sex ratio and male attractiveness: a systematic

821   review and meta-analysis. Biol Rev. 2017;92: 108–134. doi:10.1111/brv.12220

822   93.   Schielzeth H. Simple means to improve the interpretability of regression

823   coefficients. Methods Ecol Evol. 2010;1: 103–113. doi:10.1111/j.2041-

824   210X.2010.00012.x

825   94.   Nakagawa S, Schielzeth H. A general and simple method for obtaining R2

826   from generalized linear mixed-effects models. Methods Ecol Evol. 2013;4:

827   133–142. doi:10.1111/j.2041-210x.2012.00261.x

828   95.   Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing

829   and adjusting for publication bias in meta-analysis. Biometrics. 2000;56: 455–

830   463. doi:10.1111/j.0006-341X.2000.00455.x

831   96.   Duval S, Tweedie R. A nonparametric "Trim and Fill" method of accounting for

832   publication bias in meta-analysis. J Am Stat Assoc. 2000;95: 89–98.

833   doi:10.1080/01621459.2000.10473905

834   97.   Moreno SG, Sutton AJ, Ades A, Stanley TD, Abrams KR, Peters JL, et al.

835   Assessment of regression-based methods to adjust for publication bias

836   through a comprehensive simulation study. BMC Med Res Methodol. 2009;9:

837   2. doi:10.1186/1471-2288-9-2

838   98.   Team RC. R: A language and environment for statistical computing [Internet].

839   Vienna: R Foundation for Statistical Computing; 2017. Available: http://www.r-

840   project.org/

841   99.   Farine DR, Sánchez-Tójar A. aniDom: Inferring Dominance Hierarchies and

842   Estimating Uncertainty [Internet]. 2017. Available: https://cran.r-

843   project.org/package=aniDom

844   100. McDonald DB, Shizuka D. Comparative transitive and temporal orderliness in

845   dominance networks. Behav Ecol. 2013;24: 511–520.

846   doi:10.1093/beheco/ars192

847   101. Curley JP. compete: Analyzing Social Hierarchies [Internet]. 2016. Available:

848   https://cran.r-project.org/web/packages/compete/index.html

849   102. Hadfield JD. MCMC methods for multi-respoinse generalized linear mixed

850   models: The MCMCglmm R package. J Stat Softw. 2010;33: 1–22. Available:

851   http://mirror.dcc.online.pt/CRAN/web/packages/MCMCglmm/vignettes/Overvie

852   w.pdf

853   103. Hadfield JD, Nakagawa S. General quantitative genetic methods for

854   comparative biology: phylogenies, taxonomies and multi-trait models for

855   continuous and categorical characters. J Evol Biol. 2010;23: 494–508.

856   doi:10.1111/j.1420-9101.2009.01915.x

857   104. Sánchez-Tójar A, Nakagawa S, Sánchez-Fortún M, Martin DA, Ramani S,

858   Girndt A, et al. Supporting information for "Meta-analysis challenges a textbook

859   example of status signalling: evidence for publication bias" [Internet]. 2018.

860   doi:10.17605/OSF.IO/CWKXB

861

Supporting Information

863   The following supporting information is available for this article online:

864   -   Supporting Information 1:

865   o   S1 Appendix: PRISMA flow diagram.

866   o   S2 Appendix: Information about data used.

867   o   S3 Appendix: Results based on published effect sizes only.

868   o   S4 Appendix: Power analysis.

869   -   Supporting Information 2: decision spreadsheet of the systematic review.