All eggs are made equal: meta-analysis of egg sexual size dimorphism in birds

J. RUTKOWSKA*, A. DUBIEC† & S. NAKAGAWA‡

*Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland
†Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
‡Department of Zoology, University of Otago, Dunedin, New Zealand

Keywords:
egg mass; egg volume; maternal effect; offspring sex; phylogenetic meta-analysis; publication bias.

Abstract
Sex-biased resource allocation in avian eggs has gained increasing interest. The adaptive explanations of such allocation are often related to life-history strategies of the studied species. In some species, egg sexual size dimorphism (SSD) was suggested to promote future size differences between adults of each sex. In other species, egg SSD was invoked as an adaptive means by which a mother balances sex-specific nestling mortality. According to the first scenario, mothers should produce bigger eggs for the bigger sex, thus across species, adult SSD should be a significant positive predictor of egg SSD. Under the second scenario, mothers should produce bigger eggs for the smaller sex. If different species use contrasting strategies, then a universal expectation is that there should be a significant relationship between the magnitude of adult SSD and the magnitude of egg SSD, irrespective of the direction of those differences. Our aim was to examine whether the direction of egg SSD is predicted by the direction of adult SSD or whether degree of egg SSD is related to degree of adult SSD. To answer that question, we performed meta-analysis of 63 studies, which included information on egg SSD of 65 effect sizes from 51 avian species. We found that across species, adult SSD does not predict egg SSD. More importantly, the observed variation in effect sizes in our data set was largely explained by sampling error (variance). Although adult SSD is undoubtedly a prominent feature of birds, there is little evidence for egg SSD across avian species.

Introduction
Maternal effects mediated by egg size may have profound influence on offspring fitness (Williams, 1994). This influence stems from the significant effect of egg size on offspring size and juvenile survival (reviewed by Krist, 2011). The positive effects of egg size might have long-lasting consequences affecting offspring size at adulthood (Potti & Merino, 1994). In avian studies, sex-specific resource allocation has gained much interest from perspectives of how egg size modulates sexual dimorphism of the offspring (reviewed by Badyaev, 2002) and how it affects sibling competition (e.g. Blanco et al., 2003).

Thus, the two most frequently invoked adaptive explanations of egg sexual size dimorphism (SSD) are related to life-history strategies of the studied species. In some species, egg SSD was suggested to promote future size differences between the adults of each sex. It has been reported that in the white-crowned sparrow Zonotrichia leucophrys oriantha (Mead et al., 1987), the house sparrow Passer domesticus (Cordero et al., 2000) and the European blackbird Turdus merula (Martyka et al., 2010), males, which are the larger sex, hatch from larger eggs. The adaptive explanation for the observed pattern was that increased maternal investment early in life ensures potentially higher fitness returns from male offspring (e.g. Mead et al., 1987; Cordero et al., 2000).

However, not all observations of SSD follow the above pattern. For instance, in the spotless starling...
males are the bigger adult sex, but they hatch from smaller eggs (Cordero *et al.*, 2001). In the brown songlark *Cinclorhamphus cruralis*, male adults are twice as big as females, but they were reported to hatch from slightly smaller eggs (Magrath *et al.*, 2003). In the American kestrel *Falco sparverius*, female adults are larger than males, yet females hatch from lighter eggs than males (Anderson *et al.*, 1997). In those cases, egg SSD was invoked as an adaptation by which a mother reduces within-brood competition and balances nestling mortality differences between the sexes by producing bigger eggs for the smaller sex.

Under the first scenario, mothers should produce bigger eggs for the bigger sex. Thus, across species, adult SSD should be a significant positive predictor of egg SSD. According to the second scenario, mothers should produce smaller eggs for the bigger sex, and thus, with increasing adult SSD, egg SSD should decrease. If different species use contrasting strategies, then the universal expectation is a significant relationship between the absolute values of adult and egg SSD.

The possibility that adult SSD might affect egg size was established in an analyses relating adult SSD to egg size at the between-species level (Weatherhead & Teather, 1994). This review revealed that with increasing adult SSD, egg size relative to female size increases. The authors verbally refuted the possibility that the effect was driven by females laying larger eggs for the larger sex (Weatherhead & Teather, 1994). Their argument was based on the lack of egg SSD in some highly dimorphic species reported in the relevant studies (e.g. Bancroft, 1984; Weatherhead, 1985; Teather, 1989) available at that time. Currently, studies addressing egg SSD are much more numerous, and this enables us to perform the first systematic quantitative assessment on the relationship between egg size and offspring sex across many bird species.

The aim of our study is twofold. Firstly, we review the existing data on avian egg size in relation to offspring sex and apply the meta-analytical approach to answer whether egg SSD is indeed a common phenomenon. Secondly, we examine whether direction of egg SSD is predicted by the direction of adult SSD or whether degree of egg SSD is related to degree of adult SSD.

**Materials and methods**

**Data collection and inclusion criteria**

For systematic literature search, we followed the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement as much as possible (http://www.prisma-statement.org/). The literature was searched on *Web of Science* and Google Scholar (to identify studies not reported in the literature databases, such as dissertations) using the keywords: ‘egg size,’ ‘egg mass,’ ‘egg volume,’ ‘laying order,’ ‘hatching order,’ ‘sexual dimorphism,’ ‘sex ratio,’ ‘sex’ and ‘bird’ in different combinations. From the resulting list, we identified publications, in which title indicated that the study was carried out on birds, and scanned their abstracts looking for information on egg size comparisons in relation to sex. When such information was found, or the abstract indicated that collected data might enable such comparisons, the full text was consulted. Where papers were selected, their references (backward search) and citation record (forward search) were searched for other articles providing data of interest.

When available, information on estimates of mean male and female egg size, their SD (or SE) and sample sizes (the numbers of eggs, clutches and females included in studies) was extracted from the publication. If a publication (among the final selection of publications) lacked some of the required information, or information in the publication indicated that the data of interest were collected, but not presented, the corresponding author or co-authors were contacted. Contacted authors were also asked for any unpublished data.

In some studies, numerical data were extracted from graphs using WebPlotDigitizer (Rohatgi; online source). When data on sex-associated differences in egg size were presented in relation to laying order and the information on sample size was available, weighted means of egg size were calculated for male and female eggs. When the sample size for each category was not available, arithmetic mean was calculated.

If descriptive statistics were not available, we collected inferential statistics, which were associated with the difference between female and male egg size, namely these were *t* and *F* values and the associated degrees of freedom d.f. (note that the numerator d.f. of *F* values was always 1 as comparisons were between two categories, male and female eggs). In most cases, information on egg size was available in the form of either ‘volume’ or ‘mass’, and we recorded which measure was used. In one case (Badyaev *et al.*, 2006a), egg size measurements were two dimensional (an area-based measure), but this measurement was noted as volume. When both were available (e.g. Clotfelter, 1996), we extracted data on mass.

Information on adult SSD was obtained from the same study as the data on egg SSD. When it was not available in the original publication, we used the Dunnning (2008) database, Cornell Lab of Ornithology online resources and published papers on specific species. For statistical analysis, we used the natural logarithm of adult SSD.

There are a number of studies, which we had to exclude (Fig. 1), although they potentially had relevant information. Those studies and the reasons for exclusion are summarized in Data S1.
Effect size and sampling variance calculation

Although it is more customary to use standardized mean difference (e.g. Hedge’s $d$) for meta-analysis of comparisons between two means (Nakagawa & Cuthill, 2007), we used $Z_r$ (Fisher’s z-transformation of correlation coefficient) as our standardized effect size. We made that choice because estimating the sampling variances that accompany effect sizes in our data set was easier for $Z_r$ than Hedge’s $d$. We note that the use of $Z_r$ is common in ecological and evolutionary meta-analyses (Nakagawa & Santos, 2012), even when assessing differences. For example, $Z_r$ values (or correlation coefficients) have been commonly used in meta-analyses of sex ratio adjustment studies (e.g. West & Sheldon, 2002). We also note that $r$ (or $Z_r$) and $d$ values are readily convertible (Nakagawa & Cuthill, 2007).

The calculation of sampling variance for an effect size requires the sample size associated with that effect size; in the case of $Z_r$, $N-3$ is used for the calculation of sampling variance. However, in most studies, more than one egg came from any one clutch (or one female). Hence, if the number of eggs were used as $N$, we would risk introduction of a correlated data structure (i.e. pseudo-replication), increasing the rate of type I error (see Higgins & Green, 2008).

Therefore, we used clutch as our clustering unit to calculate ‘effective sample size’ (Rao & Scott, 1992); this was done using the following formulas:

$$M = N_{\text{egg}} / N_{\text{clutch}},$$
$$D = 1 + (M - 1) \text{ICC},$$
$$N_{\text{effective}} = N_{\text{egg}} / D,$$

where $N_{\text{egg}}$, $N_{\text{clutch}}$ and $N_{\text{effective}}$ are the number of eggs, the number of clutches and the effective sample size for an effect size, $M$ is the average cluster (i.e. clutch size), $D$ is known as the ‘design effect’ and ICC is intraclass correlation (Higgins & Green, 2008). Note that this series of calculations will result in $N_{\text{effective}} = N_{\text{egg}}$ if $N_{\text{egg}} = N_{\text{clutch}}$. Also, when ICC = 1, $N_{\text{effective}} = N_{\text{clutch}}$, whereas when ICC = 0, $N_{\text{effective}} = N_{\text{egg}}$. The number of clutches used in a study was often identical to the number of females. If there was no information on the number of clutches, we used the number of females.

ICC is known as ‘repeatability’ in the biological literature (Nakagawa & Schielzeth, 2010). In our case, ICC represents how consistent egg size is within a clutch in relation to egg sizes in other clutches. Unfortunately, no studies provided ICC or the information required to obtain it, and thus, we estimated ‘representative ICC’
using 3 data sets, namely Rutkowska & Cichoń (2005), Pariser et al. (2012) and Bowers et al. (2013). ICC estimates for egg size were 0.65, 0.57 and 0.68, respectively. Therefore, we decided to create two data sets: one using the ICC of 0.5 and the other 0.8. Below, we only present the results from the data set with the ICC of 0.8, which provides more conservative estimates of sampling variance. We provide the results from the data set with ICC of 0.5 in the Data S1. Notably, almost all results are qualitatively identical between the two data sets.

Correlation coefficients, \( r \) (later converted to \( Z_r \) values), for meta-analysis were calculated using standard formulas provided in Lipsey and Wilson (2001) and Nakagawa and Cuthill (2007). We note that in many cases, first, \( d \) values were calculated using those formulas, before being converted to \( r \). Calculating effect sizes from \( t \) values from paired \( t \)-tests and \( F \) values from repeated-measures ANOVA (i.e. nonindependent tests) requires a correlation value between pairs, that is, between male and female eggs from the same clutches/females (note that this correlation is different from correlation values, \( Z_r \), used for meta-analysis); see Equation 4 in Nakagawa & Cuthill (2007). Such a correlation is quantitatively very similar to ICC in this scenario (for a relevant explanation, see Nakagawa and Schielzeth, 2010). Therefore, we used correlations of 0.5 and 0.8 in the respective data sets to obtain our effect size, \( Z_r \), from \( t \) and \( F \) values of nonindependent tests (only five cases required to assume such correlations).

Our database includes 65 effect sizes from 63 studies, covering 51 species (Fig. 1; see the Table S1). From Badyaev et al. (2006a), we extracted two effect sizes from two different populations of one species, whereas from Slagsvold et al. (1992), we obtained two effect sizes from two different species. Therefore, we assumed that we had 65 independent effect sizes in our meta-analysis, although we also considered phylogenetic nonindependence in separate analysis (see Results).

**Statistical analysis**

All analyses were conducted in the R environment (version 3.0.1; R Development Core Team 2013). The random-effects model (meta-analysis) and mixed-effects model (meta-regression) with the restricted maximum likelihood (REML) estimator were fitted using the R package, metafor (Viechtbauer, 2010). Also, for phylogenetic meta-analytic models (Hadfield & Nakagawa, 2010; Nakagawa & Santos, 2012), we used the R package, MCMCglmm (Hadfield, 2010). Statistical heterogeneity was quantified using \( I^2 \) for standard meta-analysis (Higgins et al., 2003) and a modified version of \( I^2 \) for phylogenetic meta-analysis (described in Nakagawa & Santos, 2012). The values of 25%, 50% and 75% are considered to be low, medium and high, respectively (Higgins et al., 2003). This is, in a way, similar to \( r \) values of 0.1, 0.3 and 0.5, which are considered to be small, moderate and large, respectively (sensu Cohen, 1988). For \( Z_r \), corresponding values are 0.10, 0.31 and 0.55. For sensitivity analyses where we checked potential biases and publication biases in our data, we used functions from metafor.

**Results**

**Meta-analysis and meta-regression**

Egg SSD was not different from zero (random-effects model: \( \hat{b}_{[\text{meta-analytic mean}]} = 0.021, 95\% \text{ confidence interval, CI} = -0.011 \text{ to } 0.053; \) Fig. 2). Importantly,
Lack of sexual size dimorphism in avian eggs

Statistical heterogeneity measured by $I^2$ is low (12.70%). Also, this heterogeneity was not statistically significant (heterogeneity test: $Q_{64} = 62.882, P = 0.516$), which indicated that our data containing 65 effect sizes are consistent with the idea that there is no or little egg SSD. In other words, after controlling for sampling variance, there was only low or little variation left in our data. The results from our phylogenetic meta-analysis are quantitatively very similar to the aforementioned mean effects and are presented in the Data S1.

As expected from the above result, adult SSD did not predict the direction of egg SSD (meta-regression: $b_{[\text{meta-regression}]} = -0.148$, 95% CI = -0.488 to 0.190; Fig. 3a). To test whether degree of egg SSD is related to degree of adult SSD, we used a Spearman’s rank correlation test between the absolute values of effect sizes and those of adult SSD on the natural logarithm scale. This is because the absolute values of effect sizes were not normally distributed (they followed what is known as a ‘folded’ normal distribution, cf. Morrissey & Hadfield, 2012), which meant that we could not use meta-analytic models. There was a very low and nonsignificant relationship between the absolute values of effect sizes and adult SSD ($r_S = 0.067$, 95% CI = -0.179 to 0.306; Fig. 3b).

Sensitivity analysis and publication bias

To ascertain the robustness of our results, we ran a meta-regression model with three potentially confounding variables: (i) publication year (see Trikalinos & Ioannidis, 2005); unpublished effect sizes were assigned the year, 2014; this variable was centred (sensu Schielzeth, 2010); (ii) whether the effect size was based on egg mass or volume; and (iii) whether effect sizes were estimated from descriptive or inferential statistics. None of these factors predicted variation in effect size (meta-regression: $b_{[\text{Year}]} = -0.002$, 95% CI = -0.007 to 0.003, $b_{[\text{Measure}]} = 0.022$, 95% CI = -0.044 to 0.088, $b_{[\text{Statistic}]} = 0.051$, 95% CI = -0.037 to 0.138).

We also tested for publication bias using both Egger’s tests (Egger et al., 1997) and the trim-and-fill method with the ‘L0’ estimator (Duval & Tweedie, 2000a,b) because Egger’s tests are sensitive to statistical power (Higgins & Green, 2008). Although Egger’s test did not provide any evidence for publication bias ($t_{63} = 1.621, P = 0.110$), the trim-and-fill method identified asymmetry in the funnel plot and added eight potentially ‘missing’ effect sizes (Fig. 2). The meta-analytic mean after taking into account these filled values was very close to zero (random-effects model: $b_{[\text{meta-analytic mean}]} = 0.001$, 95% CI = -0.032 to 0.034), reinforcing the lack of evidence of egg SSD across avian species.

Discussion

In this study, we meta-analytically tested whether egg SSD is a general phenomenon across bird species. Our results suggest that egg SSD is rather an exception than the norm (Fig. 2). We also tested two frequently invoked explanations of egg SSD in relation to adult SSD. We did not find any evidence that adult SSD could predict egg SSD (Fig. 3). Our additional sensitivity analyses and bias corrections indicated the robustness of our results. Thus, although adult SSD has a significant effect on between-species egg size allometry (Weatherhead & Teather, 1994), we conclude that female birds, as a general rule, do not seem to adjust egg size according to the sex of their offspring.

This conclusion is in line with a number of studies that expected significant egg SSD in highly dimorphic species, but failed to find any evidence of such a trend. For instance, the Northern lapwing Vanellus vanellus exhibits pronounced adult SSD and strong sexual selection, yet its eggs are not sexually dimorphic (Lislevand et al., 2005). Note, however, that nonstatistically significant differences in egg size might only reflect the lack of statistical power in data sets; this point is discussed in detail below. The brown songlark Cinclorhamphus cruralis has the highest adult male-to-female body mass ratio in our data set, yet the reports on its egg SSD either show that daughters hatch from larger (Magrath et al., 2003) or nonsignificantly smaller eggs (Isaksson et al., 2010). Previously assumed egg SSD of the white-crowned sparrow (Mead et al., 1987) and of the house
sparrow *Passer domesticus* (Cordero et al., 2000) has also been questioned by newer the findings of Bonier et al. (2007) and Wetzel et al. (2012), respectively. Although female birds are able to lay extremely dimorphic eggs (Stein & Williams, 2013), and there are several benefits that could come along with egg SSD (Mead et al., 1987; Anderson et al., 1997; Cordero et al., 2000; Love & Williams, 2011), it seems that maternal adjustment of egg size in relation to offspring sex is not a prevailing phenomenon.

It is entirely possible that there might be a subtle size difference between male- and female-bearing eggs. However, to detect such an effect at the significance level ($z = 0.05$), one would have to measure approximately 20,000 eggs, assuming our meta-analytic mean ($r = -0.02$; equivalent to $d = -0.04$) with the statistical power of 80%. Such data sets are usually not available for wild birds. In our data set, 3313 eggs (from 205 clutches; extended data from Petrie et al. (2001)) represented the largest sample size. In poultry species, in which one would expect larger data sets, no relevant information is available. In fact, the poultry industry would be vitally interested in selecting for chicken breeds in a way that would enable a predictable association between egg sex and size. To our knowledge, such a goal has not been achieved. This might be due to stabilizing selection on ‘optimal’ egg size, which is evidenced by the highest hatching success for intermediate-sized eggs (Lerner, 1951).

The lack of clear support for sex-specific investment in the egg size across species does not preclude the possibility of differential maternal allocation at a finer scale. Firstly, egg SSD might be observed at only a certain position of the egg-laying sequence, an illustrative case being the Eurasian kestrel *Falco tinnunculus*, in which only the first eggs are sexually dimorphic (Blanco et al., 2003). Our meta-analysis bares a small risk of diluting such effect as it takes into consideration all eggs. Secondly, eggs might be identical in terms of total mass, but the sexes could differ in relative content of albumen and white, as was reported in the ring-billed gull *Larus delawarensis* (Chin et al., 2012). Finally, it is also likely that egg components, such as immunoglobulins (Saino et al., 2003; Martyka et al., 2011), antioxidants and hormones (Petrie et al., 2001; Badyaev et al., 2006b; Pariser et al., 2012; but see Rubolini et al., 2011) are distributed in a sex-specific manner. More studies on egg quality are needed before it could be assessed whether fine-scale sex-specific egg dimorphism is a prevailing strategy across birds.

Five of 65 effect sizes, corresponding to five different species, were outside the 95% CI around zero as seen in Fig. 2. Although, by random chance, we expected at least 3 species to be outside the 95% CI, egg SSD could constitute a biologically relevant phenomenon in these five cases. We find, however, no obvious ecological explanation why those data stand out. The species that are outstanding in our database are neither phylogenetically related nor particularly dimorphic as adults (the canary, *Serinus canaria* (Leitner et al., 2006), the European blackbird (Martyka et al., 2010), the parasitic jaeger, *Stercorarius parasiticus* (Janssen et al., 2006) and the blue-footed booby, *Sula nebouxii* (D’Alba et al., 2007)). As mentioned before, in the case of brown songlark, the results from two studies were inconsistent.

Differences in egg size have sometimes been discussed from the perspective of offspring sex determination and the potential interplay between egg size and offspring sex at the proximate level (e.g. Rutkowska & Cichoń, 2002; Müller et al., 2005). In the light of our current results and given that final egg size is decided after fertilization (Etches, 1996), it seems improbable that egg size is under the influence of an egg’s sex. What is more likely is that both egg size and offspring sex are influenced by a third factor. For example, in several avian species, laying order has a simultaneous effect on offspring sex, egg components and egg size (reviewed in Gil, 2008; see also Blanco et al., 2003).

In conclusion, although female birds of some species are known to be able to adjust egg size within a clutch and such adjustments could indeed be beneficial (e.g. Stein & Williams, 2013), it seems that there has not been strong enough selection for within-species (nor within-female) adjustment of egg size based on the offspring sex. We recommend that any future studies in this topic, if at all, need to have an ambitious sampling design measuring eggs (at least one male and one female egg) from 200 or more different mothers. Such a design may allow researchers to statistically detect small SSD in eggs (an effect size of $r = -0.1$, or $d = -0.2$) with a power of 80%.

In a wider context, we can see that our study adds another example, to the evolutionary and ecological literature, where meta-analysis revealed that previously alleged effects or relationships are unlikely to exist, or are likely to be too small to be biologically important (for examples, see Palmer, 1999; Jennions et al., 2013a). This highlights the importance of ‘meta-analytic thinking’, whereby we should view each empirical piece of work as a modest contribution to a specific topic (Thompson, 2002), and also of ‘effective thinking’ (Nakagawa & Cuthill, 2007), where interpretation of scientific results should focus on the magnitude of effect rather than statistical significance. Conclusions from single studies are more likely to be erroneous than meta-analytic conclusions due to type I and type II errors; the latter is common for studies with small sample sizes, which are typical of ecological and evolutionary studies. Therefore, the generalization of results from a single piece of empirical work is limited, and meta-analytic insights are often required to resolve inconsistent empirical results and to draw general conclusions (Nakagawa & Poulin, 2012; Jennions et al., 2013b).
Acknowledgments

We are grateful to all the authors who shared with us the details of their published and unpublished studies. We thank Zofia Prokop for helpful comments on the manuscript and Alistair Senior for comments and proofreading the final version. J. R. was supported by DS/WWiNoZ/INoS/757/13 and by Polish National Science Center DEC-2012/07/B/NZ8/01293. S. N. was supported by the Rutherford Discovery Fellowship (New Zealand).

References


Supporting information

Additional Supporting Information may be found in the online version of this article:
Table S1 Database used in the analyses.
Data S1 Additional methods and results.

Received 31 August 2013; revised 15 October 2013; accepted 21 October 2013