bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias

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]

Note: *this preprint was peer-reviewed by two reviewers, a reviewing editor and an editor in chief at eLife.*

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, Bielefeld University, Germany.

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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

*Corresponding author: Alfredo Sánchez-Tójar, Department of Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33615, Bielefeld, Germany. Email address: alfredo.tojar@gmail.com. Phone number: 00495211062727. Fax number: 00495211062998

Authors' current addresses (if different to affiliations):

- Alfredo Sánchez-Tójar: Department of Evolutionary Biology, Bielefeld University, Germany.
- Moisès Sánchez-Fortún: Department de Biologia Evolutiva, Ecologia i Ciències Ambientals, University of Barcelona, Spain.
- Dominic A. Martin: Biodiversity, Macroecology, and Biogeography group, University of Gottingen, Germany.

Type of article: Research Article

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The status signalling hypothesis aims to explain within-species 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 primary data from published and unpublished studies to test whether dominance rank is positively associated with bib size across studies. Contrary to previous studies, the overall effect size (i.e. meta-analytic mean) was small and uncertain. Furthermore, we found several biases in the literature that further question the support available for the status signalling hypothesis. 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, and should stimulate renewed interest in understanding within-species variation in ornamental traits.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Plumage ornamentation is a striking example of colour and pattern diversity in the animal kingdom, and has attracted considerable research (Hill, 2002). Most studies have focused on sexual selection as the key mechanism to explain this diversity in ornamentation (Andersson 1994; Dale et al. 2015). The status signalling hypothesis explains within-species variation in ornaments by suggesting that these traits signal individual dominance status or fighting ability (Rohwer, 1975). Aggressive contests are costly in terms of energy use, and risk of injuries and predation (Jakobsson, Brick, & Kullberg, 1995; Kelly & Godin, 2001; Neat, Taylor, & Huntingford, 1998; Prenter, Elwood, & Taylor, 2006; Sneddon, Huntingford, & Taylor, 1998). These costs could be reduced if individuals can predict the outcome of such contests beforehand using so-called “badges of status” – i.e. two potential competitors could decide whether to avoid or engage in aggressive interactions based on the message provided by their opponent’s signals (Rohwer, 1975).

Patches of ornamentation have been suggested to function as badges of status in a wide range of taxa, including insects (Tibbetts & Dale 2004), reptiles (Whiting, Nagy, & Bateman, 2003) and birds (Senar, 2006). The status signalling hypothesis was originally proposed to explain variation in the size of mountain sheep horns (Beninde, 1937; Geist, 1966), but the hypothesis has become increasingly important in the study of variability in plumage ornamentation in birds (Rohwer, 1975; Senar, 2006). Among the many bird species studied (Santos, Scheck, & Nakagawa, 2011), the house sparrow (*Passer domesticus*) has become the classic textbook example of status signalling (Andersson 1994; Searcy & Nowicki 2005; Senar 2006; Davies, Krebs, & West 2012). 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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Nakagawa, Ockendon, Gillespie, Hatchwell, & Burke, 2007; Santos et al., 2011).

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 (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018; Nakagawa & Poulin, 2012; Nakagawa & Santos, 2012; Senior et al., 2016). 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 (Mengersen, Gurevitch, & Schmid, 2013). Meta-analysis of primary or raw data is a specific type of meta-analysis where studies are analysed in a consistent manner (Mengersen et al., 2013). 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 (Simmonds et al., 2005). Unfortunately, meta-analysis of primary data is still rarely used in evolutionary ecology (but see Barrowman et al. 2003; Richards & Bass 2005; Krasnov et al. 2009), perhaps due to the difficulty of obtaining the primary data of previously published studies until recently (Culina, Crowther, Ramakers, Gienapp, & Visser, 2018; Schmid et al., 2003).

An important feature of any meta-analysis is to identify the existence of bias in the literature (Nakagawa & Santos 2012; Jennions et al. 2013). For example, publication

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Jennions et al. 2013). Thus, publication bias can strongly affect the estimation of the meta-analytic mean, and distort the interpretation of the hypothesis (Rothstein, Sutton, & Borenstein, 2005). Several methods have been developed to identify this and other biases (Nakagawa & Santos 2012; Jennions et al. 2013); however, such methods are imperfect and dependent on the number of effect sizes available, and therefore should be considered as types of sensitivity analysis (Nakagawa, Noble, Senior, & Lagisz, 2017; Nakagawa & Santos, 2012).

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. 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, Supplementary File 1: S1 Appendix).

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. More information is available in the data files provided (Sánchez-Tójar, Nakagawa, et al., 2018).

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.
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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

						originally reported as statistically non-significant. Groups were tested twice.
12	Dolnik & Hoi 2010	Austria (captivity)	No	4 4	31 31	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: Lendvai et al. 2004; Bókony et al. 2012
15	Tóth et al.	Hungary (captivity)	Yes[3]	3	35	Unpublished data part of: Tóth et al. 2009; Bókony et al. 2012
16	Bókony et al.	Hungary (captivity)	Yes[3]	4	26	Unpublished data part of: Bókony, Kulcsár, and Liker 2010; Bókony et al. 2012
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.

[1] for primary data = yes, groups of birds containing less than 4 individuals were not included (see Materials and Methods).[2] Note: since most studies analysed more than one group of birds, the total number of males is different from group size in most cases (see below). [3] Information for the unpublished datasets is available in S5 Table.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dominance hierarchies

Mean sampling effort was 36 interactions/individual (SD = 24), which highlights that, overall, dominance hierarchies were inferred reliably across groups (Sánchez-Tójar, Schroeder, & Farine, 2018). 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 medium in 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 is at best a weak and unreliable signal of 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 regression tests are shown. Estimates are presented as standardized effect sizes using Fisher's transformation (*Zr*). 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***	**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]

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		0.20	15	20	53	-0.12
meta 2	87					
		[-0.01,0.40]	[0,46]	[0,49]	[34,74]	[-0.55,0.28]

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

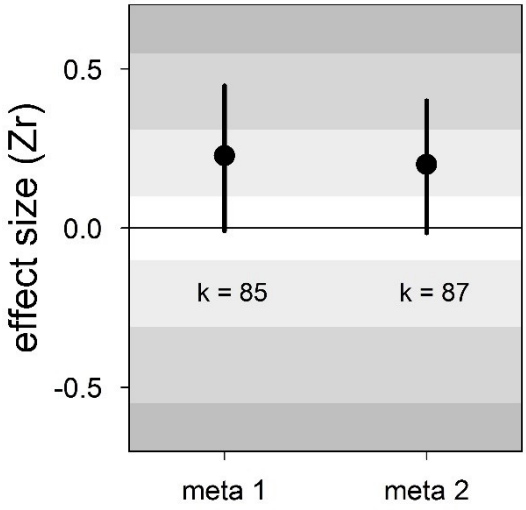


Figure 1. Forest plot showing the across-study effect size for the relationship between dominance rank and bib size in male house sparrows. Both meta 1 and meta 2 include published and unpublished estimates, with meta 2 including two non-reported estimates assumed to be zero (see section "Meta-analyses"). We show posterior means and 95% credible intervals from multilevel meta-analyses. Estimates are presented as standardized effect sizes using Fisher's transformation (*Zr*). Light, medium and dark grey show small, medium and large effect sizes, respectively (Cohen, 1988). *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) also was not an important moderator (Table 3).

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (*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]

k = number of estimates; CrI = credible intervals; $R^2_{marginal}$ = 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). However, published effect sizes were larger than unpublished ones, and

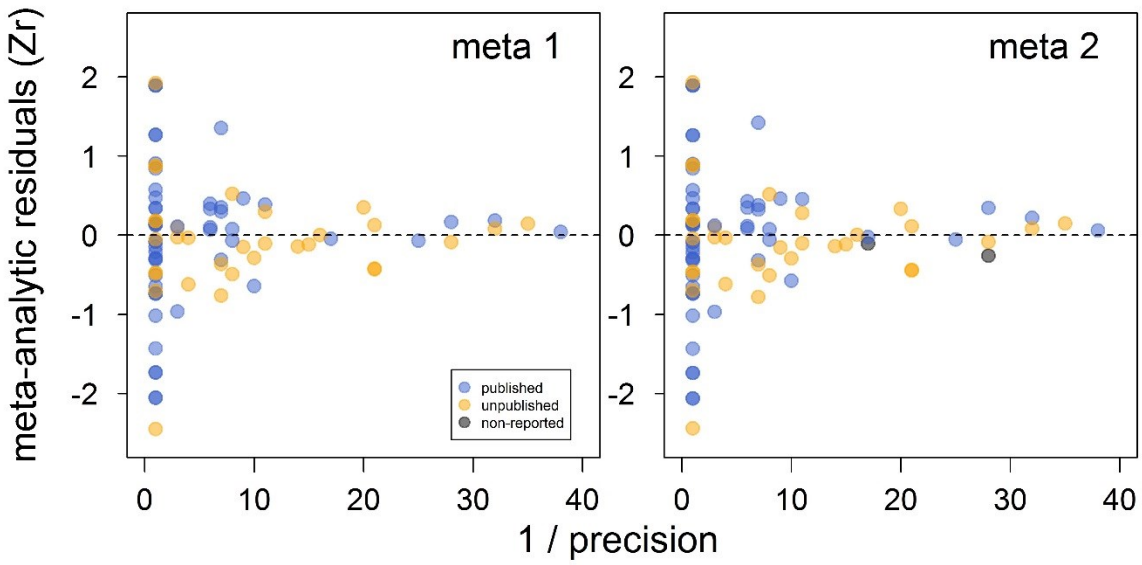
bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the latter were not different from zero (Table 4; Fig 3). Additionally, we found that the overall effect size decreased over time and approached zero (Table 4; Fig 4).

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]**
	$R^2_{marginal}$ =	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]

k = number of estimates; CrI = credible intervals; $R^2_{marginal}$ = percentage of variance explained by the moderators; [a] relative to unpublished. Year of publication was *z*-transformed.



bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

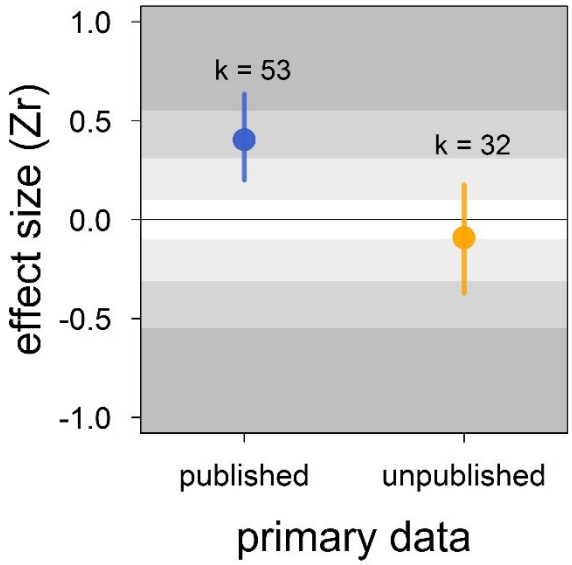


Figure 3. Published effect sizes for the status signalling hypothesis in male house sparrows are larger than unpublished ones. We show posterior means and 95% credible intervals from a multilevel meta-regression. Estimates are presented as standardized effect sizes using Fisher's transformation (*Zr*). Light, medium and dark grey show small, medium and large effects sizes, respectively (Cohen, 1988). *k* is the number of estimates.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

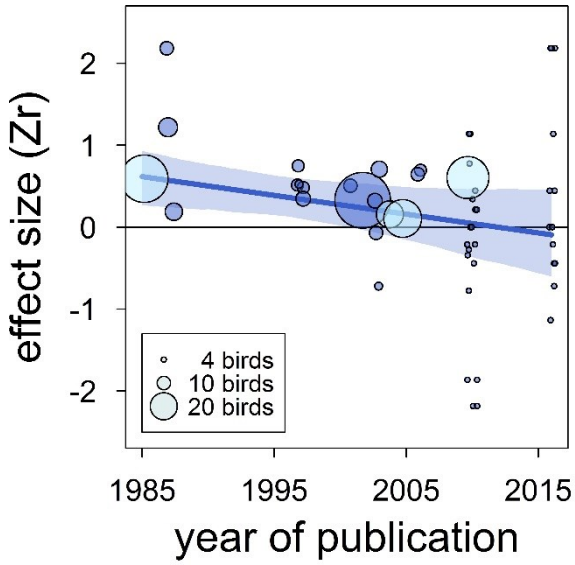


Figure 4. The overall published effect size for the status signalling hypothesis in male house sparrows has decreased over time since first described (*k* = 53 estimates from 12 publications). The solid blue line represents the model estimate, and the shading shows the 95% credible intervals of a multilevel 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 (Nakagawa et al., 2007), our updated meta-analytic mean was small, 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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

meta-analytic mean of unpublished studies was essentially zero, compared to the medium effect size detected in published studies. Second, we found that 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, and should stimulate renewed attention to hypotheses explaining within-species variation in ornamentation.

The status signalling hypothesis (Rohwer, 1975) has been extensively tested to try and explain within-species trait variation (e.g. reptiles: Whiting, Nagy, & Bateman 2003; insects: Tibbetts & Dale 2004; humans: Dixson & Vasey 2012), particularly plumage variation (Santos et al., 2011). Soon after the first empirical tests on birds, the black bib of house sparrows became a textbook example of the status signalling hypothesis (Andersson 1994; Searcy & Nowicki 2005; Senar 2006; Davies, Krebs, & West 2012), an idea that was later confirmed meta-analytically (Nakagawa et al., 2007). However, Nakagawa *et al*.'s (2007) meta-analytic mean was over-estimated because only 9 low-powered studies were available (more in Button et al. 2013). 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'

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plumage differs from that of males (Parker, 2013). 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 (Seguin & Forstmeier, 2012; Wang et al., 2018). Finally, the existence of a badge of status in a non-bird model species, the paper wasp (*Polistes dominulus*; Tibbetts & Dale 2004) has also been challenged multiple times (e.g. Cervo et al. 2008; Green & Field 2011; Green et al. 2013), generating doubts about its generality. Our findings corroborate studies showing that abundant replication is needed before any strong or general conclusion can be drawn (Aarts et al., 2015), and highlight the existence of important impediments (e.g. publication bias) to scientific progress in evolutionary ecology (Forstmeier, Wagenmakers, & Parker, 2017; Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018).

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 on 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 (D.F. Westneat & V. Bókony, *personal communication*, February, 2018), that is our unpublished datasets are not all examples of the "file drawer problem" (*sensu* Rosenthal 1979). Egger's regression tests failed to detect any funnel plot asymmetry, even in the meta-analyses based on published effect sizes only (S6 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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

effect sizes available (i.e. low power) and the moderate level of heterogeneity found in this study (Moreno et al., 2009; Sterne & Egger, 2005).

An additional type of publication bias is time-lag bias, where early studies report larger effect sizes than later studies (Trikalinos & Ioannidis, 2005). We detected evidence for such bias because the correlation between dominance rank and bib size in published studies has decreased over time and approached zero. 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*. S6 Table). Time-lag bias has been detected in other ecological studies (Poulin 2000; Jennions & Møller 2002), including a meta-analysis on status signalling across bird species (Santos et al., 2011). 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 (Santos et al., 2011). However, Santos *et al*. (2011) did not attempt to analyse unpublished data, so additional evidence is needed to determine the effect that unpublished data may 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 to 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 (Cassey et al. 2004; Jennions & Møller 2002; Palmer 2000). Publication bias leads to false conclusions if not accounted for (Rothstein et al., 2005), and is, thus, a serious impediment to scientific progress.

In addition to estimating the overall effect size for a hypothesis, meta-analyses are also used to assess heterogeneity among estimates (Higgins & Thompson 2002; Higgins et al. 2003). Understanding the sources of heterogeneity is an important step

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

towards the correct interpretation of a meta-analytic mean, and can be done using meta-regressions (Nakagawa & Santos, 2012). Here, we found that the percentage of variance that was not attributable to sampling error (i.e. heterogeneity) was moderate. This value is below the average calculated across ecological and evolutionary meta-analyses (Senior et al., 2016), 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 strongly influenced the overall effect size, possibly because of limited sample sizes.

The badge of status idea is more complex than typically portrayed (reviewed by Diep & Westneat 2013). Badges of status are expected to be particularly important in large and unstable groups of individuals where individual recognition would otherwise be difficult (Rohwer, 1975). While the evolution of badges of status in New and Old World sparrows has been related to sociality (i.e. flocking) during the non-breeding season (Tibbetts & Safran 2009), additional factors need to be involved if the signal is to function in reducing aggression but retaining honesty (Diep & Westneat 2013). Our results, however, did not show any evidence for a season-dependent effect as the moderator “season” (breeding vs. non-breeding) was not a strong predictor in our models. Badges of status are expected to function both within and between sexes (Rohwer, 1975; Senar, 2006). 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,

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a priori non-sex-specific, resources such as food, water, sand baths and roosting sites. Møller (1988, 1989) reported that female house sparrows preferentially choose males with large bibs (but see Kimball 1996), and bib size has been positively correlated with sexual behaviour (Veiga 1996; Møller 1990), 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 unfamiliar individuals (Rohwer, 1975; Senar, 2006). Although we did not have data to test whether unfamiliarity among 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 among unfamiliar individuals is difficult because the certainty of estimates of individual dominance increases over time as more contests are recorded, but so does familiarity among contestants.

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 ($I^2_{\text{population ID}}$) 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. However, context-dependence is often invoked *post hoc* to explain variation among studies, but strong evidence for it is lacking in most cases. Last, most studies testing the status signalling hypothesis in house sparrows are observational (Table 1), and

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the only two experimental studies conducted so far were inconclusive (Diep 2012; Gonzalez et al. 2002). 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. Galván et al. 2015; Galván & Alonso-Alvarez 2017) and aggression is due to pleiotropic effects of the genes involved in regulating the synthesis of melanin (reviewed by Ducrest, Keller, & Roulin 2008). Furthermore, bib size has been shown to correlate with testosterone, a hormone often involved in aggressive behaviour (Gonzalez, Sorci, Smith, & de Lope, 2001) but this relationship has not been consistently observed (Laucht, Kempenaers, & Dale, 2010). 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% for r = 0.20, Supplementary File 1: S3 Appendix) and lower than the average in behavioural ecology (Jennions & Møller 2003).

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 (Nakagawa et al., 2007), it is still limited. Also, it is likely (see above) that additional unpublished data are stored in “file drawers” *sensu* Rosenthal (1979). 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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

explaining differences among studies. Fourth, experiments will normally yield larger effect sizes than observational studies because effects of confounding factors can be reduced (Palmer, 2000). Nonetheless, our systematic review only identified two studies where the status signalling hypothesis was tested experimentally in house sparrows (Gonzalez et al. 2002; Diep 2012), preventing us from estimating the meta-analytic mean for experimental studies. Note, however, that the results of those experiments were inconclusive, and potentially affected by regression to the mean (Forstmeier et al., 2017).

In conclusion, our results challenge an established textbook example of the "badge of status" hypothesis to explain within-species 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 biases in the 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 (Forstmeier et al., 2017; Ihle, Winney, Krystalli, & Croucher, 2017; Nakagawa & Parker, 2015; Parker et al., 2016).

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 (Nakagawa et

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

al., 2007). Second, we conducted a keyword search on the Web of Science, PubMed and Scopus from 2006 to June 2017 to find studies published after Nakagawa et al. (2007), 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 (Santos et al., 2011) 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 (Supplementary File 2). We followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA: Moher et al. 2009; see “Reporting Standards Documents”). 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 χ^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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in S1-S3 Tables. In each case, the order of preference was determined prior to conducting any statistical analysis, and thus, method selection was blind to the outcome of the analyses (more details in Supplementary File 1).

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 (Møller 1987) provided all primary data in the original publication and, therefore, its author was not contacted. Last, we included our own unpublished data (S5 Table).

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) among individuals using the randomized Elo-rating method, which estimates dominance hierarchies more precisely than other methods (Sánchez-Tójar, Schroeder, et al., 2018). 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 Møller (1987). Subsequently, we estimated the Spearman's rho rank correlation (ρ) between individual rank and bib size for each group of birds. For one study (Buchanan, Evans, Roberts, Rowe, & Goldsmith, 2010), we received the already inferred dominance hierarchies for each group of birds, which we then correlated with bib size to obtain ρ.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Effect size coding

Regardless of their source (primary or summary data), we transformed all estimates (e.g. ρ, F statistics, etc) into Pearson's correlation coefficients (r), and then into standardized effect sizes using Fisher's transformation (Zr) for among-study comparison. We used the equations from Nakagawa and Cuthill (2007) and Lajeunesse (2013). 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 Zr. Zr values of 0.100, 0.310 and 0.549 were considered small, medium and large effect sizes, respectively (equivalent benchmarks from Cohen 1988). When not reported directly, the number of individuals (n) was estimated from the degrees of freedom. The variance in Zr 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, i.e. "*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).

The second meta-analysis, i.e. "*meta 2*", tested the robustness of the results of meta 1 to the inclusion of non-reported estimates from studies that 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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

missing. Thus, meta 2 was like meta 1 but included the two non-significant non-reported estimates, which were assumed to be zero (see Booksmythe et al. 2017 for a similar approach). 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 Supplementary File 1: S2 Appendix.

We investigated inconsistency across studies by estimating the heterogeneity (I^2) from our meta-analyses following Nakagawa and Santos (2012). I^2 values around 25, 50 and 75% are considered as low, moderate and high levels of heterogeneity, respectively (Higgins et al. 2003).

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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

biological moderators were mean-centred following Schielzeth (2010) 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 (Sánchez-Tójar, Schroeder & Farine, 2018). The higher this ratio, the more precisely the latent hierarchy can be inferred (Sánchez-Tójar, Schroeder & Farine, 2018). 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 (Schielzeth, 2010). The squared term was included because of the observed logarithmic relationship between sampling effort and the method's performance (Sánchez-Tójar, Schroeder & Farine, 2018). This meta-regression tested whether sampling effort had an effect on the meta-analytic mean: (i) a positive estimate would indicate that the meta-analytic mean may have been affected by the inclusion of studies with unreliable estimates of dominance rank. In contrast, (ii) a negative estimate would indicate that effect sizes were larger when based on unreliable estimates of dominance rank and hence provide evidence for the existence of publication bias.

For all meta-regressions, we estimated the percentage of variance explained by the moderators ($R^2_{marginal}$) following Nakagawa and Schielzeth (2013).

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

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

population of birds studied. We used Google maps to estimate the distance over land (i.e. avoiding large water bodies) among 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}) for each meta-analysis. 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 (Nakagawa & Santos, 2012). Second, we ran Egger's regressions using the meta-analytic residuals as the response variable, and the precision (see above) as the moderator (Nakagawa & Santos, 2012) for each meta-analysis. 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 (Nakagawa & Santos, 2012). 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),

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Duval & Tweedie 2000a,b) because this method has been advised against when significant heterogeneity is present (Moreno et al. 2009; Jennions et al. 2013), as it was the case in our meta-analyses (see section ”Results”).

Finally, we analysed temporal trends in effect sizes that could indicate “time-lag bias”. Time-lag bias is common in the literature (Jennions & Møller 2002; Poulin 2000), and occurs when the effect sizes of a specific hypothesis are negatively correlated with publication date (i.e. effect sizes decrease over time; Trikalinos & Ioannidis 2005). A decrease in effect size over time can have multiple causes. For example, initial effect sizes might be inflated due to low statistical power (“winner’s curse”) but published more easily and/or earlier due to positive selection of statistically significant results (reviewed by Koricheva, Jennions, & Lau 2013). We ran a multilevel meta-regression based on published effect sizes only, where “*year of publication*” was included as a *z*-transformed moderator (Nakagawa & Santos, 2012).

All analyses were run in R v. 3.4.0 (R Core Team 2017). We inferred individual dominance ranks from agonistic dyadic interactions using the randomized Elo-rating method from the R package “aniDom” v. 0.1.3 (Farine & Sánchez-Tójar, 2017; Sánchez-Tójar, Schroeder & Farine, 2018). 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 (Farine & Sánchez-Tójar, 2017; Sánchez-Tójar, Schroeder & Farine 2018) and the triangle transitivity (McDonald & Shizuka, 2013) using the R package “compete” 3.1.0 (Curley, 2016). We used the R

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

package "MCMCglmm" v. 2.24 (Hadfield 2010) to run the multilevel meta-analytic (meta-regression) models (Hadfield & Nakagawa 2010). 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). We report all data exclusion criteria applied and the results of all analyses conducted in our study.

Data and code availability

We provide all of the R code and data used for our analyses (Sánchez-Tójar, Nakagawa, et al., 2018).

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, Sanh K. Diep, Fabrice Helfenstein, Anna Kulcsár, Ádám Z. Lendvai, Karin M. Lindström, Thor Harald Ringsby, Alfonso Rojas Mora, Bernt-Erik Sæther, Emmi Schlicht, Erling J. Solberg, Zoltán Tóth and Jarle Tufto for providing the primary data of published and unpublished studies. We thank Wolfgang Forstmeier, Lucy Winder, and three anonymous reviewers for constructive feedback on the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. http://doi.org/10.1126/science.aac4716

Andersson, M. (1994). *Sexual selection*. New Jersey: Princeton University Press.

Andersson, S., & Åhlund, M. (1991). Hunger affects dominance among strangers in house sparrows. *Animal Behaviour*, *41*(5), 895–897. http://doi.org/10.1016/S0003-3472(05)80356-2

Barrowman, N. J., Myers, R. A., Hilborn, R., Kehler, D. G., & Field, C. A. (2003). The variability among populations of coho salmon in the maximum reproductive rate and depensation. *Ecological Applications*, *13*(3), 784–793. http://doi.org/10.1890/1051-0761(2003)013[0784:TVAPOC]2.0.CO;2

Beninde, J. (1937). *Naturgeschichte des Rothirshes. Monographie Wildsiiugetiere IV*. Leipzig.

Bókony, V., Kulcsár, A., & Liker, A. (2010). Does urbanization select for weak competitors in house sparrows? *Oikos*, *119*(3), 437–444. http://doi.org/10.1111/j.1600-0706.2009.17848.x

Bókony, V., Lendvai, Á. Z., & Liker, A. (2006). Multiple cues in status signalling: the role of wingbars in aggressive interactions of male house sparrows. *Ethology*, *112*(10), 947–954. http://doi.org/10.1111/j.1439-0310.2006.01246.x

Bókony, V., Seress, G., Nagy, S., Lendvai, Á. Z., & Liker, A. (2012). Multiple indices of body condition reveal no negative effect of urbanization in adult house sparrows. *Landscape and Urban Planning*, *104*(1), 75–84. http://doi.org/10.1016/j.landurbplan.2011.10.006

Booksmythe, I., Mautz, B., Davis, J., Nakagawa, S., & Jennions, M. D. (2017). Facultative adjustment of the offspring sex ratio and male attractiveness: a systematic review and meta-analysis. *Biological Reviews*, *92*(1), 108–134. http://doi.org/10.1111/brv.12220

Buchanan, K. L., Evans, M. R., Roberts, M. L., Rowe, L., & Goldsmith, A. R. (2010). Does testosterone determine dominance in the house sparrow *Passer domesticus*? An experimental test. *Journal of Avian Biology*, *41*(4), 445–451. http://doi.org/10.1111/j.1600-048X.2010.04929.x

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. http://doi.org/10.1038/nrn3475

Cassey, P., Ewen, J. G., Blackburn, T. M., & Møller, A. P. (2004). A survey of publication bias within evolutionary ecology. *Proceedings of the Royal Society B*, *271*(Suppl_6), S451–S454. http://doi.org/10.1098/rsbl.2004.0218

Cervo, R., Dapporto, L., Beani, L., Strassmann, J., & Turillazzi, S. (2008). On status badges and quality signals in the paper wasp *Polistes dominulus*: body size, facial colour patterns and hierarchical rank. *Proceedings of the Royal Society B*, *275*(1639), 1189–1196. http://doi.org/10.1098/rspb.2007.1779

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Taylor & Francis Inc.

Culina, A., Crowther, T. W., Ramakers, J. J. C., Gienapp, P., & Visser, M. E. (2018). How to do meta-analysis of open datasets. *Nature Ecology & Evolution*, *2*, 1053–1056. http://doi.org/10.1038/s41559-018-0579-2

Curley, J. P. (2016). compete: Analyzing Social Hierarchies. Retrieved from https://cran.r-project.org/web/packages/compete/index.html

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dale, J., Dey, C. J., Delhey, K., Kempenaers, B., & Valcu, M. (2015). The effects of life history and sexual selection on male and female plumage colouration. *Nature*, *527*(7578), 367–370. http://doi.org/10.1038/nature15509

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology* (4th ed.). Oxford: Wiley-Blackwell.

Diep, S. (2012). *The role of social interactions on the development and honesty of a signal of status*. University of Kentucky. Retrieved from http://uknowledge.uky.edu/biology_etds/9/

Diep, S. K., & Westneat, D. F. (2013). The integration of function and ontogeny in the evolution of status signals. *Behaviour*, *150*(9–10), 1015–1044. http://doi.org/10.1163/1568539X-00003066

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behavioral Ecology*, *23*(3), 481–490. http://doi.org/10.1093/beheco/arr214

Dolnik, O. V., & Hoi, H. (2010). Honest signalling, dominance hierarchies and body condition in House Sparrows *Passer domesticus* (Aves: *Passeriformes*) during acute coccidiosis. *Biological Journal of the Linnean Society*, *99*(4), 718–726. http://doi.org/10.1111/j.1095-8312.2010.01370.x

Ducrest, A. L., Keller, L., & Roulin, A. (2008). Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends in Ecology and Evolution*, *23*(9), 502–510. http://doi.org/10.1016/j.tree.2008.06.001

Duval, S., & Tweedie, R. (2000a). A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. http://doi.org/10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. http://doi.org/10.1111/j.0006-341X.2000.00455.x

Farine, D. R., & Sánchez-Tójar, A. (2017). aniDom: Inferring Dominance Hierarchies and Estimating Uncertainty. Retrieved from https://cran.r-project.org/package=aniDom

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. http://doi.org/10.1111/brv.12315

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS ONE*, *13*(7), e0200303. http://doi.org/10.1371/journal.pone.0200303

Galván, I., & Alonso-Alvarez, C. (2017). Individual quality via sensitivity to cysteine availability in a melanin-based honest signaling system. *The Journal of Experimental Biology*, *220*, 2825–2833. http://doi.org/10.1242/jeb.160333

Galván, I., Wakamatsu, K., Camarero, P. R., Mateo, R., & Alonso-Alvarez, C. (2015). Low-quality birds do not display high-quality signals: the cysteine-pheomelanin mechanism of honesty. *Evolution*, *69*(1), 26–38. http://doi.org/10.1111/evo.12549

Geist, V. (1966). The evolutionary significance of mountain sheep horns. *Evolution*, *20*(4), 558–566. http://doi.org/10.2307/2406590

Gonzalez, G., Sorci, G., Smith, L. C., & de Lope, F. (2001). Testosterone and sexual signalling in male house sparrows (*Passer domesticus*). *Behavioral Ecology and Sociobiology*, *50*(6), 557–562. http://doi.org/10.1007/s002650100399

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gonzalez, G., Sorci, G., Smith, L. C., & de Lope, F. (2002). Social control and physiological cose of cheating in status signalling male house sparrows (*Passer domesticus*). *Ethology*, *108*(4), 289–302. http://doi.org/10.1046/j.1439-0310.2002.00779.x

Green, J. P., & Field, J. (2011). Interpopulation variation in status signalling in the paper wasp *Polistes dominulus*. *Animal Behaviour*, *81*(1), 205–209. http://doi.org/10.1016/j.anbehav.2010.10.002

Green, J. P., Leadbeater, E., Carruthers, J. M., Rosser, N. S., Lucas, E. R., & Field, J. (2013). Clypeal patterning in the paper wasp *Polistes dominulus*: no evidence of adaptive value in the wild. *Behavioral Ecology*, *24*(3), 623–633. http://doi.org/10.1093/beheco/ars226

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175–182. http://doi.org/10.1038/nature25753

Hadfield, J. D. (2010). MCMC methods for multi-respoinse generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22. Retrieved from http://mirror.dcc.online.pt/CRAN/web/packages/MCMCglmm/vignettes/Overview.pdf

Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, *23*(3), 494–508. http://doi.org/10.1111/j.1420-9101.2009.01915.x

Hein, W. K., Westneat, D. F., & Poston, J. P. (2003). Sex of opponent influences response to a potential status signal in house sparrows. *Animal Behaviour*, *65*(6), 1211–1221. http://doi.org/http://dx.doi.org/10.1006/anbe.2003.2132

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. http://doi.org/10.1002/sim.1186

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. http://doi.org/10.1136/bmj.327.7414.557

Hill, G. E. (2002). *A red bird in a brown bag: the function and evolution of colorful plumage in the House Finch*. New York: Oxford University Press.

Ihle, M., Winney, I. S., Krystalli, A., & Croucher, M. (2017). Striving for transparent and credible research: practical guidelines for behavioral ecologists. *Behavioral Ecology*, *28*(2), 348–354. http://doi.org/10.1093/beheco/arx003

Jakobsson, S., Brick, O., & Kullberg, C. (1995). Escalated fighting behaviour incurs increased predation risk. *Animal Behaviour*, *49*(1), 235–239. http://doi.org/10.1016/0003-3472(95)80172-3

Jennions, M. D., Lortie, C., Rosenberg, M., & Rothstein, H. (2013). Publication and related biases. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of Meta-analysis in Ecology & Evolution* (pp. 207–236). Princenton: Princeton University Press.

Jennions, M. D., & Møller, A. P. (2002). Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biological Reviews*, *77*(2), 211–222. http://doi.org/10.1017/S1464793101005875

Jennions, M. D., & Møller, A. P. (2002). Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society B*, *269*(1486), 43–48. http://doi.org/10.1098/rspb.2001.1832

Jennions, M. D., & Møller, A. (2003). A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology*, *14*(3), 438–445.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

http://doi.org/10.1093/beheco/14.3.438

Kelly, C. D., & Godin, J.-G. J. (2001). Predation risk reduces male-male sexual competition in the Trinidadian guppy (*Poecilia reticulata*). *Behavioral Ecology and Sociobiology*, *51*(1), 95–100. http://doi.org/10.1007/s002650100410

Kimball, R. T. (1996). Female choice for male morphological traits in house sparrows, *Passer domesticus*. *Ethology*, *102*(4), 639–648. http://doi.org/10.1111/j.1439-0310.1996.tb01155.x

Koricheva, J., Jennions, M. D., & Lau, J. (2013). Temporal trends in effect sizes: causes, detection, and implications. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of Meta-analysis in Ecology & Evolution* (pp. 237–254). Princenton: Princeton University Press.

Krasnov, B. R., Vinarski, M. V., Korallo-Vinarskaya, N. P., Mouillot, D., & Poulin, R. (2009). Inferring associations among parasitic gamasid mites from census data. *Oecologia*, *160*(1), 175–185. http://doi.org/10.1007/s00442-009-1278-0

Lajeunesse, M. J. (2013). Recovering Missing or Partial Data from Studies: A Survey of Conversions and Imputations for Meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of Meta-analysis in Ecology & Evolution* (pp. 195–206). Princenton: Princeton University Press.

Laucht, S., Kempenaers, B., & Dale, J. (2010). Bill color, not badge size, indicates testosterone-related information in house sparrows. *Behavioral Ecology and Sociobiology*, *64*(9), 1461–1471. http://doi.org/10.1007/s00265-010-0961-9

Lendvai, Á. Z., Barta, Z., Liker, A., & Bókony, V. (2004). The effect of energy reserves on social foraging: hungry sparrows scrounge more. *Proceedings of the Royal Society B*, *271*(1556), 2467–2472. http://doi.org/10.1098/rspb.2004.2887

Liker, A., & Barta, Z. (2001). Male badge size predicts dominance against females in house sparrows. *The Condor*, *103*(1), 151–157. http://doi.org/10.1650/0010-5422(2001)103[0151:MBSPDA]2.0.CO;2

Lindström, K. M., Hasselquist, D., & Wikelski, M. (2005). House sparrows (*Passer domesticus*) adjust their social status position to their physiological costs. *Hormones and Behavior*, *48*(3), 311–320. http://doi.org/10.1016/j.yhbeh.2005.04.002

McDonald, D. B., & Shizuka, D. (2013). Comparative transitive and temporal orderliness in dominance networks. *Behavioral Ecology*, *24*(2), 511–520. http://doi.org/10.1093/beheco/ars192

Mengersen, K., Gurevitch, J., & Schmid, C. H. (2013). Meta-analysis of primary data. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of Meta-analysis in Ecology & Evolution* (pp. 300–312). Princenton: Princeton University Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. http://doi.org/10.1371/journal.pmed.1000097

Møller, A. (1987). Variation in badge size in male house sparrows *Passer domesticus*: evidence for status signaling. *Animal Behaviour*, *35*(6), 1637–1644. http://doi.org/10.1016/S0003-3472(87)80056-8

Møller, A. P. (1988). Badge size in the house sparrow *Passer domesticus* - Effects of intra- and intersexual selection. *Behavioral Ecology and Sociobiology*, *22*(5), 373–378. http://doi.org/10.1007/BF00295107

Møller, A. P. (1989). Natural and sexual selection on a plumage signal of status and on morphology in

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

house sparrows, *Passer domesticus*. *Journal of Evolutionary Biology*, *2*(2), 125–140. http://doi.org/10.1046/j.1420-9101.1989.2020125.x

Møller, A. P. (1990). Sexual behavior is related to badge size in the house sparrow *Passer domesticus*. *Behavioral Ecology and Sociobiology*, *27*(1), 23–29. http://doi.org/10.1007/BF00183309

Moreno, S. G., Sutton, A. J., Ades, A., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, *9*(1), 2. http://doi.org/10.1186/1471-2288-9-2

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. http://doi.org/10.1111/j.1469-185X.2007.00027.x

Nakagawa, S., Noble, D. W. A., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biology*, *15*(1), 18. http://doi.org/10.1186/s12915-017-0357-7

Nakagawa, S., Ockendon, N., Gillespie, D. O. S., Hatchwell, B. J., & Burke, T. (2007). Assessing the function of house sparrows' bib size using a flexible meta-analysis method. *Behavioral Ecology*, *18*(5), 831–840. http://doi.org/10.1093/beheco/arm050

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*, 88. http://doi.org/10.1186/s12915-015-0196-3

Nakagawa, S., & Poulin, R. (2012). Meta-analytic insights into evolutionary ecology: an introduction and synthesis. *Evolutionary Ecology*, *26*, 1085–1099. http://doi.org/10.1007/s10682-012-9593-z

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*, 1253–1274. http://doi.org/10.1007/s10682-012-9555-5

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. http://doi.org/10.1111/j.2041-210x.2012.00261.x

Neat, F. C., Taylor, A. C., & Huntingford, F. A. (1998). Proximate costs of fighting in male cichlid fish: the role of injuries and energy metabolism. *Animal Behaviour*, *55*(4), 875–882. http://doi.org/10.1006/anbe.1997.0668

Palmer, A. R. (2000). Quasireplication and the contract of error: Lessons from Sex Ratios, Heritabilities and Fluctuating Asymmetry. *Annual Review of Ecology and Systematics*, *31*, 441–480. http://doi.org/10.1146/annurev.ecolsys.31.1.441

Parker, T. H. (2013). 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. *Biological Reviews*, *88*(3), 511–536. http://doi.org/10.1111/brv.12013

Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., Chee, Y. E., … Nakagawa, S. (2016). Transparency in ecology and evolution: real problems, real solutions. *Trends in Ecology and Evolution*, *31*(9), 711–719. http://doi.org/10.1016/j.tree.2016.07.002

Poulin, R. (2000). Manipulation of host behaviour by parasites: a weakening paradigm? *Proceedings of the Royal Society B*, *267*(1445), 787–792. http://doi.org/10.1098/rspb.2000.1072

Prenter, J., Elwood, R. W., & Taylor, P. W. (2006). Self-assessment by males during energetically

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

costly contests over precopula females in amphipods. *Animal Behaviour*, *72*(4), 861–868. http://doi.org/10.1016/j.anbehav.2006.01.023

Richards, T. A., & Bass, D. (2005). Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Current Opinion in Microbiology*, *8*(3), 240–252. http://doi.org/10.1016/j.mib.2005.04.010

Ritchison, G. (1985). Plumage variability and social status in captive male house sparrows. *Kentucky Warbler*, *61*, 39–42.

Riters, L. V., Teague, D. P., & Schroeder, M. B. (2004). Social status interacts with badge size and neuroendocrine physiology to influence sexual behavior in male house sparrows (*Passer domesticus*). *Brain, Behavior and Evolution*, *63*(3), 141–150. http://doi.org/10.1159/000076240

Rohwer, S. (1975). The social significance of avian winter plumage variability. *Evolution*, *29*(4), 593–610. http://doi.org/10.2307/2407071

Rojas Mora, A., Meniri, M., Glauser, G., Vallat, A., & Helfenstein, F. (2016). Badge size reflects sperm oxidative status within social groups in the house sparrow *Passer domesticus*. *Frontiers in Ecology and Evolution*, *4*, 67. http://doi.org/10.3389/fevo.2016.00067

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. http://doi.org/10.1037/0033-2909.86.3.638

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester: Wiley.

Sánchez-Tójar, A., Nakagawa, S., Sánchez-Fortún, M., Martin, D. A., Ramani, S., Girndt, A., … Schroeder, J. (2018). Supporting information for "Meta-analysis challenges a textbook example of status signalling: evidence for publication bias." http://doi.org/10.17605/OSF.IO/CWKXB

Sánchez-Tójar, A., Schroeder, J., & Farine, D. R. (2018). A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. *Journal of Animal Ecology*, *87*(3), 594–608. http://doi.org/10.1111/1365-2656.12776

Santos, E. S. A., Scheck, D., & Nakagawa, S. (2011). Dominance and plumage traits: meta-analysis and metaregression analysis. *Animal Behaviour*, *82*(1), 3–19. http://doi.org/10.1016/j.anbehav.2011.03.022

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. http://doi.org/10.1111/j.2041-210X.2010.00012.x

Schmid, C. H., Landa, M., Jafar, T. H., Giatras, I., Karim, T., Reddy, M., … Levey, A. S. (2003). Constructing a database of individual clinical trials for longitudinal analysis. *Controlled Clinical Trials*, *24*(3), 324–340. http://doi.org/10.1016/S0197-2456(02)00319-7

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication. Reliability and deception in signaling systems*. New Jersey: Princeton University Press.

Seguin, A., & Forstmeier, W. (2012). No band color effects on male courtship rate or body mass in the zebra finch: four experiments and a meta-analysis. *PLoS ONE*, *7*(6), e37785. http://doi.org/10.1371/journal.pone.0037785

Senar, J. C. (2006). Color displays as intrasexual signals of aggression and dominance. In G. E. Hill & K. J. McGraw (Eds.), *Bird coloration: Function and evolution* (pp. 87–136). London: Harvard University Press.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S. A., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

implications. *Ecology*, *97*(12), 3293–3299. http://doi.org/10.1002/ecy.1591

Simmonds, M. C., Higgins, J. P. T., Stewart, L. A., Tierney, J. F., Clarke, M. J., & Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials*, *2*(3), 209–217. http://doi.org/10.1191/1740774505cn087oa

Sneddon, L. U., Huntingford, F. A., & Taylor, A. C. (1998). Impact of an ecological factor on the costs of resource acquisition: fighting and metabolic physiology of crabs. *Functional Ecology*, *12*(5), 808–815. http://doi.org/10.1046/j.1365-2435.1998.00249.x

Solberg, E. J., & Ringsby, T. H. (1997). Does male badge size signal status in small island populations of house sparrows, *Passer domesticus*? *Ethology*, *103*(3), 177–186. http://doi.org/10.1111/j.1439-0310.1997.tb00114.x

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

Team, R. C. (2017). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Tibbetts, E. A., & Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature*, *432*, 218–222. http://doi.org/10.1038/nature02949

Tibbetts, E. A., & Safran, R. J. (2009). Co-evolution of plumage characteristics and winter sociality in New and Old World sparrows. *Journal of Evolutionary Biology*, *22*(12), 2376–2386. http://doi.org/10.1111/j.1420-9101.2009.01861.x

Tóth, Z., Bókony, V., Lendvai, Á. Z., Szabó, K., Pénzes, Z., & Liker, A. (2009). Kinship and aggression: do house sparrows spare their relatives? *Behavioral Ecology and Sociobiology*, *63*(8), 1189–1196. http://doi.org/10.1007/s00265-009-0768-8

Trikalinos, T., & Ioannidis, J. P. A. (2005). Assessing the evolution of effect sizes over time. In H. Rothstein, A. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis* (pp. 241–259). Chichester: John Wiley.

Veiga, J. P. (1996). Mate replacement is costly to males in the multibrooded house sparrow: an experimental study. *The Auk*, *113*(3), 664–671. http://doi.org/10.2307/4088987

Wang, D., Forstmeier, W., Ihle, M., Khadraoui, M., Jerónimo, S., Martin, K., & Kempenaers, B. (2018). Irreproducible text-book knowledge: the effects of color bands on zebra finch fitness. *Evolution*, *72*(4), 961–976. http://doi.org/10.5061/dryad.cc145b6

Whiting, M. J., Nagy, K. A., & Bateman, P. W. (2003). Evolution and maintenance of social status-signalling badges. In S. F. Fox, K. McCoy, & T. A. Baird (Eds.), *Lizard Social Behavior* (pp. 47–82). Baltimore: Johns Hopkins University Press.

bioRxiv preprint doi: https://doi.org/10.1101/283150; this version posted September 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information

The following supporting information is available for this article online:

- Supplementary File 1:
    - S1 Appendix: Information about data used.
    - S2 Appendix: Results based on published effect sizes only.
    - S3 Appendix: Power analysis.
- Supplementary File 2: decision spreadsheet of the systematic review.