Facultative adjustment of the offspring sex ratio and male attractiveness: a systematic review and meta-analysis

Booksmythe, Isobel; Mautz, Brian; Davis, Jacqueline; Nakagawa, Shinichi; Jennions, Michael D

Abstract: Females can benefit from mate choice for male traits (e.g. sexual ornaments or body condition) that reliably signal the effect that mating will have on mean offspring fitness. These male-derived benefits can be due to material and/or genetic effects. The latter include an increase in the attractiveness, hence likely mating success, of sons. Females can potentially enhance any sex-biased benefits of mating with certain males by adjusting the offspring sex ratio depending on their mate’s phenotype. One hypothesis is that females should produce mainly sons when mating with more attractive or higher quality males. Here we perform a meta-analysis of the empirical literature that has accumulated to test this hypothesis. The mean effect size was small (r = 0.064–0.095; i.e. explaining <1% of variation in offspring sex ratios) but statistically significant in the predicted direction. It was, however, not robust to correction for an apparent publication bias towards significantly positive results. We also examined the strength of the relationship using different indices of male attractiveness/quality that have been invoked by researchers (ornaments, behavioural displays, female preference scores, body condition, male age, body size, and whether a male is a within-pair or extra-pair mate). Only ornamentation and body size significantly predicted the proportion of sons produced. We obtained similar results regardless of whether we ran a standard random-effects meta-analysis, or a multi-level, Bayesian model that included a correction for phylogenetic non-independence. A moderate proportion of the variance in effect sizes (51.6–56.2%) was due to variation that was not attributable to sampling error (i.e. sample size). Much of this non-sampling error variance was not attributable to phylogenetic effects or high repeatability of effect sizes among species. It was approximately equally attributable to differences (occurring for unknown reasons) in effect sizes among and within studies (25.3, 22.9% of the total variance). We discuss various practical reasons and theoretical arguments as to why small effect sizes should be expected, and why there might be relatively high variation among studies. Currently, there are no species where replicated, experimental studies show that mothers adjust the offspring sex ratio in response to a generally preferred male phenotype. Ultimately, we need more experimental studies that test directly whether females produce more sons when mated to relatively more attractive males, and that provide the requisite evidence that their sons have higher mean fitness than their daughters.

DOI: https://doi.org/10.1111/brv.12220

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-124247

Originally published at:
Facultative adjustment of the offspring sex ratio and male attractiveness: a systematic review and meta-analysis

Isobel Booksmythe\textsuperscript{1,2,*}, Brian Mautz\textsuperscript{1,2}, Jacqueline Davis\textsuperscript{1,3}, Shinichi Nakagawa\textsuperscript{4,5} and Michael D. Jennions\textsuperscript{1}

\textsuperscript{1}Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia
\textsuperscript{2}Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden
\textsuperscript{3}Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, U.K.
\textsuperscript{4}Department of Zoology, University of Otago, PO Box 56, Dunedin, New Zealand
\textsuperscript{5}Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

*Author for correspondence (E-mail: isobel.booksmythe@ieu.uzh.ch; Tel.: +41 044 6354804).

ABSTRACT

Females can benefit from mate choice for male traits (e.g. sexual ornaments or body condition) that reliably signal the effect that mating will have on mean offspring fitness. These male-derived benefits can be due to material and/or genetic effects. The latter include an increase in the attractiveness, hence likely mating success, of sons. Females can potentially enhance any sex-biased benefits of mating with certain males by adjusting the offspring sex ratio depending on their mate’s phenotype. One hypothesis is that females should produce mainly sons when mating with more attractive or higher quality males. Here we perform a meta-analysis of the empirical literature that has
accumulated to test this hypothesis. The mean effect size was small ($r = 0.064–0.095$; i.e. explaining $<1\%$ of variation in offspring sex ratios) but statistically significant in the predicted direction. It was, however, not robust to correction for an apparent publication bias towards significantly positive results. We also examined the strength of the relationship using different indices of male attractiveness/quality that have been invoked by researchers (ornaments, behavioural displays, female preference scores, body condition, male age, body size, and whether a male is a within-pair or extra-pair mate). Only ornamentation and body size significantly predicted the proportion of sons produced. We obtained similar results regardless of whether we ran a standard random-effects meta-analysis, or a multi-level, Bayesian model that included a correction for phylogenetic non-independence. A moderate proportion of the variance in effect sizes ($51.6–56.2\%$) was due to variation that was not attributable to sampling error (i.e. sample size). Much of this non-sampling error variance was not attributable to phylogenetic effects or high repeatability of effect sizes among species. It was approximately equally attributable to differences (occurring for unknown reasons) in effect sizes among and within studies ($25.3\%, 22.9\%$ of the total variance). There were no significant effects of year of publication or two aspects of study design (experimental/observational or field/laboratory) on reported effect sizes. We discuss various practical reasons and theoretical arguments as to why small effect sizes should be expected, and why there might be relatively high variation among studies. Currently, there are no species where replicated, experimental studies show that mothers adjust the offspring sex ratio in response to a generally preferred male phenotype. Ultimately, we need more experimental studies that test directly whether females produce more sons when mated to relatively more attractive males, and that provide the requisite evidence that their sons have higher mean fitness than their daughters.

**Key words:** female choice, mate choice, maternal effects, sex allocation, sex ratio adjustment, sexual selection.
(1) Environmental condition-dependent sex allocation

The birth sex ratio in many species deviates from unity, despite selection on parents to produce equal numbers of sons and daughters due to negative frequency-dependent selection (Düsing, 1884; Fisher, 1930). Although the reproductive value of individuals of the rare sex is always greater than that of the other sex, thereby tending to select for equal investment into each sex, this does not constrain the
offspring sex ratio to be even. First, if one sex is cheaper to produce and total sex allocation is equal, more offspring of the cheaper sex will be produced (West, 2009). Second, adaptive adjustment of sex allocation away from equal investment into both sexes is favoured by selection if the function relating investment to fitness returns differs for sons and daughters (i.e. if it is non-linear for at least one sex: Charnov, 1979) even if there is no variation in the resources mothers can allocate to reproduction (reviews: Frank, 1990, 1998 p.174–178). Third, facultative parental control of the offspring sex ratio is potentially advantageous when offspring fitness depends on some aspect of the breeding environment, which varies among parents (e.g. territory quality, female condition, adult size), if this condition dependence differently affects the absolute reproductive value of sons and daughters (Trivers & Willard, 1973). This is likely, as many environmental factors have sex-specific effects on offspring fitness (e.g. Jones, Nakagawa & Sheldon, 2009; Bowers, Thompson & Sakaluk, 2014). More generally, it is clear that males and females tend to have different fitness functions (review: Connallon & Clark, 2014). Given variation among parents in the quality of their breeding environment, those breeding in a relatively better environment should increase allocation towards the sex that has greater reproductive value when investment is higher. One way to do this is to bias the offspring sex ratio towards that sex. Trivers & Willard (1973) originally suggested that differences in maternal condition could lead to facultative sex ratio adjustment (henceforth, SRA) by parents. They noted that in polygynous mammals (e.g. many ungulates and pinnipeds) there is enormous variation in male mating success, and that larger, healthier males are disproportionately successful at acquiring mates when males compete. By contrast, female size has weak effects on fecundity (in many mammals the range in litter size is very low). In general, mothers in good physical condition have the ability to produce larger, healthier offspring. Trivers & Willard (1973) therefore argued that mothers in good condition should preferentially produce sons because of the relatively greater fitness gains they accrue from improving their sons’ body condition. There is empirical support for this claim from studies of ungulates (meta-analysis: Sheldon & West, 2004).
Hypotheses of condition-dependent SRA have since been broadened to include many other environmental and social factors that affect ‘breeding conditions’ and generate predictable differences in the relative fitness of sons and daughters (West, 2009, chapter 6). These scenarios include: production of daughters by high-ranking mothers when social rank is maternally inherited, as in primates (Schino, 2004); production of daughters by parasitoid insects laying eggs on larger hosts because of the stronger effect of size on female fecundity than on male mating success (meta-analysis: West & Sheldon, 2002); production of the helping sex if parents lack sufficient helpers-at-the-nest in cooperatively breeding birds and mammals (meta-analysis: West & Sheldon, 2002; but see Stubblefield & Orzack, 2013); seasonal or laying order effects if one sex requires more resources as a neonate, and food availability changes seasonally (e.g. for work on raptors see Rutz, 2012 and references therein). More generally, many factors have been posited to affect the optimal offspring sex ratio because they predict future resource availability for investment into offspring (e.g. territory quality, harem status; for a more complete list see West, 2009, p.183).

(2) Genetic condition-dependent sex ratio adjustment

There is an important condition-dependent sex allocation scenario that is not based on parental ability to predict environmentally determined resource availability. It is the argument that a male’s attractiveness or quality can differentially affect the fitness of his sons and daughters because of genetic effects (Burley, 1981, 1986). Males with larger ornaments and faster display rates, those in better body condition and/or those of larger body size are often preferred by females, gaining above-average mating success (Andersson, 1994). There is strong evidence for moderate heritability of sexually selected, preferred male traits (Prokuda & Roff, 2014; Wyman & Rowe, 2014). Burley (1981, 1986) therefore predicted that females mating with attractive males should preferentially produce sons because sons will benefit more than daughters by inheriting genes that increase their sexual attractiveness. Burley’s idea is compelling because few traits are more intimately related to reproductive success and, by extension, to fitness, than those under directional sexual selection.
(Shuster & Wade, 2003). There are also plausible grounds for assuming that some of the benefits of mating with attractive males are sex-specific, or sex-biased, accruing only or mainly to sons. Still, there are several important caveats. Preferred male traits can signal both material and genetic benefits (e.g. Hoelzer, 1989; meta-analysis: Møller & Jennions, 2001). If there are direct, fecundity-enhancing benefits of female choice due to material gains, then additional maternal resources might benefit daughters as much as, or even more than, sons (e.g. if the fecundity gains from increasing female size outweigh those of higher mating success for larger males or males who can invest more into costly ornaments). The adaptive prediction that females will produce sons when mated to attractive males implicitly assumes there are indirectly selected, genetic benefits of mate choice. Furthermore, for genetic benefits to favour SRA towards sons, the genes inherited from attractive sires must disproportionately advantage males (Burley, 1986). This is intuitively the case for arbitrarily attractive (Fisherman) male traits that are not inherited by daughters. The inheritance of genes that increase viability (so-called ‘good genes’) must, however, have a stronger effect on males than females to favour a male-biased offspring sex ratio. This might occur, for example, if sons benefit more than daughters from improved body condition because many sexual ornaments have highly condition-dependent expression (Cotton, Fowler & Pomiankowski, 2004). The benefits of an attractive father can also be greater for sons than daughters if attractiveness is similarly heritable for sons and daughters (i.e. both sexes are under sexual selection due to mate choice), but males experience greater variance in reproductive success (e.g. Du & Lu, 2010; Bowers et al., 2013), which is causally related to mate choice (i.e. sexual selection is stronger on males than females). In general, selection will favour a conditional sex allocation strategy when ornamentation imparts heritable benefits (be they due to the ornaments themselves or traits genetically correlated with ornamentation) that are greater for offspring of one sex than for the other (Fawcett et al., 2011).

If male-biased genetic benefits of mate choice exist, they can be enhanced by biasing the offspring sex ratio towards sons when mating with more attractive males, and towards daughters when mating...
with less attractive males (Burley, 1981, 1986). This form of conditional SRA has several potential fitness benefits for mothers that are driven by genetic effects. (1) It increases the production of ‘sexy sons’ (Weatherhead & Robertson, 1979; Firman, 2011; Prokop et al., 2012). There is, however, an implicit assumption that the benefits of a higher mating rate are not perfectly countered by greater naturally selected costs (e.g. higher mortality) (see Pen & Weissing, 2000), or countered by a trade-off with other sexually selected traits (e.g. reduced sperm competitiveness; see Lüpold et al., 2014). (2) Producing fewer daughters reduces the negative effects of sexually antagonistic genes when mated to an attractive male (if genes from more attractive sires make daughters of below-average fitness) (Rice & Chippindale, 2001; review: Rice, 2013). (3) Even in the absence of any such sexual antagonism, producing more daughters lessens the disadvantage of mating with an unattractive male as the value of daughters increases due to their ‘underproduction’ by females mated to attractive mates. Consequently, the fitness difference between females mated to attractive and unattractive mates declines (Fawcett et al., 2007, 2011).

(3) Models of sex ratio adjustment (SRA) in response to mate attractiveness

Greater production of sons due to genes that confer male-specific fitness benefits is distinct from most other conditional sex allocation scenarios. In the latter, the factors that favour biased sex ratios (e.g. maternal condition, habitat quality) tend to do so because they predict the availability of resources that can be invested into offspring, and the sexes differ in the effect of investment on their reproductive value (Frank, 1998). Consequently, most theoretical treatments of conditional sex allocation are not readily applicable to the mate attractiveness hypothesis. This is largely because covariances between genes for different traits (e.g. ornaments, preferences and sex allocation strategies), and changes in genetic variances have to be taken into account when modelling genetic benefits. Crucially, selection on condition-dependent SRA does not obviously affect variation in parental resource availability that favours SRA, and the sex-specific functions relating reproductive value to environmental quality remain the same. By contrast, selection on SRA in response to
additive genetic variation in attractiveness could lead to a co-evolutionary feedback that changes the fitness benefits of being more attractive because it will affect phenotypic variation in male attractiveness and female mating preferences (Fawcett et al., 2011).

To date, only four mathematical models have addressed the evolution of SRA in response to male attractiveness. Pen & Weissing (2000) provided the first formal model. They used a game theory approach with a discrete male trait (i.e. ornamented versus unornamented males) to show that an evolutionarily stable strategy for male-biased SRA when mated to ornamented males could evolve. The stability of such a strategy was, however, sensitive to the form of selection driving male trait exaggeration. Facultative SRA evolved under ‘good genes’ selection where the male trait signalled heritable viability (albeit in a simplified scenario where ‘viability’ genes did not benefit daughters). It did not evolve under a pure Fisherian process. In the latter case, all males are of equal fitness at equilibrium as natural selection perfectly balances sexual selection (ornamented males have greater mating success but higher mortality). There was therefore no fitness benefit from SRA towards sons when mated to an attractive male in this model.

Subsequently Fawcett et al. (2007) ran individual-based simulations that confirmed these initial results. Facultative SRA did not evolve when the male ornamental trait was purely Fisherian, but did evolve when it was a conditional indicator of viability, and when it was a Fisherian trait subject to a mutation bias (sensu Pomiankowski, Iwasa & Nee, 1991). Reassuringly, the results did not differ between two different proximate mechanisms of SRA. The findings were, however, slightly different if the male trait exhibited continuous rather than discrete variation. For a continuous male trait, the only scenario where facultative SRA unambiguously evolved was for a Fisherian trait subject to mutation bias. Although there was still a shift in sex allocation when the male trait was a conditional indicator, it was weak and inconsistent across simulation runs. The evolution of a stronger sex ratio bias for a Fisherian trait makes intuitive sense because such traits only benefit males. By contrast, conditional indicators signal the presence of viability genes that increase the fitness of both sexes. An additional, clear finding of Fawcett et al. (2007) was that the evolution of SRA in response to male
attractiveness was in all cases very slow, suggesting that it is under very weak selection. This is readily seen by comparing the rate of evolution of facultative SRA to the rate of change in the female preference and expression of preferred male traits in the simulations (Fawcett et al., 2007). Selection in the wild is likely to be even weaker than reported in the models, because (a) the models ignore the additional costs of actually adjusting the sex ratio (e.g. energetic costs if this requires strategic reabsorption of embryos, or time costs of delaying fertilization; Krackow, 1995); and (b) in practice, females obtain imperfect information on male genotype due to both non-heritable environmental influences on male sexual trait expression (Holman & Kokko, 2013), and female perceptual errors when assessing the expression of sexual signals (e.g. Farris & Ryan, 2011).

More recently, Fawcett et al. (2011) noted that facultative SRA in response to heritable sire attractiveness reduces sexual selection on male traits. A co-evolutionary feedback occurs because females mated to less attractive males ‘compensate’ by producing daughters. This reduces the fitness difference between females mated to more and less attractive males. The strength of selection for costly female preferences for males with more elaborate sexual traits is therefore weakened. In addition, sexual selection declines because the mean strength of the female preference in the population is lower when choosier females (those mating with more attractive males) produce fewer daughters, because only daughters express mating preferences. Both processes select for reduced investment into costly sexual traits by males. Facultative SRA effectively reduces the heritability of attractiveness and mating preferences because of the lower fitness gain of ornamentation (Fawcett et al., 2011). Fawcett et al. (2011) therefore conclude that species with elaborate sexual traits are the least likely to show facultative SRA in response to mate attractiveness. This is worrisome because empirical researchers testing for adaptive SRA have sought out species with strong sexual selection on males.

Booksmythe, Schwanz & Kokko (2013) suggested that the feedback between sexual selection and SRA based on attractiveness might be even more complicated than indicated by Fawcett et al. (2011).
Specifically, as the expression of a preferred male trait declines, it is increasingly difficult for females to distinguish and discriminate between males due to perceptual errors while mate sampling. More generally, imperfect maternal control of sex allocation is expected. They also noted that given costs of adjusting the sex ratio, this strategy might be eliminated when male trait expression is low. In an extension of Fawcett et al.’s (2011) model they allowed the ‘accuracy’ of sex allocation to evolve (i.e. the extent to which there is a ‘step-like’ transition between the production of daughters and sons; see West, Reece & Sheldon, 2002, for comments on empirical evidence that sex allocation is imprecise and not ‘step-like’). They also imposed a cost on females adjusting the sex ratio that increased with the precision of sex allocation. The simulation results were complex, suggesting there is not a simple linear relationship between the strength of SRA and male trait expression. In general, the greatest degree of SRA arose when sexual selection on male traits was intermediate in strength.

(4) The empirical evidence

Many empirical studies testing Burley’s original idea have accumulated in the last three decades (Fig. 1), but the available evidence for adjustment of the offspring sex ratio in response to mate attractiveness is still mixed. An early meta-analysis found a strong positive correlation between the proportion of sons in a brood and male attractiveness based on 11 avian studies of eight species ($r = 0.205$; West & Sheldon, 2002). This result seemed to spur a flurry of further research, but the extent to which the hypothesis is currently supported is less clear. Several studies have found a significant positive relationship between some measure of male attractiveness and the proportion of sons in a brood (e.g. Saino et al., 2002; Sato & Karino, 2010; Correa et al., 2011; Bowers et al., 2013; others have found no such relationship (Ramsay et al., 2003; Rutstein et al., 2004; Parker, 2005; Limbourg, Mateman & Lessells, 2013), and still others have reported a significant relationship in the opposite direction (Rutstein et al., 2005; Delhey et al., 2007). In addition, studies over multiple years or from different populations of the same species have reported that the relationship varies in strength and/or direction (e.g. Griffith et al., 2003; Rosivall et al., 2004; see also Parker, 2013, for a meta-analysis of
studies on blue tits). The extent to which intra-specific variation is due to sampling error or genuine biological differences is unknown. Ultimately, the diverse empirical results obtained to date reflect the challenge of identifying male traits that: (a) have a sound theoretical reason to affect the offspring sex ratio (rather than other forms of differential allocation such as preferential provisioning of eggs or juveniles of one sex; see Horváthová, Nakagawa & Uller, 2012), (b) are associated with male-biased genetic benefits, and (c) are applicable to the study population/species. In combination with the added difficulties of controlling for confounding variables, dealing with measurement error, and recurrent problems of low statistical power in behavioural ecology (Jennions & Møller, 2003) it is encouraging that any studies have obtained a significant result in the predicted direction. Even so, determining whether this has occurred more often than expected by chance remains a major challenge.

(5) The role of meta-analysis

Qualitative, narrative reviews have reported mixed evidence for adaptive sex allocation in birds and mammals, regardless of the source of selection on the offspring sex ratio (reviews: Clutton-Brock, 1986; Sheldon, 1998; Palmer, 2000; West, Herre & Sheldon 2000; West et al., 2002; Komdeur & Pen, 2002). By contrast, formal meta-analyses of hypotheses related to sex allocation in birds, mammals and insects have produced more clear-cut conclusions. Several meta-analyses have reported that the average relationship is significant in the predicted direction [e.g. helpers-at-the-nest: West & Sheldon, 2002; Griffin, Sheldon & West, 2005; but see Stubblefield & Orzack, 2013; mate attractiveness in birds: West & Sheldon, 2002; but see Parker, 2013; maternal condition in ungulates: Sheldon & West, 2004; host size in parasitoids: West & Sheldon, 2002; mate quality, laying date and season in birds: Cassey, Ewen & Møller, 2006 (correction of Ewen, Cassey & Møller, 2004); local mate competition: West, Shuker & Sheldon, 2005; split sex ratios in social insects: Meunier, West & Chapuisat, 2008; local resource competition in primates: Silk & Brown, 2008]. However, other meta-analyses, often using similar data sets, have reported that the mean relationship (effect size) is not
significantly different from the null expectation (zero) (e.g. maternal rank in primates: Brown & Silk, 2002; Schino, 2004; Silk, Willoughby & Brown, 2005; maternal condition in mammals: Cameron, 2004). Meta-analysis is a statistical technique designed to detect broad-scale patterns across empirical studies and quantitatively weigh studies by their accuracy in reporting the focal relationship. In short, estimates of a relationship are made with clearly stated confidence intervals. By contrast, narrative reviews are susceptible to vote counting (how many studies obtained a ‘yes’ or ‘no’ answer), which ignores the limitations of low power that plague most studies in evolutionary ecology. Meta-analysis has the additional advantage that it provides tools to gauge the extent to which moderating factors are associated with the observed relationships, and tools to assess the likely influence of publication biases (Jennions et al., 2012). For example, although the mean effect size did not differ from zero in Ewen et al. (2004), moderator variables had a significant influence on offspring sex ratios (Cassey et al., 2006). This suggests that, under certain conditions, there is still evidence that parents can skew sex ratios.

As new empirical data accumulate, meta-analyses need to be revised. This is especially important in fields where effect sizes tend to decline over time (see Jennions & Møller, 2002). For example, Pereira & Ioannidis (2011) have recently estimated – based on a decline in the reported mean effect sizes between meta-analyses conducted in 2005 and 2010 – that 37% of medical meta-analyses are likely to conclude falsely that there is evidence for a relationship. In addition, it is obvious to anyone who has conducted a meta-analysis that subjective decisions are made (e.g. how to extract effect sizes, how to search for papers). This can lead to researchers reporting different findings from meta-analyses on the same question, even when they use similar data sets. For a good example, see a recent dispute about the effect of the menstrual cycle on human mate choice between Wood et al. (2014) and Gildersleeve, Haselton & Fales (2014). For a topic closer to the current issue one can compare the findings of two meta-analyses on SRA in relation to helper number by Griffin et al. (2005) and Stubblefield & Orzack (2013).
The most recent meta-analysis of Burley’s (1981) hypothesis was by West & Sheldon (2002). Since then, there have been many studies with data on SRA in relation to mate attractiveness/quality (Fig. 1). The accumulation of studies allows for a more robust test of whether there is a general trend, which seems especially relevant in the light of recent theory questioning the extent to which there will be detectable selection for SRA based on male attractiveness (Fawcett et al., 2011). Here we use both standard meta-analytical and newer meta-regression approaches (including a control for phylogenetic relationships) to: (1) test whether there is a general trend for the offspring sex ratio to be male-biased when females are mated to attractive males. (2) Investigate if the relationship between the offspring sex ratio and male attractiveness depends on the type of male trait or measurement used as a proxy for attractiveness. For example, are ornaments better predictors of SRA than behaviour? (3) Test whether phylogenetic effects and differences in study design are associated with the strength of the relationship between the offspring sex ratio and male attractiveness.

II. METHODS

(1) Literature search

We used two complementary approaches to identify as many relevant studies as possible. First, we performed a key word search using the ISI Web of Knowledge search engine (for details of the databases covered see online supporting information, Appendix S1). We cross-referenced a set of key words related to sex ratio (e.g. “offspring sex ratio”, “proportion male”) to two sets of key words relating to male attractiveness: (1) general terms (e.g. attractive, ornament, “secondary sexual trait”) and (2) numerous specific secondary sexual characteristics (e.g. plumage, antler, horn, badge). The exact search term combinations used are given in Appendix S1. Second, we identified six empirical and theoretical papers that are highly cited in studies of sex allocation in response to mate attractiveness: Burley (1981), Ellegren, Gustafsson & Sheldon (1996), Frank (1990), Fawcett et al. (2007), Sheldon et al. (1999) and West & Sheldon (2002). At least one of these papers is likely to be cited by a study testing for SRA in response to male attractiveness. We performed a forward search.
to obtain papers citing any of these works. In total, the two search approaches yielded 4474 unique references (online Appendix S2). To finalize the data-checking and analysis protocols we conducted our final search on 17 July 2014. We did not add papers encountered by other means because these are more likely to be drawn from higher impact journals that might preferentially publish studies with significant results (Murtaugh, 2002). We also did not solicit unpublished data sets from colleagues. This procedure reduces the risk of biasing our estimates of effect sizes (see Jennions et al., 2013). As with an empirical study, our goal was to obtain an extensive, representative sample of studies rather than to locate every available data point. The criteria we used to decide whether a study contained potentially usable data are provided in online Appendix S1. In brief, we excluded studies of plants, humans and farmed animals. We looked for studies where there were likely to be data on both brood/litter offspring sex ratios (i.e. not just population-level ratios) and measurements of male traits that are generally assumed to be targets of female choice in many taxa (e.g. body condition, body size and, most importantly, male ornamentation).

(2) Data extraction and coding

We collected and transformed study results into a standard effect size measure ($Z_r$; see below) to allow for comparisons between studies. Reported effects were used in the following order of preference: (1) direct reports of an effect size (e.g. correlations) ($N = 13$); (2) effect measures with magnitude and direction (e.g. regression coefficients and standard errors, standardized mean differences, or descriptive statistics) ($N = 79$); (3) raw numbers (e.g. contingency tables) ($N = 12$); (4) inferential (test) statistics (often from multiple regressions) (e.g. $t$, $z$, $F$ and $\chi^2$ values, $P$ values) ($N = 161$). In some cases we could only note that there was a non-significant effect ($N = 11$). All data were converted to correlation coefficients $r$, and then to $Z_r$, following the formulae listed in Lipsey & Wilson (2001) and Nakagawa & Cuthill (2007). The variance of $Z_r$ is $1/(N–3)$ (Lipsey & Wilson, 2001). We calculated separate effect sizes for each measure of attractiveness from every independent
population and/or experiment (i.e. each unique group of parental individuals; hereafter referred to as a ‘study’). Thus, if a paper tested the relationship between a male ornament and offspring sex ratio in two distinct populations, these were considered independent studies and the paper contributed two effect sizes to the meta-analysis. In some cases, where data were analysed by year, each year could reasonably be considered an independent sample of individuals and was treated as a separate study (e.g. Griffith et al., 2003).

The included studies defined and measured male attractiveness in different ways. Namely, (a) researchers chose focal traits based on evidence from previous research that particular morphological or behavioural traits influenced male success in attracting and/or mating with females; (b) researchers assumed that sexually dimorphic traits reflect sexual selection on males such that the trait will affect male reproductive success; (c) researchers based their hypothesis on the general prediction that extra-pair males are more attractive than average; (d) in a few studies researchers directly measured female preferences (e.g. association time with different males) to score or rank test males. We were broad in our criteria for what made an ‘attractive’ trait acceptable for inclusion as some studies cited strong evidence for the effect on male mating success of a certain trait (e.g. experimental manipulation of the trait affected female mate choice), while others would include this trait precisely because it had been shown to influence sex ratios in other species. To test whether relationships varied across different attractiveness proxies, including broad classes of male phenotypic traits, we grouped them into seven categories: Ornament, Behaviour, Size, Condition, Preference, Age, and Extra-pair paternity (EPP) (see Table 1). Each effect size was assigned to one of these categories. ‘Ornament’ included measures of the size or degree of elaboration of secondary sexual traits; for example, colour measures or the size of a particular body part such as a comb, badge or elongated tail feather. Studies that had manipulated male morphology (e.g. by adding coloured leg bands, or lengthening or shortening tail feathers) were also included in this category. Similarly, ‘Behaviour’ included any measures of male sexual behaviour that had been identified in the studies as attractive to females, such as song or call attributes, courtship displays, and nest decoration collected by males. ‘Size’
included measures of structural body size and body mass. ‘Condition’ included measures of moult timing, parasite infection, and the residuals of mass on body size. ‘Preference’ measured attractiveness using female preference scores for particular males (e.g. association time in choice tests). Measures of male age or breeding experience were grouped into the category ‘Age’. Finally, EPP studies compared the sex ratio of within-pair and extra-pair offspring either within broods (extra-pair young \textit{versus} within-pair young), between broods (i.e. broods with \textit{versus} without EPP), or across broods (extra-pair and within-pair young pooled across all broods) (see Table 1). 

Studies could include multiple effect size estimates due to measurement of several male traits from the same category (e.g. two different ornaments) and/or from different categories (e.g. an ornament and a measure of condition). The presence of more than one effect size per study presents potential problems of non-independence of data for analyses. We dealt with this issue by running the meta-analysis in two ways. First, we used standard random-effects meta-analytic models, which require one effect-size estimate per study (Borenstein \textit{et al.}, 2009). Here we calculated a mean effect size for each study by conducting a within-study meta-analysis (i.e. weighting the estimates by the sample size, as this occasionally differed even within a study). This was done separately for analyses restricted to a single trait type, or by pooling effect sizes across all trait types for analyses of ‘All’ traits. Second, we ran multi-level, random-effects meta-analytic models, which allowed for the inclusion of multiple, non-independent effect size estimates per study by including the levels study and species (Nakagawa & Santos, 2012; Mengersen, Jennions & Schmid, 2013). In these analyses individual effect sizes were used and we included the trait type as a moderator.

The effect size calculated was the Fisher-transformed correlation coefficient, $Z_{r}$. By convention, researchers refer to $r = 0.1$, $0.3$ and $0.5$ as small, medium and large effect sizes (Cohen, 1988). We present $Z_{r}$ in our results, which is almost identical to $r$ at low values (e.g. $0.1003$, $0.310$ and $0.549$ are the equivalent $Z$, benchmarks). The effect size was coded as positive if an increase in the male attractiveness measure (i.e. greater ornamentation, more courtship, larger size, better condition, stronger preference score) was associated with a greater proportion of male offspring. We assumed
that older males were more attractive as this is a general assumption in bird studies (Møller & Ninni, 1998; Brooks & Kemp, 2001), and that extra-pair males were more attractive than within-pair males (Møller & Birkhead, 1994; but see Dunn, Whittingham & Pitcher, 2001; reviews: Westneat & Stewart, 2003; Forstmeier et al., 2014). The variance in $Z_r$ was calculated based on the number of broods in the study. We used the number of broods rather than unique breeding pairs as the sample size because this was the information most widely available across studies. Using different sources of $N$ among studies to weight them is problematic because it gives greater weighting to studies that used lower-level analyses (e.g. an analysis that uses $N =$ chicks). Where necessary, however, effect sizes were calculated using the source of $N$ that was appropriate for the inferential statistic the authors presented. Notably, when we converted $t$ or $z$ values from mixed-effect models, we used $N =$ broods rather than $N =$ chicks (which potentially inflates the effect size; cf. Nakagawa & Cuthill, 2007; see also Section IV). We also coded information about the study that might have influenced the effect size, for use in a moderator analysis: study species, year of publication ($z$-transformed for analysis), study design (experimental or observational), and study location (laboratory or field research). There was substantial variation among studies in the stage at which offspring sex ratio was measured; we therefore included as a moderator ‘primary’ (measured at birth/laying) versus ‘secondary’ (some juvenile mortality had occurred) sex ratio rather than excluding studies on this basis.

Seventeen papers reported insufficient information to calculate an effect size for one or more measured male traits (most commonly only reporting that the effect was ‘not significant’). We obtained the relevant information directly from the authors of 10 papers, retaining 11 unknown effects from the remaining seven papers. In addition, 28 papers contained one or more effects for which only model fit statistics or $P$ values were reported, allowing us to extract the magnitude of an effect but not its direction. We obtained the direction of the effect from the authors of three of these papers. We retained 56 ‘directionless’ effects from the remaining 25 papers. We attempted to contact all authors so ‘incomplete’ data reflects no response by an author or their inability to recalculate the effect size (i.e. data files unavailable).
In all analyses, we tried to account for the uncertainty introduced by effects of unknown magnitude and/or direction. First, we excluded all unknown and directionless effect sizes (dataset 1). Second, if this analysis showed a significant mean effect, we added the unknown effect sizes and assigned them a value of zero (i.e. assumed that they were highly ‘non-significant’) (dataset 2). Third, if the analysis was still significant, we added the directionless effect sizes and also assigned them a value of zero (dataset 3). Finally, if the analysis was still significant, we assigned all directionless effect sizes a negative direction (dataset 4). Each successive dataset represents increasingly conservative assumptions about the magnitude/direction of missing data. This allowed us to test the robustness of our findings to the value of unknown and directionless effects.

(3) Standard random-effects meta-analyses

We conducted random-effects meta-analysis for all trait categories combined (the ‘All’ analysis), as well as for each of the seven trait categories separately. These analyses assumed independence of effect sizes. Each study therefore contributed one effect size per trait category (and thus per meta-analysis). As noted above, if studies provided more than one effect size estimate per category, the weighted mean effect for the category was calculated. Likewise, for the ‘All’ analysis we calculated the weighted mean effect across all effect sizes for the study. We then assigned the variance to this combined effect size based on the study sample size (number of broods). This is the most conservative option as it assumes complete dependence between multiple estimates (e.g. that \( r = 1 \) for the correlation between within-study estimates: see p. 228 in Borenstein et al., 2009).

For all random-effects meta-analytic models, we used a restricted maximum likelihood (REML) method to estimate \( \tau^2 \) (the true between-study variance), which was combined with the within-study variance for each effect size (Borenstein et al., 2009). For the ‘All’ data set, we used meta-regression to determine the effects of four study-level moderator variables: year of publication, design (experimental/observational), location (field/laboratory), and stage at which sex ratio was measured.
(primary/secondary) on the effect size. Based on the results, we did not include these moderators in
the subsequent analyses for separate trait categories.

We also tested whether there was more heterogeneity in effect sizes among studies than could be
explained by sampling error alone using Cochran’s $Q$ statistic. This formally tests whether there is
greater variation in effect sizes among studies than expected if the true effect is identical for all
studies (Hedges & Olkin, 1985). We also summarised heterogeneity using $I^2$, which is often defined
as the ratio of true heterogeneity to the total variance among studies. Borenstein et al. (2009) note,
however, that this is not always correct because the ratio assumes a constant within-study variance,
which is not the case (sampling error varies due to studies differing in sample size). They suggest that
$\hat{I}^2$ be treated as a measure of ‘inconsistency’ in effect sizes among studies. Higgins et al. (2003) have
suggested that benchmarks for $\hat{I}^2$ of 25%, 50% and 75% indicate low, moderate and high values of
inconsistency among studies.

All statistical analyses for the above models were conducted in $R$ statistical software using the
metafor package (Viechtbauer, 2010).

(4) Multilevel meta-analyses including phylogenetic effects

While the standard meta-analyses described above included studies of vertebrate and invertebrate
taxa, the multilevel meta-analyses with and without phylogenetic effects were limited to bird species,
because there were few studies on other taxa in our data set (84.2% of species, 87.1% of included
papers and 92.0% of effect sizes were for birds). The other taxa also had relatively few species
sampled within each family. A tree including all species would be unreliable in terms of branch
lengths (i.e. distance relationships between species). By contrast, we now have avian super-trees that
include all extant species (Jezt et al., 2012; see below). For birds, our data set had enough variation to
conduct a phylogenetic meta-analysis. Studies could contribute multiple effect sizes because the
multilevel model partially corrects for non-independence by modelling multiple random effects for
both study and species. A complete correction of non-independence requires modelling the
correlations among effect sizes. For example, effect sizes for ornament and body size are likely to be correlated when these traits were measured from the same group of individuals. Given the difficulty of determining actual correlations among related effect sizes (as such correlations are almost always not reported), the multilevel models we report in the main text assume correlations of zero. Additionally, as supplementary materials (Table S2) we report on multilevel models that conservatively assume all correlations to be 0.5. Note that qualitatively the results are very similar among these models. As for the standard random-effect meta-analyses we initially excluded all directionless and unknown effect sizes (avian dataset 1) and then added them to subsequent analyses if the preceding analysis yielded a significant mean effect size (avian datasets 2–4).

To estimate the overall effect size we used four Bayesian Markov Chain Monte Carlo (MCMC) generalised linear mixed-effects models, implemented in the MCMCglmm package of R (Hadfield, 2010; Hadfield & Nakagawa, 2010). First, the multilevel meta-analytic model (model A) included study and species as random effects. We did not include ‘paper’ as a random effect because it closely overlaps with study (99 and 81 levels, respectively, as most papers comprised a single study). This model was analogous to the standard random-effects meta-analysis described above, but used MCMC estimator rather than REML estimator and had two extra random effects (study and species). Second, we extended model A by running a phylogenetic meta-analysis that included a phylogenetic tree (model B). Third, we ran a multilevel meta-regression (model C) that extended model A by including moderators (fixed effects): year of publication, seven categories of attractiveness measure (Ornament, Behaviour, Size, Condition, Preference, Age, and EPP), design (experimental or observational), and location (laboratory or field study). This was to assess whether these study characteristics impacted the effect sizes, similar to the meta-regression used in the standard random-effects meta-analysis. Finally, for completeness, we ran a meta-regression model (model D) that extended model C by including phylogenetic effects. Models C and D were both run without the intercept. This ensures that the mean estimates for the seven trait categories (conditional on the other fixed effects) were directly obtained (for details, see Section III).
For both phylogenetic models (B and D) we used two different avian phylogenetic trees: (i) a version of the Ericson backbone trees from Jetz et al. (2012), referred to as the Ericson tree, and (ii) a version of the Hackett backbone trees also from Jetz et al. (2012), referred to as the Hackett tree. These phylogenies are presented as online supporting information in Fig. S1. Both trees were ultrametric with branch lengths. Each phylogenetic model produced quantitatively similar and qualitatively identical results using either tree. We therefore only present results for the Ericson tree (for results from the Hackett tree see Table S1 and Fig. S2). As an index of phylogenetic signal we calculated phylogenetic heritability $H^2$, which can be defined as the proportion of phylogenetic variance in relation to the sum of all other variance components except sample error variance. Along with $\hat{I}^2_{\text{phylogeny}}$ (see below), $H^2$ provides the magnitude of phylogenetic signal in the data. $H^2$ is equivalent to Pagel’s $\lambda$ when the unit of analysis is at the species level (Housworth, Martins & Lynch, 2004).

For all multilevel (phylogenetic) models, we used an inverse Gamma prior ($V = 0.002$ and $\nu = 1$) for all random effects. Each model was run for 1,100,000 iterations with a thinning interval of 1000 after a burn-in of 100,000 iterations. These settings resulted in posterior distributions consisting of 1000 samples for all parameters. We assessed chain mixing by examining autocorrelation among posterior samples; they were less than 0.1 for all estimated parameters (Hadfield, 2010). We report our point estimates from models based on posterior means rather than posterior modes. We considered fixed factor estimates statistically significant if the 95% credible interval (CI, also known as the highest posterior density, which can be considered a Bayesian equivalent of the frequentist confidence interval) did not overlap zero. To quantify heterogeneity in multilevel meta-analytic models, we calculated a modified version of the $\hat{I}^2$ statistics, following Nakagawa & Santos (2012). Basically, this partitions the proportion of ‘unknown’ variance not attributable to sampling variance (i.e. traditional $\hat{I}^2$) into the contribution of random factors. In the current case these are the variance in effect sizes due to phylogenetic relatedness, differences among species, differences among studies, and differences in within-study variation (also called ‘residual’ variation).
percentages of total variation due to these four sources equals the traditional $I^2$ of Higgins et al. (2003).

(5) Bias detection

A funnel plot reports the relationship between effect sizes and a measure of their variance (e.g. sampling variance, standard error or sample size). Variability in the observed effect size should decrease as sample size increases (i.e. as the estimate becomes more accurate due to lower sampling error). This yields a funnel-shaped plot. Almost all statistical tests for publication bias are based on the assumption that the funnel plot should be symmetrical (Jennions et al., 2013). Asymmetry is assumed to reflect publication bias. We used funnel plots for visual inspection of funnel asymmetry, and then conducted Egger’s regression (Egger et al., 1997) to test statistically for publication bias (i.e. funnel plot asymmetry) for each category of attractiveness measure. We then conducted trim-and-fill tests (the L0 model; Duval & Tweedie, 2000a, b) as another test for publication bias. Asymmetry due to publication bias is only predicted to appear as missing studies on the left of the plot (i.e. fewer significant results in the direction opposite to a true non-zero effect), but trim-and-fill can detect asymmetry on either side. We present results when studies were missing on the left or right. We do so because, in a broader context, the proportion of tests that identify missing studies on the left versus right can be viewed as an indicator of the extent to which asymmetry is due to publication bias (left only) rather than other factors that might generate asymmetry on either side. Small sample sizes limit the detection of significant asymmetry using Egger’s regression (Sterne & Egger, 2005), and funnel asymmetry can occur by chance especially when the sample size is low. Consequently, care is required when interpreting these publication bias analyses. Tests for bias were only run on ‘Data 1’ data sets (i.e. excluding unknown and directionless effect-size estimates). For the multi-level models (i.e. the avian-only data sets) we conducted analyses of bias on meta-analytic residuals (sensu Nakagawa & Santos, 2012). This ensures that we meet assumptions of independence, and that we reduce the effect of heterogeneity due to moderators that
could induce funnel asymmetry. All publication bias analyses were conducted using functions in the R package metafor.

III. RESULTS

(1) Attrition of studies

We scanned the title and abstract of all 4474 papers generated by our search protocol to identify studies potentially suitable for inclusion in the meta-analysis. We then read and assessed the methods and results sections of short-listed studies to determine their eligibility for inclusion. We identified 118 eligible papers. After full-text screening, 25 of these papers were excluded (Fig. 2, online Table S2). In total, 276 effect sizes for the meta-analysis were computed from 113 independent studies within 93 published papers. Of these, 254 effect sizes were from studies of birds (48 species). The remaining studies contributed nine effect sizes from four insect species, eight effect sizes from one species of fish, four effect sizes from three reptile species, and one effect size from a mammal. The full data set, and the data set reduced to one effect size per trait category per study are provided in online Appendix S3.

(2) Standard random-effects meta-analyses

The mean effect of male attractiveness on offspring sex ratio for traits pooled across all categories was small, positive and statistically significant for ‘dataset 1’ (Fig. 3A, Table 2). Females mated to attractive males produced more sons. This result was robust to the implementation of the successively more conservative datasets 2, 3 and 4 (all $P < 0.005$; Table 2). There were, however, reasons for concern about the effect of publication bias. The Egger’s regression test revealed significant asymmetry in the funnel plot ($t_{79} = 3.45, P = 0.001$; Table 3), and trim-and-fill estimated 17 ($\pm 5.9$ S.E.) studies ‘missing’ from the left-hand side of the plot (see Fig. 4A). Correcting for this bias substantially reduced the mean effect size estimate, which no longer differed from zero ($P = 0.149$; Table 3).
The heterogeneity due to variation between studies that could not be attributed to sampling error, $I^2$, was moderate at 56.20% (sensu Higgins et al., 2003). Given the wide range of taxa, traits and study designs, one might expect a higher level of heterogeneity. To try to explain some of this heterogeneity, we ran a meta-regression model including the moderators year of publication, design, location and stage at which sex ratio was measured. They did not, however, explain a significant proportion of the observed heterogeneity ($Q_1 = 2.848$, $P = 0.584$; $b_{[\text{year}]} = 0.014$, 95% confidence interval, CI $[-0.324, 0.061]$, $b_{[\text{Experiment-Observation}]} = 0.026$ $[-0.080, 0.133]$, $b_{[\text{Field-Laboratory}]} = 0.082$ $[-0.022, 0.185]$, $b_{[\text{Primary-Secondary}]} = 0.014$ $[-0.082, 0.109]$). Given this result, we did not run meta-regression models for the other trait categories.

When we examined the effects for each trait category separately, we found a significantly positive, albeit small, effect for both male ornamentation and male size (Fig. 3A, Table 2). Females mated to more ornamented males or to larger males produced a higher proportion of sons than those mating with less ornamented or smaller mates. For both categories, this result held across datasets 1–3, and was only statistically non-significant in the most conservative ‘dataset 4’ analysis (where directionless effect sizes were assigned negative values) (Table 2). For ornaments, correction for a potential publication bias identified using trim-and-fill resulted in a mean effect size estimate that was non-significant ($P = 0.350$; Table 3). By contrast, for male size, trim-and-fill identified 7 (±3.6 S.E.) studies ‘missing’ from the right of the funnel plot. Correcting for this asymmetry increased the mean effect size estimate to 0.15 (±0.03 S.E.; $P < 0.001$). The mean effect size estimate did not differ from zero for the other five trait types (behaviour, body condition, female preference score, male age, and male extra-pair or within-pair status) (Fig. 3A, Table 2). For male age, the Egger’s test suggested a publication bias, and trim-and-fill analysis suggested there were four missing studies. Even so, the addition of putative effect sizes still did not yield a significant effect size (Table 3). Trim-and-fill analyses also suggested there were three missing studies on the left for body condition, and two on the right for preference scores. Again, the addition of these studies did not yield significant mean effect size estimates (both $P > 0.089$; Table 3). Funnel plots are shown in Fig. 4.
The estimates of heterogeneity for the individual trait categories (dataset 1 analyses) ranged from $I^2 = 42.56$ to $73.43\%$ (Table 2). This is moderate to high (sensu Higgins et al., 2003), but it should be noted that uncertainty in estimates of $I^2$ is high when sample sizes are small. These results, along with those for all trait categories combined, suggest that over half the variation in effect sizes was due to systematic between-study differences. The sources of this variation are explored in Section IV.

Notably, however, there was no heterogeneity in the effect size estimates for extra-pair mating status ($I^2 = 0$), indicating all variation was due to sampling error, although the number of studies in this analysis was rather limited.

**(3) Multilevel meta-analyses including phylogenetic effects**

For both the simple and phylogenetic multilevel meta-analyses we found small but significantly positive mean effect sizes for ‘avian dataset 1’ analyses (Table 4; Fig. 3A,B,C). These estimates were very similar to those from the standard random-effects meta-analysis for all traits pooled. The simple multilevel meta-analytic model (model A) had 95% credible intervals that did not span zero for avian datasets 2 and 3, but did for the most conservative dataset 4 (but even then, the lower CI was only just below zero at $-0.002$; Table 4). For the phylogenetic multilevel meta-analysis (model B), the mean effect had a 95% credible interval that included zero for avian dataset 2, so we did not run datasets 3 or 4.

In both multilevel meta-analytic models we can partition the variance not due to sampling error (i.e. heterogeneity). Here, $I^2_{\text{study}}$ is the variance due to differences among studies while $I^2_{\text{species}}$ is that among species. In the phylogenetic models, $I^2_{\text{species}}$ is the variation due to the non-phylogenetic portion of species-level differences (associated with repeatability of effect sizes among species), while $I^2_{\text{phylogeny}}$ is the variance due to phylogenetic relatedness. $I^2_{\text{effect size}}$ is the residual variance that is not due to the sampling error. $I^2_{\text{total}}$ is the sum of these values. In the simple multilevel meta-analysis, the largest component of variance for ‘avian dataset 1’ (effects of known direction and magnitude) was at the effect-size level (24.5%) then the study level (21.0%). Species had little
influence (6.0%) (Table 4), suggesting that repeatability is very low, although low replication within species (only 13 of 57 species were represented in more than one study, and only six species in more than two studies) likely contributes to this result. For the phylogenetic, multilevel meta-analysis of avian dataset 1, the largest component was also at the effect-size level (23.2%) followed by the study level (16.3%), while 10.9% of heterogeneity was due to phylogeny (Table 4).

The meta-regression models (C and D) used trait category as one of four moderators. The effect size estimates for each trait type were comparable to those from the standard random-effects models (compare Fig. 3A with Fig. 3B or C). Again, the other moderators had no significant effect ($b_{[\text{year}]} = 0.005$, 95% credible interval, CI $[-0.036, 0.039]$, $b_{[\text{Experiment-Observation}]} = 0.003 [-0.106, 0.116]$, $b_{[\text{Field-Laboratory}]} = 0.079 [-0.074, 0.231])$.

Egger’s regression test did not indicate any funnel asymmetry based on meta-analytic residuals from the multilevel meta-regression models (simple: $t_{193} = 1.036, P = 0.302$; phylogenetic: $t_{193} = 0.994, P = 0.321$; see funnel plots, Fig. 4). However, the trim-and-fill tests indicated that eight ($\pm 8.6$ SE; multilevel) and five ($\pm 8.4$ SE; phylogenetic) effect sizes were missing, and that we might have overestimated the simple and phylogenetic model means by 0.0093 and 0.0054, respectively. Shifting the CIs for ‘avian dataset 1’ by these values would result in a credible interval that included zero for the phylogenetic meta-analysis, but not for the simple multilevel meta-analysis. Given this result, and the importance of controlling for phylogeny, our analyses did not provide robust evidence for an offspring sex-ratio bias related to male attractiveness.

### IV. DISCUSSION

#### (1) General findings

We used meta-analysis to test whether females mated to attractive males have a more male-biased offspring sex ratio (Burley, 1981). There was a small but statistically significant relationship in the predicted direction. However, male attractiveness accounted for less than 1% of the variation in offspring sex ratios based on estimates from either standard random-effects meta-analysis or...
multilevel meta-analysis. The estimated mean effect sizes were $r = 0.096$ and 0.064, respectively.

Similarly, when we divided the data into different trait categories we found small, but significant, effects of male attractiveness on the sex ratio for two trait types: male ornamentation and male body size. The means for all seven traits were positive, but there were no significant effects for male behavioural displays, body condition, age, extra-pair status or female preference score. Studies measuring ornaments and body size were the most common (these categories contained 37 and 30 studies, respectively, for dataset 1), while there were relatively few studies for any of the other five trait types (6–16 studies; Table 2). The absence of a significant effect for these five trait types might reflect low statistical power. This is plausible given the estimated magnitude of the effects: no matter how the data were analysed, and even when significant, the mean $r$ was always $< 0.2$. Alternatively, the lack of significantly positive effect size estimates for some traits might reflect a true null relationship, and/or that unmeasured variation obscured a true relationship. In general though, it is clear that the offspring sex ratio is, on average, only weakly associated with various measures of male attractiveness.

The significant mean effect size observed when combining data from all trait types could be seen as strong evidence for the hypothesis, first proposed by Burley (1981), that there is facultative maternal SRA in response to mate attractiveness/quality. It is, however, prudent to consider the robustness of meta-analytic results to narrow-sense publication bias (i.e. unpublished studies) and/or selective reporting of data (i.e. failure to report non-significant results, or to do so in a way that hinders the ability to calculate effect sizes; see Cassey et al., 2004; review: Jennions et al., 2013). We used increasingly conservative data sets (datasets 2 to 4) to include unknown effect sizes in our analyses, treating them as neutral or opposite to the test hypothesis. Depending on the analysis, the significance of the mean effect tended to vary in its sensitivity to the inclusion of these effects. Specifically, for the standard random effects meta-analysis the estimated mean effect stayed significantly positive for all traits combined across datasets 2, 3 and 4; for ornaments across datasets 2 and 3; and for body size
across datasets 2 and 3. For the simple multi-level analysis the estimated mean effect remained
significantly positive for all traits combined across avian datasets 1–3.
Crucially, the mean effect was not significant for avian dataset 2 when phylogenetic relatedness
between species was taken into account (multilevel model B). This analysis is arguably the most
important we present. Although it is based only on bird studies, these represent 92% of the available
effect sizes. It is appropriate to take phylogeny into account when conducting a meta-analysis that
involves reasonably closely related taxa (Chamberlain et al., 2012). There is therefore justifiable
concern about the robustness of the main conclusion to selective reporting. In addition, a publication
bias test for ‘missing’ studies indicated a strong asymmetry in the funnel plot for all traits. If the
putative missing studies were included, the estimated mean effect size was only $r = 0.034$, and did
not differ significantly from zero. Again, this suggests that the finding of a significant mean effect
size is sensitive to narrow-sense publication bias.
Another perspective is that our most important analysis is that based on ornamental male traits. A
strict interpretation of the mate attractiveness hypothesis relies on the existence of paternal traits
whose heritable benefits are partly male-limited (which is most obviously applicable to ornaments).
The initial ‘dataset 1’ analysis indicated a significant effect of ornaments on SRA. At face value, this
might reflect the use of traits that are likely to be under the strongest sex-specific selection of any of
our categories. A stronger sex bias in the heritable benefits of male traits theoretically leads to
stronger SRA results: for example, SRA in response to variation in Fisherian traits (given mutation
bias) is predicted to result in more strongly biased offspring sex ratios than is SRA in response to
viability indicator traits signalling genes that tend to benefit both sexes (Pen & Weissing, 2000;
Fawcett et al., 2007). Again, however, there is cause for concern about publication bias. The
significant effect for ornaments was robust to selective reporting (i.e. datasets 2 and 3 still yielded a
significant mean effect size), but not to narrow-sense publication bias. A trim-and-fill analysis
identified seven ‘missing’ studies that, when imputed, yielded a non-significant mean effect size
estimate of $r = 0.029$. 
The only other trait type with a significant mean effect size was male body size. Why this might be the case is unclear. In birds, which comprise most of our data, there is little evidence that females actively prefer larger males during mate choice. It is noteworthy, however, that body size frequently correlates with major components of male fitness (e.g. success when competing for mates and/or resources, survival, and EPP) in birds (e.g. Kölliker et al., 1999; Kissner, Weatherhead & Francis, 2003). More generally, body size often predicts male attractiveness and mating success in other taxa (Fairbairn, Blanckenhorn & Székely, 2007). If sons benefit more than daughters from increased body size then, whether or not females prefer larger males, it could still be adaptive for females to engage in SRA in response to the size of their mate. This is a subtly different explanation from one based on female assessment of mate attractiveness, but only if females do not prefer larger males. It seems possible that the role of male size in attractiveness has been underestimated in birds – or should at least be tested for experimentally. Interestingly, the results for body size are robust to publication bias. The significant effect was not sensitive to selective reporting (i.e. datasets 2 and 3 still yielded a significant mean effect size). It was also robust to narrow-sense publication bias. A trim-and-fill analysis identified seven ‘missing’ studies but their putative values were positive: when imputed, the mean effect size was $r = 0.151 (P < 0.001)$.

(2) Reasons for a small effect: sex allocation decisions are influenced by many factors

There were moderate levels of among-study variation not attributable to sampling error (i.e. heterogeneity) in our data sets (Table 2). This indicates that fairly large differences among studies were unaccounted for. None of our three main moderator variables (year of publication, laboratory versus field studies, or observational versus experimental studies) had a significant influence on the relationship between male attractiveness and the offspring sex ratio. This was true for both the standard random-effects model and the multilevel analyses. The inclusion of phylogenetic information did little to explain heterogeneity, which remained around 50% (Table 4). The unexplained variance among studies highlights the difficulty of testing specific hypotheses for SRA.
Many selective factors have been suggested to affect SRA and, for some of them, meta-analyses have formally shown that they explain a significant amount of variation in offspring sex ratios (review: West, 2009) (Table 5). These putative forces of selection on SRA are not mutually exclusive; nor, more importantly here, do they act independently of any influence of mate attractiveness on SRA. If multiple sources of selection on SRA interact they could create extremely complex outcomes. This makes it difficult to predict the optimal offspring sex ratio. For example, the benefits of SRA based on mate attractiveness appear to be seasonally dependent (Griffith et al., 2003; Korsten et al., 2006; Delhey et al., 2007), context-dependent (Pryke et al., 2011), or only reliably indicated by specific age classes of males (Taff et al., 2011).

There are numerous other obstacles to testing SRA hypotheses directly. One is the ubiquitous problem of inferring causality from observational studies where there are potential confounding factors that actually generate the observed pattern. For example, assortative mating such that females in good condition tend to mate with more attractive males seems likely in many seasonally breeding bird species. If this occurs, then a male-biased offspring sex ratio might be partly (or entirely) driven by maternal condition (i.e. Trivers & Willard, 1973), rather than by male-biased genetic benefits. Similarly, males might vary in their ability to provide parental care, which could affect SRA (e.g. Bowers et al., 2013). If parental ability is correlated with male attractiveness and sons and daughters differ in the benefits of increased investment, this could lead to adaptive SRA that is correlated with, but not causally related to, attractiveness. Some observational studies in our data set do, however, try to control statistically for confounding maternal and environmental variables (e.g. Addison, Kitaysky & Hipfner, 2008; Taff et al., 2011).

Another potentially common problem arises if the benefits of SRA depend on a male–female interaction. For instance, one of the clearest examples of adaptive SRA – mainly because it involves several different experimental manipulations of male colouration – is in Gouldian finches (*Erythrura gouldiae*). In this species there is severe genetic incompatibility between red and black colour morphs, which has a more detrimental effect on daughters than sons. Females express strong mating
preferences for males of the same colour morph. When paired with a male bearing the phenotype of the other morph they exhibit adaptive SRA towards sons (Pryke & Griffith, 2009; Pryke et al., 2011). This species was excluded from our meta-analysis because we only investigated directional female preferences for male traits (i.e. we excluded individual-specific preferences). A type of male–female interaction that might arise in many species and select for SRA based on a non-directional preference is associated with inbreeding. There is evidence that pair relatedness can affect adaptive SRA (e.g. Sardell & DuVal, 2014). If widespread, this could confound patterns of SRA predicted solely on a male’s general attractiveness.

Even when genetic incompatibility is absent, SRA might still depend on both female and male qualities. Studies of zebra finches (Taeniopygia guttata) suggest that assortative mating accounts for the strong sex ratio biases based on artificial leg colour bands originally detected by Burley (Rutstein et al., 2005). If higher quality females can better bear the costs of producing the more attractive/expensive sex (Katsuki et al., 2012), assortative mating between high-quality females and attractive males could amplify adaptive SRA based solely on male identity. One caveat to this is that the influence of sexually antagonistic genes on adaptive SRA has been poorly explored (but see Calsbeek & Sinervo, 2004; Alonzo & Sinervo, 2007; Blackburn, Albert & Otto, 2010). Given sexual antagonism, it is plausible that assortative mating might eliminate SRA. If genes affecting female quality have more beneficial effects on daughters’ fitness (e.g. Leimar, 1996), while genes that affect male attractiveness more strongly elevate sons’ fitness (Pischedda & Chippindale, 2006), then assortative mating could lead to sex-specific maternal and paternal effects on offspring that nullify the advantages of SRA. If the relative fitness of offspring of each sex remains similar due to ‘symmetrical’ effects of maternal and paternal genes on offspring of the opposite sex (but see Connallon & Jakubowski, 2007; Cox & Calsbeek, 2010), a more equitable sex ratio is predicted. This would occur even if attractive males sire sons of above-average fitness when randomly assigned a mate (i.e. even if there is still a general male-biased benefit of sire attractiveness).
Finally, in addition to SRA, there are other, non-mutually exclusive sex allocation responses to the sources of selection described in Table 5, including mate attractiveness. SRA is not the only form of adaptive sex allocation based on changes in investment patterns. There is no evidence in birds that mothers can detect the sex of an embryo and then differentially invest into eggs on this basis (meta-analysis: Rutkowska, Dubiec & Nakagawa, 2014). Mothers can, however, differentially invest into total offspring production (i.e. produce larger broods and/or bigger eggs when mated to an attractive male; Burley, 1986; review: Sheldon, 2000; meta-analysis: Horváthová et al., 2012) and differentially allocate resources towards one sex without any change in the sex ratio or number of offspring per breeding attempt (e.g. Saino et al., 2003). These varied strategic responses are forms of sex allocation that are not always distinguished in the literature, either from each other or from SRA (Carranza, 2002).

(3) Reasons for a small effect: measuring sexual selection is tricky

Although there are clear predictions for adaptive SRA based on male attractiveness (West & Sheldon, 2002), it is often unclear (a) what traits actually determine male attractiveness or quality, and (b) if these traits actually affect offspring fitness through male-biased genetic benefits. Identifying targets of mate choice is a perennial problem in comparative analyses or meta-analyses of sexual selection. What makes a male attractive? Many researchers measure several different male traits that are putatively preferred by females. Thus, even within a single empirical study, there can be several reported relationships between ‘attractiveness’ and the offspring sex ratio (e.g. Griffith et al., 2003; Bowers et al., 2013). This could be statistically problematic if several male traits are included in a multiple regression as potential predictors of the offspring sex ratio. This was often the case for studies in our data set: many of our effect sizes were taken from generalized linear models (GLMs) or generalized-linear mixed-effects models (GLMMs) with several predictors (exclusion of such data would have halved our sample size). The resultant effect size estimates are therefore the effect of the focal trait controlling for its relationships with other measured traits, and for the
relationship between those traits and the offspring sex ratio (i.e. we reported partial correlations). For a review of the potential statistical biases that occur when using partial rather than bivariate correlations see Aloe (2015), and references therein. The more strongly the measured traits are correlated the greater the discrepancy between partial and bivariate relationships. In general, however, collinearity between measured traits is modest (otherwise they would not be included as independent predictors in a model). A lack of correlation among measured traits does, however, highlight the likelihood that some of them might contribute little if anything to net attractiveness. More generally, female use of multiple cues for mate choice (review: Candolin, 2003) means that effect sizes for individual traits imperfectly capture the true relationship being tested: that between net male attractiveness and the offspring sex ratio.

Interpretation of what constitutes an attractive male trait has also expanded since Burley’s original hypothesis was formulated. For example, it has been argued that whether a male is a within-pair or extra-pair mate is informative about his relative attractiveness to females (reviews: Westneat & Stewart, 2003; but see Forstmeier et al., 2014). In addition, traits used by females for mate choice are not necessarily those that directly influence adaptive SRA, even if there are male-biased genetic benefits. For example, male attractiveness and competiveness are often imperfectly correlated (see Kelly, 2008), or even negatively correlated (Hunt et al., 2009). Even if researchers correctly identify heritable traits that determine male attractiveness, these are not necessarily the ones that females assess for the largest potential SRA fitness gains (e.g. if the marginal fitness gains from being more ‘dominant’ during fights for territories outweigh those of being more attractive to females when holding a territory, and both attractiveness and dominance are heritable).

Determining the relationship between fitness and attractiveness is a major challenge for two reasons. First, it is difficult to measure fitness, and it is often simply assumed that attractiveness predicts net fitness. But do males bearing attractive traits and experiencing higher mating rates actually have greater net fitness than less attractive males? Studies often report an association between male attractiveness/quality and an increase in some components of reproductive success (e.g. number of
mates in a single breeding season) without testing whether natural selection (Hine, McGuigan & Blows, 2011) or even other forms of sexual selection (e.g. sperm competitiveness; Lüpold et al., 2014) negate the sexually selected benefits of traits that elevate attractiveness. Second, the mate attractiveness hypothesis for adaptive SRA assumes that male-derived fitness benefits exist, and are to some extent sex-limited. In the context of genetic benefits this basically means that male attractiveness is heritable. Many studies have reported significant additive genetic variation for male sexual traits (Prokuda & Roff, 2014), but it remains unclear if net attractiveness (and, ideally, net fitness) is heritable. Genetic covariance among traits affecting fitness can result in the absence of genetic variation in the direction of selection on net attractiveness and, ultimately, fitness (Walsh & Blows, 2009).

(4) Updated results and future improvements

Our updated meta-analyses indicate a substantially weaker relationship between measures of mate attractiveness and offspring sex ratios than the \( r = 0.205 \) reported by West & Sheldon (2002). A decline in effect size over time is a common phenomenon in many research fields, including ecology and evolution (Jennions & Møller, 2002). There are many explanations for a temporal decline, including an early publication bias against weaker effect sizes (i.e. non-significant results) and the inclusion of less appropriate study systems over time (review: Koricheva, Jennions & Lau, 2013). In general, we argue that the observed small effect sizes are expected for many reasons. Six factors that we discussed are: (a) poor control for other selective factors that affect SRA (see Table 5); (b) other forms of conditional sex allocation (e.g. sex-biased provisioning); (c) assortative mating by mate quality; (d) difficulty identifying preferred male traits that directly influence sex allocation – ideally researchers should use measures of net male attractiveness; (e) uncertainty that attractiveness increases net male fitness; (f) uncertainty that the benefits of mate choice are genetic and greater for sons than daughters. In addition, there are at least four other factors that might lower effect sizes: (g) high measurement error due to logistic challenges (e.g. measuring behavioural traits accurately); (h)
the inclusion of inappropriate study species (e.g. species with high fecundity are less likely to skew the sex ratio; for a review and explanation see Frank, 1990); (i) difficulty in obtaining appropriate effect size estimates – we had to use data from multiple regression models that included several prospective measures of sire attractiveness because simple bivariate correlations were rarely available; and (j) stochastic noise (e.g. female perceptual errors when assessing males, and constraints on sex-biasing mechanisms).

Aside from considering the problems described above, future studies should broaden their taxonomic scope. Approximately 84% of the studies in our meta-analysis were on birds. This limits the generality of our conclusions, and evaluation of the wider relevance of the SRA mate attractiveness hypothesis. In general, specific sex allocation hypotheses have been tested with a focus on taxa that seem especially amenable (e.g. local mate competition in fig wasps, or helper production in birds and mammals). The disproportionate use of birds to test the SRA mate attractiveness hypothesis might be due to high sexual dimorphism and extravagance of male ornamental traits in many species that suggest strong sexual selection through female choice. More importantly, recent theory suggests that SRA due to mate attractiveness is least likely to occur when there is very strong sexual selection (Fawcett et al., 2011; Booksmythe et al., 2013). Finally, many avian life histories do not conform to key assumptions of mathematical models of sex allocation/adjustment, such as non-overlapping generations, semelparity (or a fixed total investment at every reproductive event), a short and discrete bout of parental investment and investment (potential for control) by one parent only, and no cost of adjustment (review: Komdeur & Pen, 2002). Although some of these limitations will apply to non-avian taxa as well, researchers might wish to reconsider their choice of study species and look to taxa other than birds to test the SRA mate attractiveness hypothesis.

During our attempts to extract effect sizes from papers it became obvious that many studies are guilty of post hoc analyses, and fail to clearly specify study goals prior to analysis. Many statistical tests are better seen as exploratory, rather than acting to test a hypothesis with the stated likelihood of type I error (for an approachable case study see Simmons, Nelson & Simonsohn, 2011). There was often an
excess of male traits being examined, the inclusion of covariates seemed arbitrary, and the actual statistical test used (e.g. GLM or contingency table) and whether data were pooled or analysed at the brood level was often not justified. All these problems increase the danger of P-hacking (selective collection, manipulation, and/or reporting of data to achieve statistical significance) and erroneous reporting of significant results, which inflates effect size estimates (see Simonsohn, Nelson & Simmons, 2014; Head et al. 2015). They also made it difficult to locate the information needed to extract effect sizes. The SRA hypothesis is straightforward so it is reasonable to expect greater consistency in how researchers test for a relationship between male attractiveness and offspring sex ratio. We mention this not to criticise colleagues (we are guilty of the same ‘misconduct’ sensu Simmons et al., 2011, in some of our own papers), but simply to highlight wider calls by many others (e.g. Garamszegi et al., 2009) for the need to improve standards of reporting in evolutionary ecology studies. It is very possible, for example, that the conclusions of our meta-analysis might have differed if researchers had always used composite measures of male attractiveness (rather than analysing multiple traits) or integrated fitness measures such as female responses, had only controlled for covariates with well-supported biological relevance to their study system and for which they could provide evidence of effects on sex ratios (e.g. level of local mate competition in parasitoids), and had fully reported the direction and magnitude of all tests, regardless of whether or not the effect was significant.

Finally, we should note that there is at least one species where well-replicated experimental studies confirm that mothers adjust the offspring sex ratio in response to male phenotype: head colour in Gouldian finches (Pryke & Griffith, 2009; Pryke et al., 2011). As such, there is evidence that SRA in response to male attractiveness has evolved. In our meta-analysis, however, the small mean effect size and its sensitivity to narrow-sense publication bias (the ‘file drawer’ problem, where a greater proportion of non-significant than significant results go unpublished; Rosenthal, 1979; Jennions et al., 2013) and/or selective reporting, indicates little support for the hypothesis that such SRA is widespread. Ultimately, the strongest evidence has (and will continue to) come from experiments that
involve direct manipulation of male attractiveness and/or random assignment of males to females. To date there are only a handful of appropriate studies (e.g. Burley, 1981, 1986; Sheldon et al., 1999; Pike & Petrie, 2005; Ferree, 2007). In species where it is feasible, the study design of Sato & Karino (2010), whereby a female’s social environment is manipulated so that the same male is assessed to be either relatively attractive or unattractive, seems a particularly powerful approach. Once species in which SRA has been shown robustly to occur are identified, the next step is to test whether this is because sons sired by relatively more attractive males are of higher mean fitness than are daughters. Both lines of evidence are needed to test the SRA mate attractiveness hypothesis fully.

Given the many difficulties with the interpretation of observational studies (i.e. inability to predict the sex ratio with certainty due to potentially confounding factors) we suggest that the most fruitful way forward is to conduct experiments that directly rank males according to their mean attractiveness to females, and then randomly assign females a mate of low or high attractiveness. If these experiments are conducted on species where there is little evidence for direct material assistance by males (i.e. no male parental care or mate provisioning) it should be possible to test the hypothesis that genetic benefits that accrue more strongly to sons than daughters can drive greater sex allocation to males.

V. CONCLUSIONS

(1) The mean effect size for the relationship between male attractiveness and offspring sex ratio based on all trait types was small ($r < 0.10$) but significant. Assuming no publication bias this result was consistent using standard random-effects meta-analysis or multilevel meta-analysis models (including a control for phylogenetic relatedness in a data set restricted to birds).

(2) These significant mean effect size estimates from all traits combined were, however, sensitive to publication bias – due to a possible ‘file drawer’ effect and/or selective reporting of significant results within studies. Correction for possible bias reduced effect size estimates to non-significant values.
(3) The mean effect sizes for seven specific male trait types were significant only for ornamental male traits and measures of male body size. The mean effect remained significant for body size even after correcting for potential publication bias.

(4) There was moderate heterogeneity in effect size estimates. The unexplained variation between studies indicates that there are moderating variables that have yet not been identified. We detected no effect of our three test moderators: year of publication, experimental versus observational study, and laboratory versus field study. There was a very weak phylogenetic signal for birds (i.e. related species did not have similar effect sizes).

(5) There was a strong taxonomic bias in our data set. Most studies were on bird species (~ 84%).

Given the lack of robust evidence in birds, future studies on other groups (invertebrates, fish, reptiles) are needed to shed light on whether the mate attractiveness SRA hypothesis will have any greater support in other taxa.

(6) In sum, our results provide very weak support for the hypothesis that females facultatively adjust the sex ratio of their offspring in response to the attractiveness of their mate.

VI. ACKNOWLEDGEMENTS

We thank the following authors who responded to email requests and, where possible, sent data: E. Adkins-Regan, J. Aparicio, R. Bradbury, S. Correa, K. Delhey, A. Dreiss, D. Dowling, P. Dunn, J. Graves, D. Green, L. Gustafsson, M. Hjernquist, A. Husby, P. Korsten, K. Kraaijeveld, D. Leech, S. Leitner, C. Lessells, B. MacDougall-Shackleton, N. Perlut, A. Peters, J. Potti, A. Roulin, T. Schmoll, M. Soma, T. Szekely, C. Taff, D. Westneat, N. Yamaguchi. (Apologies to anyone we have inadvertently omitted.) Funding was provided by the Australian Research Council (M.D.J., J.D., B.M.), European Research Council grant AdG-294333 to G. Arnqvist (I.B.), the Rutherford Discovery Fellowship, New Zealand (S.N.), and an Australian Research Council Future Fellowship (S.N.).
VII. REFERENCES

Studies marked with * contributed effect sizes to the meta-analyses; studies marked with † are key papers from which forward searches were conducted.


Offspring sex ratio is related to male body size in the great tit (Parus major). Behavioral Ecology 10, 68-72.


SARDELL, R.J. & DUVAL, E.H. (2014). Differential allocation in a lekking bird: females lay larger eggs and are more likely to have male chicks when they mate with less related males. *Proceedings of the Royal Society of London Series B* 281, 20132386.


VIII. SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article, DOI: 10.1111/brv.12220

**Fig. S1.** The phylogenetic trees used for the phylogenetic meta-analysis and meta-regression analysis: (1) a tree with Ericson backbone and (2) a tree with Hackett backbone.

**Fig. S2.** Forest plot of the regression coefficients of the phylogenetic meta-regression with the Hackett tree.

**Table S1.** Results of phylogenetic (random-effects) meta-analyses using the Hackett tree.

**Table S2.** Results of multilevel and phylogenetic (random-effects) meta-analyses, incorporating correlations among effect sizes that come from the same set of individuals.

**Table S3.** Summary of the 25 studies excluded on full-text screening.

**Appendix S1.** Literature search and selection procedure.

**Appendix S2.** EndNote library (SexRatioAdjustment.enlp): all unique records obtained from the Web of Science literature searches outlined in Appendix S1.

**Appendix S3.** Raw data file (SexRatioAdjustmentData.xlsx): all effect sizes included in the meta-analyses.
Figures

Fig. 1. Papers published per year since Burley (1981) on adjustment of offspring sex ratios in response to paternal traits. Only papers eligible for inclusion in our meta-analyses are shown; four included papers from 2014 are not shown as our search only encompassed half of this year.
Fig. 2. Preferred reporting items for systematic reviews and meta analyses (PRISMA) flow chart of our literature search and data collection process (see http://www.prisma-statement.org/).

Records identified through database searching
\((N = 5319)\)

Records after duplicates removed
\((N = 4474)\)

Records screened
\((N = 4474)\)

Records excluded
\((N = 4356)\)

Full-text articles assessed for eligibility
\((N = 118)\)

Studies included in quantitative synthesis (meta-analysis)
\((N = 93)\)

Full-text articles excluded \((N = 25)\) – see Table S3 for additional detail:
- did not measure sire traits \(n = 3\);
- non-comparable measure of ‘attractiveness’ \(N = 6\);
- no direct test of male trait effect on offspring sex ratio \(N = 7\);
- did not measure offspring sex ratio \(N = 4\);
- analysis at population level \(N = 4\);
- used data from another included paper \(N = 1\)
Fig. 3. Forest plots of (A) the meta-analytic means of the eight random-effects meta-analyses, the multilevel meta-analysis, and the phylogenetic meta-analysis (the latter two used the avian-only data set; the phylogenetic analysis used the Ericson backbone tree), (B) the regression coefficients of the multilevel meta-regression, and (C) of the phylogenetic meta-regression. EPP, extra-pair paternity. 

(A) Mean estimates of different meta-analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>k</th>
<th>Effect size (Z_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>EPP</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Ornament</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Multilevel (avian data only)</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>Phylogenetic (avian data only)</td>
<td>195</td>
<td></td>
</tr>
</tbody>
</table>

(B) Multilevel regression (avian data only)

<table>
<thead>
<tr>
<th>Variable</th>
<th>k</th>
<th>Effect size (Z_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>EPP</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Ornament</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>36–159</td>
<td></td>
</tr>
<tr>
<td>Multilevel (avian data only)</td>
<td>170–25</td>
<td></td>
</tr>
<tr>
<td>Phylogenetic (avian data only)</td>
<td>195</td>
<td></td>
</tr>
</tbody>
</table>

(C) Phylogenetic regression (avian data only)

<table>
<thead>
<tr>
<th>Variable</th>
<th>k</th>
<th>Effect size (Z_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>EPP</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Ornament</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>36–159</td>
<td></td>
</tr>
<tr>
<td>Multilevel (avian data only)</td>
<td>170–25</td>
<td></td>
</tr>
<tr>
<td>Phylogenetic (avian data only)</td>
<td>195</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 4. Funnel plots (effect sizes plotted against their corresponding precision, the inverse of the standard error) of (A–H) the eight random-effects meta-analyses, (I) the multilevel meta-analysis and (J) the phylogenetic multilevel meta-analysis (Ericson backbone tree). Plots for the multilevel analyses (I, J) use the meta-analytic residuals (see Nakagawa & Santos, 2012); these analyses used the avian-only data set. EPP, extra-pair paternity.
Table 1. Summary of 93 papers coded for the meta-analyses. $N$ studies reports the number of independent populations/experiments for which the relationship between brood sex ratios and some measure of male attractiveness was tested. $N$ broods gives the range of sample sizes for each study; even where a paper contained only one study, different numbers of broods may have contributed to the tests for different measured traits. For EPP (extra-pair paternity): ‘within broods’ compares the sex ratio of offspring sired by a female’s social and extra-pair mate, within the same brood; ‘between broods’ compares the sex ratio of broods with and without any extra-pair offspring; ‘across all broods’ compares the probability of being male for individual offspring sired by a within- or extra-pair male.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Species</th>
<th>$N$ studies</th>
<th>Measures of attractiveness tested</th>
<th>$N$ broods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abroe et al. (2007)</td>
<td>Common yellowthroat, Geothlypis trichas</td>
<td>1</td>
<td>EPP: within broods; Ornament: mask size, bib size, bib colour; Size</td>
<td>69–138</td>
</tr>
<tr>
<td>Addison et al. (2008)</td>
<td>Rhinoceros auklet, Cerorhinca moncerata</td>
<td>1</td>
<td>Size</td>
<td>25</td>
</tr>
<tr>
<td>Aparicio &amp; Cordero (2001)</td>
<td>Lesser kestrel, Falco naumanni</td>
<td>1</td>
<td>Condition; Ornament: wing patch size</td>
<td>75</td>
</tr>
<tr>
<td>Bell et al. (2014)</td>
<td>Blue tit, Cyanistes caeruleus</td>
<td>2</td>
<td>Condition</td>
<td>37, 41</td>
</tr>
<tr>
<td>Benito et al. (2013)</td>
<td>Common tern, Sterna hirundo</td>
<td>1</td>
<td>Behaviour: arrival date; Size</td>
<td>81</td>
</tr>
<tr>
<td>Blanchard et al. (2007)</td>
<td>Wandering albatross, Diomedea exulans</td>
<td>1</td>
<td>Condition; Size</td>
<td>63</td>
</tr>
<tr>
<td>Bowers et al. (2013)</td>
<td>Collared flycatcher, Ficedula albicollis</td>
<td>1</td>
<td>Age; EPP: across all broods; Ornament: wing patch size, forehead patch size</td>
<td>160</td>
</tr>
<tr>
<td>Bradbury et al. (1997)</td>
<td>European starling, Sturnus vulgaris</td>
<td>1</td>
<td>Age; Size</td>
<td>108</td>
</tr>
<tr>
<td>Burley (1981)</td>
<td>Zebra finch, Taeniopygia guttata</td>
<td>1</td>
<td>Ornament: leg band colour</td>
<td>20</td>
</tr>
<tr>
<td>Burley (1986)</td>
<td>Zebra finch, Taeniopygia guttata</td>
<td>1</td>
<td>Ornament: leg band colour</td>
<td>18</td>
</tr>
<tr>
<td>Calsbeek &amp; Sinervo (2004)</td>
<td>Side-blotched lizard, Uta stansburiana</td>
<td>2</td>
<td>Size</td>
<td>33, 41</td>
</tr>
<tr>
<td>Cockburn &amp; Double (2008)</td>
<td>Superb fairy wren, Malurus cyaneus</td>
<td>1</td>
<td>Condition; EPP: across all broods</td>
<td>993</td>
</tr>
<tr>
<td>Correa et al. (2011)</td>
<td>Japanese quail, Coturnix japonica</td>
<td>1</td>
<td>Behaviour: strutting, neck holding, mounting</td>
<td>26</td>
</tr>
<tr>
<td>Cotton et al. (2014)</td>
<td>Stalk-eyed fly, Teldeopsis dalmanni</td>
<td>1</td>
<td>Ornament: eyespan, relative eyespan; Size</td>
<td>130</td>
</tr>
<tr>
<td>Cox et al. (2011)</td>
<td>Brown anole, Anolis sagrei</td>
<td>1</td>
<td>Condition</td>
<td>67</td>
</tr>
<tr>
<td>Cunningham &amp; Russell (2001)</td>
<td>Mallard, Anas platyrhynchos</td>
<td>1</td>
<td>Female preference</td>
<td>32</td>
</tr>
<tr>
<td>Debiak et al. (2014)</td>
<td>Common loon, Gaviaius immer</td>
<td>1</td>
<td>Size</td>
<td>14</td>
</tr>
<tr>
<td>Delhey et al. (2007)</td>
<td>Blue tit, Cyanistes caeruleus</td>
<td>4</td>
<td>Ornament: crown brightness, chroma, hue, UV chroma; Size</td>
<td>20, 22, 40, 70</td>
</tr>
<tr>
<td>Delmore et al. (2008)</td>
<td>Tree swallow, Tachycineta bicolor</td>
<td>1</td>
<td>Age; Condition; EPP: between broods; Size</td>
<td>50</td>
</tr>
<tr>
<td>Dietrich-Bischoff et al. (2006)</td>
<td>Coal tit, Parus ater</td>
<td>1</td>
<td>Age; EPP: within broods, between broods; Ornament: bib saturation, bib size; Size</td>
<td>235–457</td>
</tr>
<tr>
<td>Dolan et al. (2009)</td>
<td>Eastern kingbird, Tyrannus tyrannus</td>
<td>1</td>
<td>Behaviour: song latency, song rate; EPP: between broods; Size</td>
<td>42–72</td>
</tr>
<tr>
<td>Dowling &amp; Mulder (2006)</td>
<td>Red-capped robin, Petroica goodenovii</td>
<td>1</td>
<td>EPP: within broods; Ornament: breast &amp; forehead patch brightness, hue, chroma; Size</td>
<td>44</td>
</tr>
<tr>
<td>Dreiss et al. (2006)</td>
<td>Blue tit, Cyanistes caeruleus</td>
<td>1</td>
<td>Behaviour: song characteristics; EPP: within broods; Ornament: crown brightness, hue, UV chroma; Size</td>
<td>84</td>
</tr>
<tr>
<td>Drullion &amp; Dubois (2011)</td>
<td>Zebra finch, Taeniopygia guttata</td>
<td>1</td>
<td>Female preference</td>
<td>6</td>
</tr>
<tr>
<td>Reference</td>
<td>Species</td>
<td>N studies</td>
<td>Measures of attractiveness tested</td>
<td>N broods</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------</td>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Ellegren et al. (1996)</td>
<td>Collared flycatcher, <em>Ficedula albicollis</em></td>
<td>1</td>
<td>Ornament: forehead patch size; Size</td>
<td>79</td>
</tr>
<tr>
<td>Griffith et al. (2003)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>3</td>
<td>Age; Ornament: plumage brightness, UV chroma</td>
<td>15–16, 22–23, 50–57</td>
</tr>
<tr>
<td>Grindstaff et al. (2001)</td>
<td>Dark-eyed junco, <em>Junco hyemalis</em></td>
<td>1</td>
<td>Female preference</td>
<td>48</td>
</tr>
<tr>
<td>Hjernquist et al. (2009)</td>
<td>Collared flycatcher, <em>Ficedula albicollis</em></td>
<td>1</td>
<td>Condition; Ornament: forehead patch size</td>
<td>74</td>
</tr>
<tr>
<td>House et al. (2011)</td>
<td>Dung beetle, <em>Onthophagus taurus</em></td>
<td>1</td>
<td>Size</td>
<td>50</td>
</tr>
<tr>
<td>Husby et al. (2006)</td>
<td>House sparrow, <em>Passer domesticus</em></td>
<td>1</td>
<td>Age; Ornament: badge size</td>
<td>15</td>
</tr>
<tr>
<td>Johnson et al. (2009)</td>
<td>House wren, <em>Troglydtes aedon</em></td>
<td>1</td>
<td>EPP: within broods, between broods</td>
<td>476</td>
</tr>
<tr>
<td>Katsuki et al. (2012)</td>
<td>Broad-horned flour beetle, <em>Gnathocerus cornutus</em></td>
<td>1</td>
<td>Ornament: mandible length</td>
<td>50</td>
</tr>
<tr>
<td>Kempenaers et al. (1997)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>1</td>
<td>EPP: within broods, between broods</td>
<td>57–141</td>
</tr>
<tr>
<td>Köllicker et al. (1999)</td>
<td>Great tit, <em>Parus major</em></td>
<td>1</td>
<td>Age; Ornament: breast stripe size; Size</td>
<td>57</td>
</tr>
<tr>
<td>Korpimäki et al. (2000)</td>
<td>Kestrel, <em>Falco tinnunculus</em></td>
<td>1</td>
<td>Condition; Size</td>
<td>80</td>
</tr>
<tr>
<td>Korsten et al. (2006)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>2</td>
<td>Ornament: UV chroma</td>
<td>26, 44</td>
</tr>
<tr>
<td>Kraaijeveld et al. (2007)</td>
<td>Black swan, <em>Cygnus atratus</em></td>
<td>1</td>
<td>EPP: across all broods; Ornament: number of curled wing feathers; Size</td>
<td>36</td>
</tr>
<tr>
<td>Lee &amp; Hatchwell (2011)</td>
<td>Vinous-throated parrotbill, <em>Proactomis webbianus</em></td>
<td>1</td>
<td>EPP: between broods; Size</td>
<td>36</td>
</tr>
<tr>
<td>Leech et al. (2001)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>1</td>
<td>Age; Condition; EPP: within broods; Size</td>
<td>103–154</td>
</tr>
<tr>
<td>Lessells et al. (1996)</td>
<td>Great tit, <em>Parus major</em></td>
<td>1</td>
<td>Age; Size</td>
<td>52–92</td>
</tr>
<tr>
<td>Limbourg et al. (2013)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>1</td>
<td>Age; Ornament: UV chroma</td>
<td>33</td>
</tr>
<tr>
<td>Magrath et al. (2002)</td>
<td>Fairy martin, <em>Petrochelidon ariel</em></td>
<td>1</td>
<td>Age; Condition; EPP: between broods; Size</td>
<td>29–56</td>
</tr>
<tr>
<td>Neto et al. (2011)</td>
<td>Savi’s warbler, <em>Locustella luscinoides</em></td>
<td>1</td>
<td>EPP: between broods; Size</td>
<td>102</td>
</tr>
<tr>
<td>Oddie &amp; Reim (2002)</td>
<td>Great tit, <em>Parus major</em></td>
<td>1</td>
<td>Age; Condition; Size</td>
<td>143–173</td>
</tr>
<tr>
<td>Olsson et al. (2005b)</td>
<td>Sand lizard, <em>Lacerta agilis</em></td>
<td>1</td>
<td>Ornament: badge size</td>
<td>58</td>
</tr>
<tr>
<td>Pariset et al. (2012)</td>
<td>Zebra finch, <em>Taeniopygia guttata</em></td>
<td>1</td>
<td>Ornament: leg band colour</td>
<td>21</td>
</tr>
<tr>
<td>Parker (2005)</td>
<td>Red junglefowl (chicken), <em>Gallus gallus</em></td>
<td>1</td>
<td>Ornament: comb size</td>
<td>76</td>
</tr>
<tr>
<td>Reference</td>
<td>Species</td>
<td>N studies</td>
<td>Measures of attractiveness tested</td>
<td>N broods</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Perlut et al. (2014)</td>
<td>Savannah sparrow, <em>Passerculus sandwichensis</em></td>
<td>1</td>
<td>EPP: between broods; Size</td>
<td>26–48</td>
</tr>
<tr>
<td>Polo et al. (2004)</td>
<td>Spotless starling, <em>Sturnus unicolor</em></td>
<td>1</td>
<td>Behaviour: nest decoration</td>
<td>149</td>
</tr>
<tr>
<td>Questiau et al. (2000)</td>
<td>Bluethroat, <em>Luscinia svecica</em></td>
<td>1</td>
<td>Age; Ornament: throat patch size; Size</td>
<td>27–39</td>
</tr>
<tr>
<td>Radford &amp; Blakey (2006)</td>
<td>Great tit, <em>Parus major</em></td>
<td>1</td>
<td>Age; Condition; Ornament: breast stripe size; Size</td>
<td>84</td>
</tr>
<tr>
<td>Rosivall et al. (2004)</td>
<td>Collared flycatcher, <em>Ficedula albicollis</em></td>
<td>1</td>
<td>Age; Ornament: forehead patch size, wing patch size; Size</td>
<td>54–57</td>
</tr>
<tr>
<td>Roulin et al. (2010)</td>
<td>Barn owl, <em>Tyto alba</em></td>
<td>1</td>
<td>Ornament: spot size</td>
<td>44</td>
</tr>
<tr>
<td>Rutstein et al. (2005)</td>
<td>Zebra finch, <em>Taeniopygia guttata</em></td>
<td>1</td>
<td>Ornament: leg band colour</td>
<td>101</td>
</tr>
<tr>
<td>Saino et al. (1999)</td>
<td>Barn swallow, <em>Hirundo rustica</em></td>
<td>1</td>
<td>Condition; EPP: across all broods; Ornament: tail length; Size</td>
<td>19–54</td>
</tr>
<tr>
<td>Saino et al. (2002)</td>
<td>Barn swallow, <em>Hirundo rustica</em></td>
<td>1</td>
<td>Age; Ornament: tail length</td>
<td>76</td>
</tr>
<tr>
<td>Saino et al. (2003)</td>
<td>Barn swallow, <em>Hirundo rustica</em></td>
<td>1</td>
<td>Ornament: tail length</td>
<td>46</td>
</tr>
<tr>
<td>Sheldon et al. (1999)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>2</td>
<td>Age; Ornament: plumage hue, brightness, UV chroma</td>
<td>20, 57</td>
</tr>
<tr>
<td>Stauss et al. (2005)</td>
<td>(a) Great tit, <em>Parus major</em></td>
<td>2</td>
<td>Condition</td>
<td>(a) 95</td>
</tr>
<tr>
<td></td>
<td>(b) Blue tit, <em>Cyanistes caeruleus</em></td>
<td></td>
<td></td>
<td>(b) 88</td>
</tr>
<tr>
<td>Svensson &amp; Nilsson (1996)</td>
<td>Blue tit, <em>Cyanistes caeruleus</em></td>
<td>1</td>
<td>Age</td>
<td>33</td>
</tr>
<tr>
<td>Szekely et al. (2004)</td>
<td>Kentish plover, <em>Charadrius alexandrinus</em></td>
<td>1</td>
<td>Age; Size</td>
<td>59</td>
</tr>
<tr>
<td>Taff et al. (2011)</td>
<td>Common yellowthroat, <em>Geothlypis trichas</em></td>
<td>2</td>
<td>Age; EPP: within broods; Ornament: plumage yellow &amp; UV brightness, carotenoid chroma, mask size, bib size; Size</td>
<td>27–71, 44</td>
</tr>
<tr>
<td>Watt et al. (2001)</td>
<td>Guppy, <em>Poecilia reticulata</em></td>
<td>1</td>
<td>Ornament: orange spot size; Female preference; Size</td>
<td>17–24</td>
</tr>
<tr>
<td>Westerdahl et al. (1997)</td>
<td>Great reed warbler, <em>Acrocephalus arundinaceus</em></td>
<td>1</td>
<td>Behaviour: song repertoire; EPP: within broods; Size</td>
<td>9–63</td>
</tr>
<tr>
<td>Westneat et al. (2002)</td>
<td>House sparrow, <em>Passer domesticus</em></td>
<td>1</td>
<td>Age; Condition; Ornament: bib size; Size</td>
<td>93–142</td>
</tr>
<tr>
<td>Whittingham &amp; Dunn (2000)</td>
<td>Tree swallow, <em>Tachycineta bicolor</em></td>
<td>1</td>
<td>Condition</td>
<td>40</td>
</tr>
<tr>
<td>Reference</td>
<td>Species</td>
<td>N studies</td>
<td>Measures of attractiveness tested</td>
<td>N broods</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Wilk et al. (2008)</td>
<td>Collared flycatcher, <em>Ficedula albicollis</em></td>
<td>1</td>
<td>EPP: across all broods</td>
<td>78</td>
</tr>
<tr>
<td>Yamaguchi et al. (2004)</td>
<td>Varied tit, <em>Parus varius</em></td>
<td>1</td>
<td>Ornament: forehead &amp; breast patch size; Size</td>
<td>59</td>
</tr>
<tr>
<td>Zann &amp; Runciman (2003)</td>
<td>Zebra finch, <em>Taeniopygia guttata</em></td>
<td>3</td>
<td>Condition; Ornament: breast band size, leg band colour; Size</td>
<td>20, 24, 68</td>
</tr>
<tr>
<td>Zielinska et al. (2010)</td>
<td>House martin, <em>Delichon urbicum</em></td>
<td>1</td>
<td>Condition; Ornament: rump patch colour &amp; size; Size</td>
<td>129</td>
</tr>
</tbody>
</table>
Table 2. Results from the random-effects meta-analyses for the eight different trait categories; $k =$ number of effect sizes, $m =$ number of species and $n =$ number of broods in each data set used for these meta-analyses (for the descriptions of Dataset, $Q$ and $I^2$, see text Section II); EPP, extra-pair paternity.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Dataset</th>
<th>$k$</th>
<th>$m$</th>
<th>$n$</th>
<th>Mean ($Z_r$)</th>
<th>Lower CI (2.5%)</th>
<th>Upper CI (97.5%)</th>
<th>$z$ value</th>
<th>$P$ ($z$)</th>
<th>$Q$</th>
<th>d.f. ($Q$)</th>
<th>$P$ ($Q$)</th>
<th>$I^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1</td>
<td>81</td>
<td>49</td>
<td>7512</td>
<td>0.096</td>
<td>0.057</td>
<td>0.134</td>
<td>4.815</td>
<td>&lt;0.001</td>
<td>177.40</td>
<td>80</td>
<td>&lt;0.001</td>
<td>56.20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>84</td>
<td>50</td>
<td>7807</td>
<td>0.072</td>
<td>0.040</td>
<td>0.104</td>
<td>4.432</td>
<td>&lt;0.001</td>
<td>136.70</td>
<td>83</td>
<td>&lt;0.001</td>
<td>37.07</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>94</td>
<td>57</td>
<td>8788</td>
<td>0.061</td>
<td>0.033</td>
<td>0.089</td>
<td>4.238</td>
<td>&lt;0.001</td>
<td>138.35</td>
<td>93</td>
<td>0.002</td>
<td>30.02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>94</td>
<td>57</td>
<td>8788</td>
<td>0.039</td>
<td>0.012</td>
<td>0.067</td>
<td>2.817</td>
<td>0.005</td>
<td>135.46</td>
<td>93</td>
<td>0.003</td>
<td>26.77</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>1053</td>
<td>0.015</td>
<td>–0.096</td>
<td>0.127</td>
<td>0.273</td>
<td>0.785</td>
<td>19.60</td>
<td>8</td>
<td>0.012</td>
<td>58.81</td>
</tr>
<tr>
<td>Behaviour</td>
<td>1</td>
<td>13</td>
<td>11</td>
<td>1159</td>
<td>0.078</td>
<td>–0.005</td>
<td>0.161</td>
<td>1.849</td>
<td>0.064</td>
<td>23.72</td>
<td>12</td>
<td>0.022</td>
<td>43.21</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>14</td>
<td>12</td>
<td>1908</td>
<td>0.062</td>
<td>–0.046</td>
<td>0.170</td>
<td>1.118</td>
<td>0.264</td>
<td>38.59</td>
<td>13</td>
<td>&lt;0.001</td>
<td>73.29</td>
</tr>
<tr>
<td>EPP</td>
<td>1</td>
<td>16</td>
<td>14</td>
<td>3159</td>
<td>0.016</td>
<td>–0.021</td>
<td>0.052</td>
<td>0.851</td>
<td>0.395</td>
<td>15.26</td>
<td>15</td>
<td>0.433</td>
<td>0.000</td>
</tr>
<tr>
<td>Ornament</td>
<td>1</td>
<td>37</td>
<td>18</td>
<td>2394</td>
<td>0.073</td>
<td>0.017</td>
<td>0.130</td>
<td>2.534</td>
<td>0.011</td>
<td>61.86</td>
<td>36</td>
<td>0.005</td>
<td>42.56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>38</td>
<td>18</td>
<td>2418</td>
<td>0.072</td>
<td>0.016</td>
<td>0.128</td>
<td>2.528</td>
<td>0.012</td>
<td>61.94</td>
<td>37</td>
<td>0.006</td>
<td>41.31</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>43</td>
<td>21</td>
<td>2911</td>
<td>0.060</td>
<td>0.012</td>
<td>0.109</td>
<td>2.451</td>
<td>0.014</td>
<td>63.61</td>
<td>42</td>
<td>0.017</td>
<td>34.68</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>43</td>
<td>21</td>
<td>2911</td>
<td>0.042</td>
<td>–0.006</td>
<td>0.089</td>
<td>1.708</td>
<td>0.088</td>
<td>64.32</td>
<td>42</td>
<td>0.015</td>
<td>33.26</td>
</tr>
<tr>
<td>Preference</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>172</td>
<td>0.148</td>
<td>–0.180</td>
<td>0.477</td>
<td>0.885</td>
<td>0.376</td>
<td>17.26</td>
<td>5</td>
<td>0.004</td>
<td>73.43</td>
</tr>
<tr>
<td>Size</td>
<td>1</td>
<td>30</td>
<td>27</td>
<td>1757</td>
<td>0.097</td>
<td>0.033</td>
<td>0.162</td>
<td>2.950</td>
<td>0.003</td>
<td>49.02</td>
<td>29</td>
<td>0.011</td>
<td>42.67</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>31</td>
<td>28</td>
<td>1865</td>
<td>0.093</td>
<td>0.031</td>
<td>0.155</td>
<td>2.926</td>
<td>0.003</td>
<td>50.10</td>
<td>30</td>
<td>0.012</td>
<td>41.86</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>44</td>
<td>36</td>
<td>3207</td>
<td>0.060</td>
<td>0.018</td>
<td>0.102</td>
<td>2.778</td>
<td>0.005</td>
<td>57.30</td>
<td>43</td>
<td>0.071</td>
<td>25.75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>44</td>
<td>36</td>
<td>3207</td>
<td>0.025</td>
<td>–0.020</td>
<td>0.069</td>
<td>1.092</td>
<td>0.275</td>
<td>61.98</td>
<td>43</td>
<td>0.030</td>
<td>32.41</td>
</tr>
</tbody>
</table>
Table 3. Results of Egger’s regression tests and trim-and-fill (TAF) tests; * indicates missing effect sizes in the unexpected direction (for the descriptions of $Q$ and $I^2$, see text Section II); EPP, extra-pair paternity.

<table>
<thead>
<tr>
<th>Trait</th>
<th>$t$:</th>
<th>d.f.</th>
<th>$P (t)$</th>
<th>Missing k: TAF</th>
<th>Mean ($Zr$)</th>
<th>Lower CI (2.5%)</th>
<th>Upper CI (97.5%)</th>
<th>$z$ value</th>
<th>$P (z)$</th>
<th>$Q$</th>
<th>d.f. (Q)</th>
<th>$P (Q)$</th>
<th>$I^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>3.450</td>
<td>79</td>
<td>0.001</td>
<td>17</td>
<td>0.034</td>
<td>−0.012</td>
<td>0.080</td>
<td>1.443</td>
<td>0.149</td>
<td>287.08</td>
<td>97</td>
<td>&lt;0.001</td>
<td>72.99</td>
</tr>
<tr>
<td>Age</td>
<td>3.650</td>
<td>7</td>
<td>0.008</td>
<td>4</td>
<td>−0.102</td>
<td>−0.239</td>
<td>0.036</td>
<td>−1.447</td>
<td>0.148</td>
<td>44.40</td>
<td>12</td>
<td>&lt;0.001</td>
<td>80.17</td>
</tr>
<tr>
<td>Behaviour</td>
<td>0.918</td>
<td>11</td>
<td>0.378</td>
<td>0</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Condition</td>
<td>0.949</td>
<td>12</td>
<td>0.361</td>
<td>3</td>
<td>−0.009</td>
<td>−0.129</td>
<td>0.111</td>
<td>−0.143</td>
<td>0.886</td>
<td>63.89</td>
<td>16</td>
<td>&lt;0.001</td>
<td>81.71</td>
</tr>
<tr>
<td>EPP</td>
<td>0.293</td>
<td>14</td>
<td>0.774</td>
<td>0</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Ornament</td>
<td>1.515</td>
<td>35</td>
<td>0.139</td>
<td>7</td>
<td>0.029</td>
<td>−0.031</td>
<td>0.089</td>
<td>−0.009</td>
<td>0.350</td>
<td>92.54</td>
<td>43</td>
<td>&lt;0.001</td>
<td>53.75</td>
</tr>
<tr>
<td>Preference</td>
<td>−0.873</td>
<td>4</td>
<td>0.432</td>
<td>2*</td>
<td>0.304</td>
<td>−0.046</td>
<td>0.654</td>
<td>1.701</td>
<td>0.089</td>
<td>27.09</td>
<td>7</td>
<td>&lt;0.001</td>
<td>78.40</td>
</tr>
<tr>
<td>Size</td>
<td>−0.984</td>
<td>28</td>
<td>0.334</td>
<td>7*</td>
<td>0.151</td>
<td>0.087</td>
<td>0.215</td>
<td>4.618</td>
<td>&lt;0.001</td>
<td>74.02</td>
<td>36</td>
<td>&lt;0.001</td>
<td>52.33</td>
</tr>
</tbody>
</table>
Table 4. Results of multilevel and phylogenetic (random-effects) meta-analyses; \( k \) = number of effect sizes, \( m \) = number of species and \( n \) = number of experiments/populations (or studies) in each data set used for these meta-analyses (for the descriptions of Dataset, \( H^2 \) and \( I^2 \), see the text Section II).

<table>
<thead>
<tr>
<th>Type</th>
<th>Dataset (avian)</th>
<th>( k )</th>
<th>( m )</th>
<th>( n )</th>
<th>Mean ( (Z_z) )</th>
<th>Lower CI (2.5%)</th>
<th>Upper CI (97.5%)</th>
<th>( I^2 ) study (%)</th>
<th>( I^2 ) species (%)</th>
<th>( I^2 ) effect size (%)</th>
<th>( I^2 ) phylogeny (%)</th>
<th>( I^2 ) total (%)</th>
<th>( H^2 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilevel</td>
<td>1</td>
<td>195</td>
<td>42</td>
<td>85</td>
<td>0.064</td>
<td>0.030</td>
<td>0.103</td>
<td>21.03</td>
<td>6.03</td>
<td>24.50</td>
<td>–</td>
<td>51.56</td>
<td>–</td>
</tr>
<tr>
<td>(non-phylogenetic)</td>
<td>2</td>
<td>205</td>
<td>43</td>
<td>88</td>
<td>0.058</td>
<td>0.026</td>
<td>0.093</td>
<td>17.53</td>
<td>5.84</td>
<td>23.79</td>
<td>–</td>
<td>47.16</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>254</td>
<td>48</td>
<td>99</td>
<td>0.041</td>
<td>0.017</td>
<td>0.066</td>
<td>11.04</td>
<td>5.59</td>
<td>12.01</td>
<td>–</td>
<td>28.64</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>254</td>
<td>48</td>
<td>99</td>
<td>0.025</td>
<td>–0.002</td>
<td>0.052</td>
<td>12.37</td>
<td>6.08</td>
<td>18.84</td>
<td>–</td>
<td>37.29</td>
<td>–</td>
</tr>
<tr>
<td>Phylogenetic</td>
<td>1</td>
<td>195</td>
<td>42</td>
<td>85</td>
<td>0.074</td>
<td>0.005</td>
<td>0.155</td>
<td>16.26</td>
<td>5.45</td>
<td>23.21</td>
<td>10.88</td>
<td>55.79</td>
<td>9.82</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>205</td>
<td>43</td>
<td>88</td>
<td>0.069</td>
<td>–0.009</td>
<td>0.142</td>
<td>15.34</td>
<td>5.06</td>
<td>21.40</td>
<td>10.47</td>
<td>52.26</td>
<td>9.77</td>
</tr>
</tbody>
</table>
Table 5. Factors proposed to affect the relative fitness of sons and daughters differentially, leading to adaptive sex ratio adjustment strategies. Examples are given of empirical studies that report a significant effect of the factor; note that for all factors we also located studies that find no relationship. References to meta-analyses are provided where possible.

<table>
<thead>
<tr>
<th>Influencing factor</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal condition</td>
<td>Trivers &amp; Willard (1973); Clutton-Brock et al. (1984); Kruuk et al. (1999); meta-analyses: Cameron (2004); Sheldon &amp; West (2004)</td>
</tr>
<tr>
<td>Mate attractiveness/quality</td>
<td>Burley (1981, 1986) (and see Table 1, this study); meta-analyses: West &amp; Sheldon (2002); Ewen et al. (2004); Cassey et al. (2006)</td>
</tr>
<tr>
<td>Local resource enhancement (cooperative breeders: number of helpers)</td>
<td>Taylor (1981); Gowaty &amp; Lennartz (1985); meta-analyses: West &amp; Sheldon (2002); Griffin et al. (2005); Stubblefield &amp; Orzack (2013)</td>
</tr>
<tr>
<td>Local resource competition</td>
<td>Taylor (1981); Cockburn et al. (1985); meta-analysis: Silk &amp; Brown (2008)</td>
</tr>
<tr>
<td>Local mate competition</td>
<td>Hamilton (1967); Werren (1983); Shuker &amp; West (2004); meta-analysis: West et al. (2005)</td>
</tr>
<tr>
<td>Resource quality/availability: Diet quality, composition</td>
<td>Warner et al. (2008); Pryke &amp; Rollins (2012); Rutz (2012)</td>
</tr>
<tr>
<td>Habitat/territory quality</td>
<td>Romano et al. (2012); Bell et al. (2014)</td>
</tr>
<tr>
<td>Host size for parasitoid larval development</td>
<td>Godfray (1994); meta-analysis: West &amp; Sheldon (2002)</td>
</tr>
<tr>
<td>Seasonal effects:</td>
<td>Saino et al. (2008); Graham et al. (2011); Barclay (2012); meta-analysis: Ewen et al. (2004); Cassey et al. (2006)</td>
</tr>
<tr>
<td>Laying/hatch order</td>
<td>Berkeley &amp; Linklater (2010)</td>
</tr>
<tr>
<td>Clutch size</td>
<td>Cichon et al. (2003); Lezalova et al. (2005); Lislevand et al. (2005)</td>
</tr>
<tr>
<td>Social dominance rank (maternal)</td>
<td>Nevison (1997); Maestripieri (2002); meta-analyses: Brown &amp; Silk (2002); Schino (2004); Silk et al. (2005)</td>
</tr>
<tr>
<td>Parental relatedness/compatibility</td>
<td>Pryke &amp; Griffith (2009); Sardell &amp; DuVal (2014)</td>
</tr>
<tr>
<td>Parental age/breeding experience/pair bond duration</td>
<td>Green (2002); Benito et al. (2013)</td>
</tr>
<tr>
<td>Female mating status/extent of polygyny (sole/primary/secondary female, harem size)</td>
<td>Weatherhead (1983); Nishiumi (1998); Trnka et al. (2012)</td>
</tr>
</tbody>
</table>