Meta-analysis challenges a textbook example of status signalling:

evidence for publication bias

Alfredo Sánchez-Tójar¹,²*, Shinichi Nakagawa³, Moisès Sánchez-Fortún¹,⁴, Dominic A. Martin², Sukanya Ramani¹,⁵, Antje Girndt¹,², Veronika Bókony⁶, Bart Kempenaers⁷, András Liker⁸, David F. Westneat⁹, Terry Burke⁴, Julia Schroeder¹,²

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.
Short title: Challenging a textbook example of status signalling

*Corresponding author: Alfredo Sánchez-Tójar, Max Planck Institute for Ornithology, Eberhard-Gwinner-Strasse, House 11, 82319, Seewiesen, Germany. Email address: alfredo.tojar@gmail.com. Phone number: 00498157932445. Fax number: 00498157932344

Type of article: Meta-research article
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 We conducted a systematic review, and obtained raw data from published and unpublished studies to test whether dominance rank is positively associated with bib size across studies.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 First, the overall effect size of unpublished studies was zero, compared to the medium effect size detected in published studies.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].
Meta-analysis is a powerful tool to quantitatively test the overall (across-study) effect size (i.e. the “meta-analytic mean”) for a specific hypothesis. Meta-analyses are therefore able to provide more robust conclusions than single studies, and are increasingly used in evolutionary ecology [19–22]. Traditional meta-analyses combine summary data across different studies, where design and methodology are study-specific (e.g. effect sizes among studies are typically adjusted for different fixed effects). These differences among studies are expected to increase heterogeneity, and therefore, the uncertainty of the meta-analytic mean [23]. Meta-analysis of primary or raw data is a specific type of meta-analysis where studies can be analysed in a consistent manner [23]. This type of meta-analysis allows methodology to be standardized so that comparable effect sizes can be obtained across studies and is, therefore, considered the gold standard in disciplines such as medicine [24]. Unfortunately, meta-analysis of primary data is still rarely used in evolutionary ecology (but see [25–27]), perhaps due to the difficulty of obtaining the primary data of previously published studies [28].

An important feature of any meta-analysis is to identify the existence of bias in the literature [20,29]. For example, publication bias occurs whenever particular effect sizes (e.g. larger ones) are more likely found in the literature than others (e.g. smaller ones). This tends to be the case when statistical significance and/or direction of effect sizes determines whether results were submitted or accepted for publication [29]. Thus, publication bias can strongly affect the estimation of the meta-analytic mean, and distort the interpretation of the hypothesis [30]. Several methods have been developed to identify this and other biases [20,29]; however, such methods are imperfect, dependent on the number of effect sizes available, and therefore should be considered as types of sensitivity analysis [20,31].
Here we meta-analytically assessed the textbook example of the status signalling hypothesis in the house sparrow. Specifically, we combined summary and primary data from published and unpublished studies to test the prediction that dominance rank is positively associated with bib size across studies. We found that the meta-analytic mean was small, uncertain and overlapped zero. Hence, our results challenge the status signalling function of the male house sparrow’s bib. Also, we identified several biases in the published literature that call for substantial changes in scientific publication culture. Finally, we discuss potential biological explanations for our results, and provide advice for future studies testing the status signalling hypothesis.

**Results**

Overall, we obtained the primary data for seven of 13 (54%) published studies, and we provided data for six additional unpublished studies (Table 1, S1 and S2 Appendix).

### Table 1. Studies used in the meta-analyses and meta-regressions testing the across-study relationship between dominance rank and bib size in male house sparrows.

| Study ID | Reference   | Population ID      | Primary data? | Number of groups | Total number of males | Comments                                      |
|----------|-------------|--------------------|---------------|------------------|-----------------------|-----------------------------------------------|
| 1        | Ritchison   | Kentucky (captivity)| No            | 3                | 35                    |                                               |
| 2        | Möller      | Denmark (wild)     | Yes           | 3                | 37                    | Estimate originally reported as statistically non-significant. |
| 3        | Andersson & Åhlund | Sweden (captivity) | No            | 10               | 20                    |                                               |
| Study                        | Location      | Captivity Status | Yes/No | Group Size |
|------------------------------|---------------|------------------|--------|------------|
| Solberg & Ringsby 1997       | Norway        | Yes              | 5      | 44         |
| Liker & Barta 2001          | Hungary       | Yes              | 1      | 10         |
| Gonzalez et al. 2002        | Spain         | No               | 8      | 41         |
| Hein et al. 2003            | Kentucky      | Yes              | 4      | 39         |
| Riters et al. 2004          | Wisconsin     | No               | 4      | 20         |
| Lindström et al. 2005       | New Jersey    | No               | 4      | 28         |
| Bókony et al. 2006          | Hungary       | Yes              | 2      | 19         |
| Buchanan et al. 2010        | Scotland      | No               | 14     | 56         |
| Dolnik & Hoi 2010           | Austria       | No               | 4      | 31         |
| Rojas Mora et al. 2016      | Switzerland   | Yes              | 14     | 56         |
| Lendvai et al.              | Hungary       | Yes³             | 4      | 46         |
| Tóth et al.                 | Hungary       | Yes³             | 3      | 35         |

Author shared processed data, but group ID was unavailable, so data were not re-analysed.

Groups were tested twice.

Post-breeding estimates originally reported as statistically non-significant.

Groups were tested twice.

Pre-infection estimates originally reported as statistically non-significant.

Unpublished data part of: [45,46]

Unpublished data part of: [46,47]
Dominance hierarchies

Mean sampling effort was 36 interactions/individual (SD = 24), which highlights that, overall, dominance hierarchies were inferred reliably across groups [49]. The mean Elo-rating repeatability was 0.92 (SD = 0.07) and the mean triangle transitivity was 0.63 (SD = 0.28). Thus, the dominance hierarchies observed across groups of house sparrows were of medium both steepness and transitivity.

Meta-analytic mean

Our meta-analyses revealed a small overall effect size with large 95% credible intervals that overlapped zero (Table 2; Fig 1). Additionally, the overall heterogeneity ($I^2_{overall}$) was moderate (53%; Table 2). Thus, our results suggested that generally, bib size does not predict dominance status in male house sparrows.
Table 2. Results of the multilevel meta-analyses on the relationship between dominance rank and bib size in male house sparrows. Additionally, the results of the Egger's regressions are shown. Estimates are presented as standardized effect sizes using Fisher's transformation ($Z_r$). Both meta 1 and meta 2 include published and unpublished estimates, with meta 2 including two non-reported estimates assumed to be zero (see section “Meta-analyses”).

| Meta-analysis | k | mean | $I^2$ population ID | $I^2$ study ID | $I^2$ overall | Egger's regression |
|---------------|---|------|---------------------|---------------|--------------|-------------------|
|               |   |      | [95% CrI]           | (%)           | [95% CrI]    | [95% CrI]         | [95% CrI]        |
| meta 1        | 85| 0.23 | [0.48]              | [33,73]       | [33,73]      | [-0.59,0.27]     |
|               |   |      | [-0.01,0.45]        | [33,73]       | [33,73]      | [-0.59,0.27]     |
| meta 2        | 87| 0.20 | [0.49]              | [34,74]       | [34,74]      | [-0.55,0.28]     |
|               |   |      | [-0.01,0.40]        | [34,74]       | [34,74]      | [-0.55,0.28]     |

$k = \text{number of estimates}; \text{Crl} = \text{credible intervals}; I^2 = \text{heterogeneity}.$

Figure 1. Forest plot showing the across-study effect size for the relationship between dominance rank and bib size in male house sparrows. Both meta 1 and meta 2 include published and unpublished estimates, with meta 2 including two non-reported estimates assumed to be zero (see section “Meta-analyses”). We show posterior means and 95% credible intervals from multilevel meta-analyses. Estimates are presented as standardized effect sizes using Fisher’s transformation ($Z_r$). Light, medium and dark grey show small, medium and large effect sizes, respectively [50]. $k$ is the number of estimates.
Moderators of the relationship between dominance rank and bib size

None of the three biological moderators studied (season, group composition and type of interactions) explained differences among studies (Table 3). Sampling effort (i.e. the ratio of interactions to individuals recorded) was not an important moderator either (Table 3).

Table 3. Results of the multilevel meta-regressions testing the effect of several moderators on the relationship between dominance rank and bib size in male house sparrows. Estimates are presented as standardized effect sizes using Fisher’s transformation ($Z_r$).

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

$k =$ number of estimates; CrI = credible intervals; $R^2_{\text{marginal}} = $ percentage of variance explained by the moderators. The factors season (non-breeding: 0, breeding: 1), group composition (mixed-sex: 0, male-only: 1), and type of interactions (all: 0, aggressive-only: 1) were mean-centred, and the covariates “sampling effort” and its squared term were z-transformed.
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 ($Z_r$). Precision = square root of the inverse of the variance.
Table 4. Results of the multilevel meta-regressions testing for time-lag and publication bias in the literature on status signalling in male house sparrows. Estimates are presented as standardized effect sizes using Fisher’s transformation (Zr). Credible intervals not overlapping zero are highlighted in 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] |
|                 |           | \( R^2_{\text{marginal}} = 29 [0, 66] \) |
| published vs.   | intercept | -0.09 [-0.37, 0.18] |
| unpublished (k = 85) | published\(^a\) | 0.50 [0.19, 0.81] |
|                 |           | \( R^2_{\text{marginal}} = 38 [0, 68] \) |

\( k \) = number of estimates; \( \text{Crl} \) = credible intervals; \( R^2_{\text{marginal}} \) = percentage of variance explained by the moderators; \(^a\) relative to unpublished. Year of publication was z-transformed.

Figure 3. Published effect sizes for the status signalling hypothesis in male house sparrows are larger than unpublished ones. We show posterior means and 95% credible intervals from a multilevel meta-regression. Estimates are presented as standardized effect sizes using Fisher’s transformation (Zr). Light, medium and dark grey show small, medium and large effects sizes, respectively [50]. \( k \) is the number of estimates.
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 ($Z_r$). 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.
for the status signalling hypothesis. First, the meta-analytic mean of unpublished studies was zero, compared to the medium effect size detected in published studies. Second, we found evidence for a time-lag bias. The effect size estimated in published studies has been decreasing over time, and recently published effects were on average no longer distinguishable from zero. Our findings call for reconsidering this textbook example in evolutionary and behavioural ecology, raise important concerns about the validity of the current scientific publishing culture, and should stimulate renewed attention to hypotheses explaining within-species variation in ornamentation.

The status signalling hypothesis [4] has been extensively tested to try and explain intraspecific trait variation (e.g. reptiles: [11]; insects: [10]; humans: [51]), particularly plumage variation [15]. Soon after the first empirical tests on birds, the black bib of house sparrows became a textbook example of the status signalling hypothesis [2,12,16,17], an idea that was later confirmed meta-analytically [18]. However, Nakagawa et al.’s [18] meta-analytic mean was over-estimated because only 9 low-powered studies were available (more in [52]). Here we updated that meta-analysis with newly published and unpublished data. Our results showed that the overall effect size is much smaller and much more uncertain than previously thought. The status signalling hypothesis is thus no longer a compelling explanation for the evolution of bib size across populations of house sparrows.

Similar contradicting conclusions have been reported for other model species. An exhaustive review and meta-analysis on plumage coloration of blue tits (Cyanistes caeruleus) revealed that, after dozens of publications studying the function of plumage ornamentation in this species, the only robust conclusion is that females' plumage differs from that of males [53]. Another example is the long-believed effect
of leg bands of particular colours on the perceived attractiveness of male zebra
finches (*Taeniopygia guttata*), which has been also experimentally and meta-
analytically refuted [54,55]. Finally, the existence of a badge of status in a non-bird
model species, the paper wasp (*Polistes dominulus*; [10]) has also been challenged
multiple times (e.g. [56–58]), generating doubts about its generality. Our findings
corroborate studies showing that abundant replication is needed before any strong or
general conclusion can be drawn [59], and highlight the existence of important
impediments to scientific progress in ecology.

Indeed, our results showed that the published literature on status signalling in house
sparrows is likely a biased subsample. The main evidence for this is that the mean
effect size of unpublished studies was essentially zero and clearly different from the
mean effect size based of published studies, which was of medium size.

Furthermore, this moderator (i.e. unpublished vs. published) explained a large
percentage of the model’s variance. In some of our own unpublished datasets, the
relationship between dominance rank and bib size was never formally tested
(Westneat & Bókony, *personal communication*), suggesting that these unpublished
datasets are not strictly speaking examples of the “file drawer problem” (*sensu* [60]).

Egger’s regressions failed to detect any funnel plot asymmetry, even in the meta-
analyses based on published effect sizes only (S4 Table). However, because
unpublished data indeed existed (i.e. those obtained for this study), the detection
failure was likely the consequence of the limited number of available estimates [61].

An additional common bias in the published literature is the time-lag bias [62]. We
detected evidence for such bias because the correlation between dominance rank
and bib size in published studies has decreased over time. Year of publication
explained a large percentage of variance, and accounting for year of publication
resulted in a strong reduction of the mean effect size across published studies (Table 4 vs. S4 Table). Time-lag bias has been detected in other ecological studies [63,64], including a meta-analysis on status signalling across bird species [15]. In the latter study, a positive overall (across-species) effect size persisted regardless of the time-lag bias, and no strong evidence for other types of biases was found [15]. However, Santos et al. [15] did not attempt to analyse unpublished data, so additional evidence is needed to determine the effect that unpublished data have on the overall validity of the status signalling hypothesis across bird species. If effect sizes based on unpublished data for other species were of similar magnitude than those obtained for house sparrows, the validity of the status signalling hypothesis across species would need reconsideration. The existence of publication bias in ecology has long been recognized [65–67]. Publication bias leads to false conclusions if not accounted for [30], and is, thus, a serious impediment to scientific progress.

In addition to estimating the overall effect size for a hypothesis, meta-analyses can also assess differences among estimates or heterogeneity [68,69]. Understanding the sources of heterogeneity is an important step towards the correct interpretation of a meta-analytic mean, and it can be done using meta-regressions [20]. Here, we found that the percentage of variance that was not attributable to sampling error (i.e. heterogeneity) was moderate. This value is below what has been quantified in ecological and evolutionary meta-analyses [21], and indicates that we accounted for large differences among estimates. Our meta-regressions based on biological moderators explained 20-23% of the variance (Table 3). However, none of the biological moderators that we tested influenced the overall effect size, but this might be because of limited sample sizes.
The badge of status idea is more subtle than typically portrayed [70]. While the evolution of badges of status in New and Old World sparrows has been related to sociality during the non-breeding season [71], additional factors need to be involved if the signal is to function in reducing aggression but retaining honesty. A recent study on black-crested titmice (*Baeolophus atricristatus*) suggested that badges of status might function only in high-competition scenarios [72]. However, time of the year was not a strong predictor in our models, though most studies were conducted in captivity, where food is normally provided *ad libitum*. Badges of status are expected to function both within and between sexes [4,12]. Indeed, we found little evidence that the status signalling function of bib size differed between male-only and mixed-sex flocks. Interestingly, when competing for resources, possessing a badge of status would be beneficial for both males and females. However, males but not female house sparrows have a bib. This sexual dimorphism suggests that the bib’s function is likely more important when competing for resources other than essential, *a priori* non-sex-specific, resources such as food, water, sand baths and roosting sites. [73,74] reported that female house sparrows preferentially choose males with large bibs (but see [75]), and bib size has been positively correlated with sexual behaviour [76,77], which suggests that the bib may play a role in mate choice. Furthermore, the original status signalling hypothesis posits that the main benefit of using badges of status would be to avoid fights, which should be particularly important when interacting with non-familiar individuals [4,12]. Although we did not have data to test whether unfamiliarity between contestants is an important pre-requisite for the status signalling hypothesis, we found no change in mean effect size when only obviously aggressive interactions were studied. In practice, testing whether the bib is important in mediating aggression between unfamiliar individuals
is difficult because the certainty of estimates of individual dominance increases over
time as more contests are recorded, but so does familiarity between individuals.

Our analyses have several potential limitations. First, although the number of studies
included in this meta-analysis is more than double that of the previous meta-analysis
[18], it is still limited. Also, it is likely (see above) that additional unpublished data are
stored in “file drawers” sensu [60]. Second, most tests included in this study were still
low-powered in terms of group size (median = 6 individuals/estimate, range = 4-41),
and the sample size is inflated because some of the published studies pooled
individuals from different groups (Fig 4). Third, although our results showed little
evidence of an effect of sampling effort on the overall effect size, the quality of the
data on dominance and bib size may still be a potential factor explaining differences
between studies. Additionally, data from studies not specifically designed to test the
status signalling hypothesis (such as some of the unpublished data) may be less
prone to confirmation bias [78], although other unknown factors might affect the
adequacy of such data for testing the hypothesis. Fourth, experimental effect sizes
will normally be larger, because effects of confounding factors can be reduced [67].

However, our systematic review only identified two studies where the status
signalling hypothesis was tested experimentally in house sparrows [37,79],
preventing us from estimating the meta-analytic mean for experimental studies.

Note, however, that the results of those experiments were inconclusive, and could be
partially explained by the phenomenon known as regression to the mean [78].

There are some additional explanations for the small and uncertain effect detected
by our meta-analyses. First, different populations might be under different selective
pressures regarding status signalling. Indeed, the population-specific heterogeneity
($I^2_{population ID}$) estimated in our meta-analyses was 15-16%, suggesting that
population-dependent effects might exist. Second, although none of the moderators had a strong influence on the overall effect size, the study-specific heterogeneity estimated in our meta-analyses ($I^2_{\text{study ID}} = 20-21\%$) suggests that the uncertainty observed could still be explained by the status signal being context-dependent. Although context-dependence is often invoked to explain variation between studies, there is little evidence for it. Last, most studies testing the status signalling hypothesis in house sparrows are observational (Table 1), and the only two experimental studies conducted so far were inconclusive [37,79]. Thus, it cannot be ruled out that the weak correlation observed between dominance status and bib size is driven by a third, unknown variable. In this respect, it has been proposed that the association between melanin-based coloration (such as the bib; e.g. [80,81]) and aggression is due to pleiotropic effects of the genes involved in regulating the synthesis of melanin (reviewed by [82]). Furthermore, bib size has been shown to correlate with testosterone, a hormone often involved in aggressive behaviour ([83]; but see [84]). Future studies should shift the focus towards understanding the function of bib size in wild populations and increase considerably the number of birds studied per group. The latter is essential because the statistical power of published tests of the status signalling hypothesis in house sparrows is alarmingly low (power = 8.5\%, S4 Appendix) and lower than the average in behavioural ecology [85]. In conclusion, our results challenge an established textbook example of the “badge of status” hypothesis to explain variation in ornament size. In house sparrows, we find no evidence that bib size consistently acts as a badge of status across studies and populations, and thus, this can no longer be considered a textbook example of the status signalling hypothesis. Furthermore, our analyses highlight the existence of publication and time-lag biases in the published literature, further undermining the
validity of past conclusions. Bias against the publication of small (“non-significant”) effects hinders scientific progress. We thus join the call for a change in incentives and scientific culture in ecology and evolution [78,86–88].

Materials and Methods

Systematic review

We used several approaches to maximize the identification of relevant studies. First, we included all studies reported in a previous meta-analysis that tested the relationship between dominance rank and bib size in house sparrows [18]. Second, we conducted a keyword search on the Web of Science, PubMed and Scopus from 2006 to June 2017 to find studies published after [18], using the combination of keywords [“bib/badge”, “sparrow”, “dominance/status/fighting”]. Third, we screened all studies on house sparrows used in a meta-analysis that tested the relationship between dominance and plumage ornamentation across species [15] to identify additional studies that we may have missed in our keyword search. We screened titles and abstracts of all articles and removed the irrelevant articles before examining the full texts. We followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA: [89]; see S1 Appendix). We only included articles in which dominance was directly inferred from agonistic dyadic interactions over resources such as food, water, sand baths or roosting sites (S1 Table).

Summary data extraction

Some studies had more than one effect size estimate per group of birds studied. When the presence of multiple estimates was due to the use of different statistical analyses on the same data, we chose a single estimate based on the following order of preference: (1) direct reports of effect size per group of birds studied (e.g.
correlation coefficient), (2) inferential statistics (e.g. $t$, $F$ and $\chi^2$ statistics) from analyses where group ID was accounted for and no other fixed effects were included, (3) direct reports of effect size where individuals from different groups where pooled together, (4) inferential statistics from models including other fixed effects. When the presence of multiple estimates was due to the use of different methods to estimate bib size and dominance rank on the same data, we chose a single estimate per group of birds or study based on the order of preference shown in S2 Appendix.

**Primary data acquisition**

We requested primary data (i.e. agonistic dyadic interactions and bib size measures) of all relevant studies identified by our systematic review. Additionally, we asked authors to share, if available, any unpublished data that could be used to test the relationship between dominance rank and bib size in house sparrows. We emailed the corresponding author, but if no reply was received, we tried contacting all the other authors listed. One study [33] provided all primary data in the original publication and, therefore, its author was not contacted. Last, we included our own unpublished data (S2 Appendix).

Most studies recorded data from more than one group of birds (Table 1). For each primary dataset obtained, we inferred the dominance hierarchy of each group of birds from the observed agonistic dyadic interactions (wins and losses) between individuals using the randomized Elo-rating method, which estimates dominance hierarchies more precisely than other methods [49]. We then used the provided measures of individual bib size (e.g. area outlined from pictures) or, if possible, calculated bib area from length and width measures following [33]. Subsequently, we
estimated the Spearman’s rho rank correlation ($\rho$) between individual rank and bib size for each group of birds. For one study [42], we received the already inferred dominance hierarchies for each group of birds, which we then correlated with bib size to obtain $\rho$.

**Effect size coding**

Regardless of their source (primary or summary data), we transformed all estimates (e.g. $\rho$, $F$ statistics, etc) into Pearson’s correlation coefficients ($r$), and then into standardized effect sizes using Fisher’s transformation ($Z_r$) for between-study comparison. We used the equations from [90] and [91]. Since log(0) is undefined, $r$ values equal to 1.00 and -1.00 were transformed to 0.975 and -0.975, respectively, before calculating $Z_r$. $Z_r$ values of 0.100, 0.310 and 0.549 were considered small, medium and large effect sizes, respectively (equivalent benchmarks from [50]).

When not reported directly, the number of individuals ($n$) was estimated from the degrees of freedom. The variance in $Z_r$ was calculated as: $V_{Zr} = 1/(n-3)$. Estimates ($k$) based on less than four individuals were discarded ($k = 33$ estimates discarded).

**Meta-analyses**

We ran two multilevel meta-analyses to test whether dominance rank and bib size were positively correlated across studies. The first meta-analysis, “meta 1”, included published and unpublished (re-)analysed effect sizes (i.e. effect sizes estimated from the studies we obtained primary data from), plus the remaining published effect sizes obtained from summary data (i.e. effect sizes for which primary data were unavailable).

Second, three studies reported “statistically non-significant” results without showing either the magnitude or the direction of the estimates (Table 1). Receipt of primary
data allowed us to recover some but not all the originally non-reported estimates.

Two “non-significant” estimates were still missing. To test the robustness of the results to those two cases of selective reporting, we ran an additional meta-analysis (see [92] for a similar approach). This second meta-analysis, “meta 2”, was like meta 1 but included the two non-significant non-reported estimates, which were assumed to be zero. Note that non-significant estimates can be either negative or positive, and thus, assuming that they were zero may have either underestimated or overestimated them, something we cannot know from non-reported estimates. Meta-analyses based on published studies only are shown in S3 Appendix.

We investigated inconsistency across studies by estimating the heterogeneity ($I^2$) from our meta-analyses following [20]. $I^2$ values around 25, 50 and 75% are considered as low, moderate and high levels of heterogeneity, respectively [69].

**Meta-regressions**

We tested if season, group composition and/or the type of interactions recorded had an effect on the meta-analytic mean. For that, we ran two multilevel meta-regressions that included the following moderators (hereafter “biological moderators”): (1) “season”, referring to whether the study was conducted during the non-breeding (September-February) or the breeding season (March-August); (2) “group composition”, referring to whether birds were kept in male-only or in mixed-sex groups; and, (3) “type of interactions”, referring to whether the dyadic interactions recorded were only aggressive (e.g. threats and pecks), or also included interactions that were not obviously aggressive (e.g. displacements). Because only three of 19 studies were conducted in the wild ($k = 12$ estimates; Table 1), we did not
include a moderator testing for captive versus wild environments. The three
biological moderators were mean-centred following [93] to aid interpretation.
The ratio of agonistic dyadic interactions recorded to the total number of interacting
individuals observed (hereafter “sampling effort”) is a measure of sampling effort that
correlates positively and logarithmically with the ability to infer the latent dominance
hierarchy [49]. The higher this ratio, the more precisely the latent hierarchy can be
inferred [49]. For the subset of studies for which the primary data of the agonistic
dyadic interactions were available (12 out of 19 studies; Table 1), we ran a multilevel
meta-regression including sampling effort and its squared term as z-transformed
moderators [93]. The squared term was included because of the observed
logarithmic relationship between sampling effort and the method’s performance [49].
This meta-regression tested whether sampling effort had an effect on the meta-
analytic mean.
For all meta-regressions, we estimated the percentage of variance explained by the
moderators ($R^2_{\text{marginal}}$) following [94].

**Random effects**

All meta-analyses and meta-regressions included the two random effects “population
ID” and “study ID”. Population ID was related to the geographical location of the
population of birds studied. We used Google maps to estimate the distance over
land (i.e. avoiding large water bodies) between populations, and assumed the same
population ID when the distance was below 50 km (13 populations; Table 1). Study
ID encompassed those estimates obtained within each specific study (19 studies).
Two studies tested the prediction twice for the same groups of birds (Table 1) and,
within each population, some individuals may have been sampled more than once.
However, we could not include “group ID” and/or “individual ID” as additional random effects due to either limited sample size or because the relevant data were not available.

Detection of publication bias

For the meta-analyses, we assessed publication bias using two methods that are based on the assumption that funnel plots should be symmetrical. First, we visually inspected asymmetry in funnel plots of meta-analytic residuals against the inverse of their precision (defined as the square root of the inverse of $V_{Zr}$). Funnel plots based on meta-analytic residuals (the sum of effect-size-level effects and sampling-variance effects) are more appropriate than those based on effect sizes when multilevel models are used [20]. Second, we ran Egger’s regressions using the meta-analytic residuals as the response variable, and the precision (see above) as the moderator [20]. If the intercept of such a regression does not overlap zero, estimates from the opposite direction to the meta-analytic mean might be missing and hence we consider this evidence of publication bias [20]. Further, we tested whether published estimates differed from unpublished estimates. For that, we ran a multilevel meta-regression that included population ID and study ID as random effects, and “unpublished” (two levels: yes (0), no (1)) as a moderator. This meta-regression was based on meta 1 (i.e. it did not include the two non-reported estimates). We did not use the trim-and-fill method [95,96] because this method has been advised against when significant heterogeneity is present [29,97], as it was the case in our meta-analyses (see below).

Finally, we analysed temporal trends in effect sizes that could indicate “time-lag bias”. Time-lag bias is common in the literature [63,64], and occurs when the effect
sizes of a specific hypothesis are negatively correlated with publication date (i.e. effect sizes decrease over time; [62]). We ran a multilevel meta-regression based on published effect sizes only, where “year of publication” was included as a z-transformed moderator [20].

All analyses were run in R v. 3.4.0 [98]. We inferred individual dominance ranks from agonistic dyadic interactions using the randomized Elo-rating method from the R package “aniDom” v. 0.1.3 [49,99]. Additionally, we described the dominance hierarchies observed in the groups of house sparrows for which primary data was available. For that we estimated the uncertainty of the dominance hierarchies using the R package “aniDom” v. 0.1.3 [49,99] and the triangle transitivity [100] using the R package “compete” 3.1.0 [101]. We used the R package “MCMCglmm” v. 2.24 [102] to run the multilevel meta-analytic (meta-regression) models [103]. For each meta-analysis and meta-regression, we ran three independent MCMC chains for 2 million iterations (thinning = 1800, burn-in = 200,000) using inverse-Gamma priors (V = 1, nu = 0.002). Model chains were checked for convergence and mixing using the Gelman-Rubin statistic. The auto-correlation within the chains was < 0.1 in all cases. For each meta-analysis and meta-regression, we chose the model with the lowest DIC value to extract the posterior mean and its 95% highest posterior density intervals (hereafter 95% credible interval).

Acknowledgments

AST and AG are members of and grateful for the support of the International Max Planck Research School (IMPRS) for Organismal Biology. We thank Katherine L. Buchanan, Fabrice Helfenstein, Anna Kulcsár, Ádám Z. Lendvai, Karin M. Lindström,
Thor Harald Ringsby, Alfonso Rojas Mora, Bernt-Erik Sæther, Erling J. Solberg, Zoltán Tóth and Jarle Tufto for sharing the primary data of published studies, and Emmi Schlicht for maintenance of the house sparrow database at Seewiesen. We thank Wolfgang Forstmeier and Lucy Winder for constructive feedback on the manuscript. This work was supported by the Volkswagen Foundation (to JS), the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme FP7/2007-2013 (CIG PCIG12-GA-2012-333096; to JS) the U.S. National Science Foundation (to DFW) and the Natural Environment Research Council (NERC: NE/N013832/1; to TB). The authors declare that they have no conflict of interest.

Data Accessibility

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

References

1. Hill GE. A red bird in a brown bag: the function and evolution of colorful plumage in the House Finch. New York: Oxford University Press; 2002.

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

3. Dale J, Dey CJ, Delhey K, Kempenaers B, Valcu M. The effects of life history and sexual selection on male and female plumage colouration. Nature. 2015;527: 367–370. doi:10.1038/nature15509

4. Rohwer S. The social significance of avian winter plumage variability. Evolution. 1975;29: 593–610. doi:10.2307/2407071

5. Prenter J, Elwood RW, Taylor PW. Self-assessment by males during energetically costly contests over precopula females in amphipods. Anim
6. Kelly CD, Godin J-GJ. Predation risk reduces male-male sexual competition in the Trinidadian guppy (*Poecilia reticulata*). Behav Ecol Sociobiol. 2001;51: 95–100. doi:10.1007/s002650100410

7. Sneddon LU, Huntingford FA, Taylor AC. Impact of an ecological factor on the costs of resource acquisition: fighting and metabolic physiology of crabs. Funct Ecol. 1998;12: 808–815. doi:10.1046/j.1365-2435.1998.00249.x

8. Neat FC, Taylor AC, Huntingford FA. Proximate costs of fighting in male cichlid fish: the role of injuries and energy metabolism. Anim Behav. 1998;55: 875–882. doi:10.1006/anbe.1997.0668

9. Jakobsson S, Brick O, Kullberg C. Escalated fighting behaviour incurs increased predation risk. Anim Behav. 1995;49: 235–239. doi:10.1016/0003-4722(95)80172-3

10. Tibbetts EA, Dale J. A socially enforced signal of quality in a paper wasp. Nature. 2004;432: 218–222. doi:10.1038/nature02949

11. Whiting MJ, Nagy KA, Bateman PW. Evolution and maintenance of social status-signalling badges. In: Fox SF, McCoy K, Baird TA, editors. Lizard Social Behavior. Baltimore: Johns Hopkins University Press; 2003. pp. 47–82.

12. Senar JC. Color displays as intrasexual signals of aggression and dominance. In: Hill GE, McGraw KJ, editors. Bird coloration: Function and evolution. London: Harvard University Press; 2006. pp. 87–136.

13. Beninde J. Naturgeschichte des Rothirsches. Monographie Wildsüugetiere IV. 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.
23. Mengersen K, Gurevitch J, Schmid CH. Meta-analysis of primary data. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in Ecology & Evolution. Princeton: Princeton University Press; 2013. pp. 300–312.

24. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2: 209–217. doi:10.1191/1740774505cn087oa

25. Barrowman NJ, Myers RA, Hilborn R, Kehler DG, Field CA. The variability among populations of coho salmon in the maximum reproductive rate and depensation. Ecol Appl. 2003;13: 784–793. doi:10.1890/1051-0761(2003)013[0784:TVAPOC]2.0.CO;2

26. Richards TA, Bass D. Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. Curr Opin Microbiol. 2005;8: 240–252. doi:10.1016/j.mib.2005.04.010

27. Krasnov BR, Vinarski M V., Korallo-Vinarskaya NP, Mouillot D, Poulin R. Inferring associations among parasitic gamasid mites from census data. Oecologia. 2009;160: 175–185. doi:10.1007/s00442-009-1278-0

28. Schmid CH, Landa M, Jafar TH, Giatras I, Karim T, Reddy M, et al. Constructing a database of individual clinical trials for longitudinal analysis. Control Clin Trials. 2003;24: 324–340. doi:10.1016/S0197-2456(02)00319-7

29. Jennions MD, Lortie C, Rosenberg M, Rothstein H. Publication and related
biases. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-
analysis in Ecology & Evolution. Princenton: Princeton University Press; 2013.

30. Rothstein H, Sutton A, Borenstein M. Publication bias in meta-analysis:
prevention, assessment and adjustments. Chichester: Wiley; 2005.

31. Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-
analysis: ten appraisal questions for biologists. BMC Biol. 2017;15: 18.
doi:10.1186/s12915-017-0357-7

32. Ritchison G. Plumage variability and social status in captive male house
sparrows. Kentucky Warbler. 1985;61: 39–42.

33. Møller A. Variation in badge size in male house sparrows Passer domesticus:
evidence for status signaling. Anim Behav. 1987;35: 1637–1644.
doi:10.1016/S0003-6472(87)80056-8

34. Andersson S, Åhlund M. Hunger affects dominance among strangers in house
sparrows. Anim Behav. 1991;41: 895–897. doi:10.1016/S0003-
3472(05)80356-2

35. Solberg EJ, Ringsby TH. Does male badge size signal status in small island
populations of house sparrows, Passer domesticus? Ethology. 1997;103: 177–
186. doi:10.1111/j.1439-0310.1997.tb00114.x

36. Liker A, Barta Z. Male badge size predicts dominance against females in
house sparrows. Condor. 2001;103: 151–157. doi:10.1650/0010-
5422(2001)103[0151:MBSPDA]2.0.CO;2

37. Gonzalez G, Sorci G, Smith LC, de Lope F. Social control and physiological
38. Hein WK, Westneat DF, Poston JP. Sex of opponent influences response to a potential status signal in house sparrows. Anim Behav. 2003;65: 1211–1221. doi:http://dx.doi.org/10.1006/anbe.2003.2132

39. Riters L V., Teague DP, Schroeder MB. Social status interacts with badge size and neuroendocrine physiology to influence sexual behavior in male house sparrows (Passer domesticus). Brain Behav Evol. 2004;63: 141–150. doi:10.1159/000076240

40. Lindström KM, Hasselquist D, Wikelski M. House sparrows (Passer domesticus) adjust their social status position to their physiological costs. Horm Behav. 2005;48: 311–320. doi:10.1016/j.yhbeh.2005.04.002

41. Bókony V, Lendvai ÁZ, Liker A. Multiple cues in status signalling: the role of wingbars in aggressive interactions of male house sparrows. Ethology. 2006;112: 947–954. doi:10.1111/j.1439-0310.2006.01246.x

42. Buchanan KL, Evans MR, Roberts ML, Rowe L, Goldsmith AR. Does testosterone determine dominance in the house sparrow Passer domesticus? An experimental test. J Avian Biol. 2010;41: 445–451. doi:10.1111/j.1600-048X.2010.04929.x

43. Dolnik O V., Hoi H. Honest signalling, dominance hierarchies and body condition in House Sparrows Passer domesticus (Aves: Passeriformes) during acute coccidiosis. Biol J Linn Soc. 2010;99: 718–726. doi:10.1111/j.1095-8312.2010.01370.x
44. Rojas Mora A, Meniri M, Glauser G, Vallat A, Helfenstein F. Badge size reflects sperm oxidative status within social groups in the house sparrow Passer domesticus. Front Ecol Evol. 2016;4: 67. doi:10.3389/fevo.2016.00067

45. Lendvai ÁZ, Barta Z, Liker A, Bókony V. The effect of energy reserves on social foraging: hungry sparrows scrounge more. Proc R Soc B. 2004;271: 2467–2472. doi:10.1098/rspb.2004.2887

46. Bókony V, Seress G, Nagy S, Lendvai ÁZ, Liker A. Multiple indices of body condition reveal no negative effect of urbanization in adult house sparrows. Landsc Urban Plan. 2012;104: 75–84. doi:10.1016/j.landurbplan.2011.10.006

47. Tóth Z, Bókony V, Lendvai ÁZ, Szabó K, Pénzes Z, Liker A. Kinship and aggression: do house sparrows spare their relatives? Behav Ecol Sociobiol. 2009;63: 1189–1196. doi:10.1007/s00265-009-0768-8

48. Bókony V, Kulcsár A, Liker A. Does urbanization select for weak competitors in house sparrows? Oikos. 2010;119: 437–444. doi:10.1111/j.1600-0706.2009.17848.x

49. Sánchez-Tójar A, Schroeder J, Farine DR. A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. J Anim Ecol. 2017; doi:10.1111/1365-2656.12776

50. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New Jersey: Taylor & Francis Inc; 1988.

51. Dixson BJ, Vasey PL. Beards augment perceptions of men’s age, social status, and aggressiveness, but not attractiveness. Behav Ecol. 2012;23: 481–490. doi:10.1093/beheco/arr214
52. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al.
Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14: 365–376. doi:10.1038/nrn3475

53. Parker TH. What do we really know about the signalling role of plumage colour in blue tits? A case study of impediments to progress in evolutionary biology. Biol Rev. 2013;88: 511–536. doi:10.1111/brv.12013

54. Seguin A, Forstmeier W. No band color effects on male courtship rate or body mass in the zebra finch: four experiments and a meta-analysis. PLoS One. 2012;7: e37785. doi:10.1371/journal.pone.0037785

55. Wang D, Forstmeier W, Ihle M, Khadraoui M, Jerónimo S, Martin K, et al. Irreproducible text-book knowledge: the effects of color bands on zebra finch fitness. Evolution. 2018; doi:10.5061/dryad.cc145b6

56. Cervo R, Dapporto L, Beani L, Strassmann J., Turillazzi S. On status badges and quality signals in the paper wasp Polistes dominulus: body size, facial colour patterns and hierarchical rank. Proc R Soc B. 2008;275: 1189–1196. doi:10.1098/rspb.2007.1779

57. Green JP, Field J. Interpopulation variation in status signalling in the paper wasp Polistes dominulus. Anim Behav. 2011;81: 205–209. doi:10.1016/j.anbehav.2010.10.002

58. Green JP, Leadbeater E, Carruthers JM, Rosser NS, Lucas ER, Field J. Clypeal patterning in the paper wasp Polistes dominulus: no evidence of adaptive value in the wild. Behav Ecol. 2013;24: 623–633. doi:10.1093/beheco/ars226
59. Aarts AA, Anderson JE, Anderson CJ, Attridge PR, Attwood A, Axt J, et al. Estimating the reproducibility of psychological science. Science. 2015;349: aac4716. doi:10.1126/science.aac4716

60. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull. 1979;86: 638–641. doi:10.1037/0033-2909.86.3.638

61. Sterne J, Egger M. Regression methods to detect publication and other bias in meta-analysis. In: Rothstein H, Sutton A, Borenstein M, editors. Publication Bias in Meta-Analysis2. Chichester: John Wiley; 2005. pp. 99–110.

62. Trikalinos T, Ioannidis JPA. Assessing the evolution of effect sizes over time. In: Rothstein H, Sutton A, Borenstein M, editors. Publication Bias in Meta-Analysis. Chichester: John Wiley; 2005. pp. 241–259.

63. Poulin R. Manipulation of host behaviour by parasites: a weakening paradigm? Proc R Soc B. 2000;267: 787–792. doi:10.1098/rspb.2000.1072

64. Jennions MD, Møller AP. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. Proc R Soc B. 2002;269: 43–48. doi:10.1098/rspb.2001.1832

65. Cassey P, Ewen JG, Blackburn TM, Moller AP. A survey of publication bias within evolutionary ecology. Proc R Soc B. 2004;271: S451–S454. doi:10.1098/rsbl.2004.0218

66. Jennions MD, Møller AP. Publication bias in ecology and evolution: an empirical assessment using the "trim and fill" method. Biol Rev. 2002;77: 211–222. doi:10.1017/S1464793101005875

67. Palmer AR. Quasireplication and the contract of error: Lessons from Sex
Ratios, Heritabilities and Fluctuating Asymmetry. Annu Rev Ecol Syst. 2000;31: 441–480. doi:10.1146/annurev.ecolsys.31.1.441

68. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21: 1539–1558. doi:10.1002/sim.1186

69. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327: 557–560. doi:10.1136/bmj.327.7414.557

70. Diep SK, Westneat DF. The integration of function and ontogeny in the evolution of status signals. Behaviour. 2013;150: 1015–1044. doi:10.1163/1568539X-00003066

71. Tibbetts EA, Safran RJ. Co-evolution of plumage characteristics and winter sociality in New and Old World sparrows. J Evol Biol. 2009;22: 2376–2386. doi:10.1111/j.1420-9101.2009.01861.x

72. Queller PS, Murphy TG. Seasonal variation in the utility of a status signaling system: plumage ornament predicts foraging success only during periods of high competition. PLoS One. 2017;12: e0185584. doi:10.1371/journal.pone.0185584

73. Møller AP. Badge size in the house sparrow Passer domesticus - Effects of intra- and intersexual selection. Behav Ecol Sociobiol. 1988;22: 373–378. doi:10.1007/BF00295107

74. Møller AP. Natural and sexual selection on a plumage signal of status and on morphology in house sparrows, Passer domesticus. J Evol Biol. 1989;2: 125–140. doi:10.1046/j.1420-9101.1989.2020125.x

75. Kimball RT. Female choice for male morphological traits in house sparrows,
Veiga JP. Mate replacement is costly to males in the multibrooded house sparrow: an experimental study. Auk. 1996;113: 664–671. doi:10.2307/4088987

Møller AP. Sexual behavior is related to badge size in the house sparrow Passer domesticus. Behav Ecol Sociobiol. 1990;27: 23–29. doi:10.1007/BF00183309

Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely false-positive findings – a practical guide. Biol Rev. 2017;92: 1941–1968. doi:10.1111/brv.12315

Diep S. The role of social interactions on the development and honesty of a signal of status [Internet]. University of Kentucky. 2012. Available: http://uknowledge.uky.edu/biology_etds/9/

Galván I, Wakamatsu K, Camarero PR, Mateo R, Alonso-Alvarez C. Low-quality birds do not display high-quality signals: the cysteine-pheomelanin mechanism of honesty. Evolution. 2015;69: 26–38. doi:10.1111/evo.12549

Galván I, Alonso-Alvarez C. Individual quality via sensitivity to cysteine availability in a melanin-based honest signaling system. J Exp Biol. 2017;220: 2825–2833. doi:10.1242/jeb.160333

Ducrest AL, Keller L, Roulin A. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. Trends Ecol Evol. 2008;23: 502–510. 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, Koiricheva 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.
Lajeunesse MJ. Recovering Missing or Partial Data from Studies: A Survey of Conversions and Imputations for Meta-analysis. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in Ecology & Evolution. Princenton: Princeton University Press; 2013. pp. 195–206.

Booksmythe I, Mautz B, Davis J, Nakagawa S, Jennions MD. Facultative adjustment of the offspring sex ratio and male attractiveness: a systematic review and meta-analysis. Biol Rev. 2017;92: 108–134. doi:10.1111/brv.12220

Schielzeth H. Simple means to improve the interpretability of regression coefficients. Methods Ecol Evol. 2010;1: 103–113. doi:10.1111/j.2041-210X.2010.00012.x

Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol. 2013;4: 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56: 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Duval S, Tweedie R. A nonparametric “Trim and Fill” method of accounting for publication bias in meta-analysis. J Am Stat Assoc. 2000;95: 89–98. doi:10.1080/01621459.2000.10473905

Moreno SG, Sutton AJ, Ades A, Stanley TD, Abrams KR, Peters JL, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol. 2009;9:
98. Team RC. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2017. Available: http://www.r-project.org/

99. Farine DR, Sánchez-Tójar A. aniDom: Inferring Dominance Hierarchies and Estimating Uncertainty [Internet]. 2017. Available: https://cran.r-project.org/package=aniDom

100. McDonald DB, Shizuka D. Comparative transitive and temporal orderliness in dominance networks. Behav Ecol. 2013;24: 511–520. doi:10.1093/beheco/ars192

101. Curley JP. compete: Analyzing Social Hierarchies [Internet]. 2016. Available: https://cran.r-project.org/web/packages/compete/index.html

102. Hadfield JD. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. J Stat Softw. 2010;33: 1–22. Available: http://mirror.dcc.online.pt/CRAN/web/packages/MCMCglmm/vignettes/Overview.pdf

103. Hadfield JD, Nakagawa S. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol. 2010;23: 494–508. doi:10.1111/j.1420-9101.2009.01915.x

104. Sánchez-Tójar A, Nakagawa S, Sánchez-Fortún M, Martin DA, Ramani S, Girndt A, et al. Supporting information for “Meta-analysis challenges a textbook example of status signalling: evidence for publication bias” [Internet]. 2018.
Supporting Information

The following supporting information is available for this article online:

- Supporting Information 1:
  - S1 Appendix: PRISMA flow diagram.
  - S2 Appendix: Information about data used.
  - S3 Appendix: Results based on published effect sizes only.
  - S4 Appendix: Power analysis.

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