Decomposing phenotypic skew and its effects on the predicted response to strong selection

Decomposing phenotypic skew and its effects on the predicted response to strong selection

Joel L. Pick[1,2*], Hannah E. Lemon[1], Caroline E. Thomson[1] & Jarrod D. Hadfield[1]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom
[2]Centre of Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway

[*] Corresponding Author: joel.l.pick@gmail.com

The major frameworks for predicting evolutionary change assume that a phenotype's underlying genetic and environmental components are normally distributed. However, the predictions of these frameworks may no longer hold if distributions are skewed. Despite this, phenotypic skew has never been decomposed, meaning the fundamental assumptions of quantitative genetics remain untested. Here, we demonstrate that the substantial phenotypic skew in the body size of juvenile blue tits (Cyanistes caeruleus) is driven by environmental factors. Although skew had little impact on our predictions of selection response in this case, our results highlight the impact of skew on the estimation of inheritance and selection. Specifically, the non-linear parent-offspring regressions induced by skew, alongside selective disappearance, can strongly bias estimates of heritability. The ubiquity of skew and strong directional selection on juvenile body size implies that heritability is commonly overestimated, which may in part explain the discrepancy between predicted and observed trait evolution.

Quantitative genetics describes how traits respond to selection in terms of selection and inheritance. Typically we use two equations to describe this, the breeder's equation ([1] Chapter 12) and Lande's gradient equation ([2] Eq 7). The breeder's equation gives the predicted response to selection as the heritability (h^2) multiplied by the selection differential (S), whereas Lande's gradient equation describes the response to selection as the additive genetic variance (V_A) of the trait multiplied by the selection gradient (β). Although these frameworks are generally thought to be interchangeable, they only converge when phenotypes (and their genetic and environmental components) are normally distributed or fitness functions (the relationship between a trait and fitness) are linear ([3] Chapter 29). Given that fitness functions are highly unlikely to be linear in practice[4;5], any deviation from normality can lead to problems with the application of these equations. Consequently, normality is seen as a fundamental assumption in quantitative genetics[6–8], yet to our knowledge has not been directly tested, despite the major consequences it has for how traits are predicted to respond to selection[9–17].

The most natural interpretation of heritability in the context of the breeder's equation is the slope of a *linear* parent-offspring (PO) regression[12–14;18;19], whilst S (the covariance between a trait and fitness) describes the *linear* relationship between a phenotype and fitness. The accuracy of the breeder's equation relies heavily on the linearity of both of these functions - if both are non-linear, the residuals from the linear functions may be correlated, creating a 'spurious response to selection'[14]. The linearity of the parent-offspring relationship breaks down when the amount of skew (asymmetry) differs between genetic and environmental components[20;21], with genetic and environmental skew causing curvature in opposite directions (Figure 1). Formally, a distribution is skewed when it has a non-zero third central moment.

Whilst the gradient equation is robust to environmental skew, it doesn't correctly describe the response to selection in the presence of genetic skew if the fitness function is non-linear ([11] Eq 42). Environmental skew, through it's contribution to phenotypic skew, can, however, impact the estimation of β when it is approximated using Lande-Arnold regression[5;17;22].

Although extensions to these two equations have been derived that allow for the non-linearity of the PO-regression[12] and the non-normality of genetic values[11], the majority of the work in this area remains theoretical. Non-linearity in PO-regressions has been demonstrated in

the lab[12;23–28] and ad-hoc methods have been used to test for skew at the genetic level[29;30]. Nevertheless, to our knowledge, no study has 1) relaxed the normality assumptions when making statistical inferences to examine the origin and extent of skew at different levels, and 2) explored how observed patterns of natural selection interact with skew to determine how well these two equations predict selection response in the wild.

Juvenile body size is under strong, persistent, directional selection across taxa[31], yet is known to show little response to this selection[32]. We show that juvenile body size is highly negatively skewed (long tail of small individuals) across bird species, but the origin of this skew is unknown. To determine this, we developed statistical methods to decompose the phenotypic distribution into a set of skew-t distributions, and predict the shape of PO-regression based on the estimated skew. We applied these methods to data from a long-term cross-fostering experiment of a wild bird population. By estimating survival selection acting on juvenile body size, we tested the robustness of the predicted response to selection from the breeder's and gradient equations.

Results

Prevalence of Phenotypic Skew

Across 27 species of birds, tarsus length (a common measure of structural size) was substantially negatively skewed (long tail of small individuals) in juveniles (coefficient of skew: -1.054 [-1.394, -0.686], pMCMC$<$0.001), but not adults (-0.302 [-0.641, 0.052], pMCMC$=$0.086), with tarsus length being significantly more skewed in juveniles than adults (difference $=$ -0.752 [-1.124, -0.366], pMCMC$<$0.001; Figure 2).

Decomposing Phenotypic Skew

Using data on four juvenile body size traits (tarsus length, head-bill length, mass and wing length), measured on 15 day old chicks from a long-term cross-fostering experiment on a wild population of blue tits, we decomposed phenotypic skew into genetic, between- and within-nest environmental components. We used a mixed model approach with skew-t distributed random effects which allowed the extent and direction of skew to vary between these levels. There was considerable phenotypic skew in all four traits, with the coefficient of skew ranging from -0.51 to -1.60 (Figure 3). There was little evidence of genetic skew in any trait (Figure 3, Tables S5, S8, S11 and S12 and further discussion in supplementary methods). Phenotypic skew was instead driven by considerable environmental skew at both between- and within-nest levels, with the relative magnitude of this skew varying between traits (Figure 3, Tables S6, S9, S12 and S15).

Given the environmental origin of the negative phenotypic skew, we would expect a convex PO-regression for all traits[20] (Figure 1c). By deriving a method to compute this non-linear PO-regression (Equation 1), we can show that for all traits the slope in the lower tail of the distributions is close to zero, but becomes steeper with increasing body size (Figure 3).

Selection on Juvenile Body Size

To quantify selection acting on body size, we estimated the linear and quadratic effects of body size on survival from both day 15 to fledging and fledging to local recruitment in a bivariate probit event-history model. As expected, all traits showed significant positive linear effects of body size on survival at both stages, with survival increasing at larger body sizes (Figure 4, Tables S16-19). Interestingly, all quadratic effects of juvenile size on survival between day 15 and fledging were positive, with these effects being suggestive and significant for mass and wing length, respectively (Figure 4, Tables S16-19), indicating an accelerating effect of size on offspring survival. In contrast, negative quadratic effects were typical for survival from fledging to recruitment although this effect was only suggestive in the case of tarsus length (Figure 4, Tables S16-19). The fitness functions over both events were generally concave (Figure 4), which would indicate stabilising selection, but the hypothesis that the optimal trait value lay outside of the observed phenotypic range for any trait could not be rejected (proportion of iterations with an internal optimum: tarsus 0.853; head-bill 0.543; mass 0.757; wing 0.017).

Using these fitness functions, we were able to estimate selection gradients (β) for each trait by taking the partial derivative of the individual relative fitness function with respect to the trait and averaging it over the trait's distribution. However, β is more frequently approximated using a Lande-Arnold regression of fitness on a trait[22] and phenotypic skew can bias this approximation when the fitness function is not linear or quadratic (as is the case for survival functions)[22]. To test this, we calculated the expected estimates of β that would be obtained from the Lande-Arnold approach without (β_1) and with (β_2) a quadratic term fitted[22;33;34], over the posterior distribution of the survival models (Equations 10 and 11). Figure 4 shows that generally there is little meaningful difference between estimates, with the exception of wing length, where there is suggestive evidence that β_1 would underestimate β by approximately 30% (β_1/β: 0.702 [0.494, 0.881], pMCMC=0.010).

Predicted Response to Selection

In the absence of genetic skew, the correct response to selection is given by Lande's gradient equation ($V_A\beta$), which for these traits gives: tarsus: 0.085mm [0.034, 0.127]; head-bill: 0.069mm [0.037, 0.102]; mass: 0.094g [0.052, 0.139]; wing: 0.175mm [0.077, 0.280]. The breeder's equation is equal to the gradient equation when the Lande-Arnold regression without the quadratic term gives good estimates of the selection gradient, irrespective of whether the PO-regression is linear or not (i.e if $\beta_1 = \beta$ then $h^2S = V_A\beta$;[3] Chapter 29). Given the similarity between β and β_1 for tarsus, head-bill and mass, the breeder's equation will therefore give accurate predictions of the selection response for these traits. However, it underestimates the response to selection in wing length by approximately 30%, as the proportional change in the predicted response to selection is equal to β_1/β (shown above).

Selection Bias and Heritability Estimation

The heritability in the breeder's equation is the heritability *before* selection (h_b^2) which can be interpreted as the slope of the PO-regression averaged over all individuals irrespective of their fitness. However, direct estimates of the PO-regression can only be obtained from individuals

that survive to become parents and so to some extent measure the heritability *after* selection (h_a^2; note the terms heritability before and after selection are used in a broader sense than in[14], and capture a different bias; see[3] p171 for a clear explanation of Heywood's usage). Since larger individuals are more likely to survive, and the PO-regression is steeper for these individuals, direct estimates of the PO-regression are likely to be upwardly biased estimates of heritability. To demonstrate this, we obtained direct estimates of the PO-regression from the 182 individuals (118 male and 64 female) that were measured as chicks and survived to produce offspring that were also measured. Although the estimated linear regression (blue line in Figure 5) is similar to the predicted non-linear PO-regression (red line in Figure 5) for the large surviving individuals (the linear and non-linear regressions fit the data equally well for all traits; tarsus $p = 0.195$, head-bill $p = 0.087$, mass $p = 0.060$ and wing $p = 0.052$), the two diverge substantially at small body sizes (Figure 5). In order to directly compare h_a^2 and h_b^2, we used the parameters of the quantitative genetic and survival models described above to calculate h_a^2 as the linear PO-regression weighted by the fitness of the parents (Equation 16) and h_b^2 as V_A/V_P. For tarsus, head-bill and mass, h_a^2 was substantially and significantly higher than h_b^2, with a proportional increase in h_a^2 of over 60% for head-bill and mass (h_a^2/h_b^2: tarsus 1.223 [1.137, 1.333], pMCMC=0.002; head-bill 1.664 [1.421, 1.951], pMCMC<0.001; mass 1.645 [1.325, 2.046], pMCMC<0.001; wing 1.584 [0.373, 2.551], pMCMC=0.372).

Estimates of h_b^2 will only be accurate if they do not suffer from the same selection bias present in PO-regression. Our experimental cross-fostering design means that the majority of information used to estimate V_A in our analysis comes from the comparison of siblings (569 nests have chicks from at least 2 clutches), rather than parents and offspring (182 parent-offspring comparisons). Sibling comparisons are made before selection, and so should not suffer from the same selection bias as parent-offspring comparisons. However, many wild bird pedigrees rely largely on information from parent-offspring relationships to estimate genetic effects - without partial cross-fostering and using social pedigrees (no within-nest variation in relatedness), sibling comparisons provide little information on genetic effects because they are confounded with common environment (nest) effects. As both PO-regression and the animal model assume that the relationship between offspring and parental phenotypes is linear, animal models relying mainly on the information from parent-offspring comparisons may also be biased. To test this, we simulated data using the parameters from our quantitative genetic and selection models for mass, assuming social and genetic monogamy, with and without skew and with and without partial cross-fostering. As expected, environmental skew caused h^2 estimated from PO-regressions to be consistently and substantially upwardly biased by a similar amount as we observed in our data, regardless of cross-fostering (estimated/simulated: no cross-fostering 1.609; cross-fostering 1.616). Without cross-fostering (information mainly from parent-offspring comparisons), estimates of V_A, and so heritability, from animal models were upwardly biased, although less than in the PO-regressions (estimated/simulated: V_A 1.226, h^2 1.228), whereas cross-fostering (information mainly from sibling comparisons) led to the correct estimation of V_A and h^2 (estimated/simulated: 1.012 and 1.015 respectively; Table 1).

Discussion

A common assumption in quantitative genetics is that phenotypes, and their underlying genetic and environmental components, are normally distributed. Here we demonstrate that this assumption is commonly violated, and in four morphological traits the observed negative phenotypic skew is driven by environmental, rather than genetic, skew. There was strong directional viability selection acting on all four traits, with non-linear fitness functions. Under these conditions the breeder's equation may give inaccurate predictions for the response to selection, but Lande's gradient equation - which only assumes genetic values are normally distributed - is expected to be accurate[11]. However, this assumes that the methods used to obtain estimates of β and V_A are robust to deviations from normality. Here we empirically demonstrate that common methods used to estimate both metrics can produce biased estimates in the presence of environmental skew.

Perhaps the most striking result is the apparent absence of genetic skew. Theory shows that directional selection can generate genetic skew, but the direction of the skew differs between models. Under the infinitesimal (Gaussian descendants[35]) model (assumed in our analyses), directional selection can drive a Gaussian distribution of breeding values to be skewed in the direction of selection through the build up of linkage disequilibrium[11;36;37]. However, stabilising selection may mitigate this ([11] Eq 46) and the breeding value distribution quickly returns to normality if selection ceases (the skew quarters each generation for unlinked loci; [36] p149). Finite allele models also generate genetic skew through changes in allele frequency. Under the rare-alleles model, directional selection after a long period of stabilising selection generates skew in the direction of selection[10;11] but sustained long term directional selection (with new mutations, on average, having effects in the opposite direction) is expected to drive skew in the opposite direction to selection[38;39]. Given juvenile body size appears to be under sustained positive directional selection[31] and gene knockout studies in mice show that loss-of-function mutations reduce size more often than increase it[40], we would predict negative genetic skew in our system. However, these models predict that the amount of skew generated through selection should be small, consistent with our finding of no or negligible genetic skew. Other processes, such as few loci, alleles of large effect, extreme allele frequencies or substantial non-additive gene action, particularly directional dominance, could generate greater levels of skew[21;38;41;42]. This seems unlikely for body size, which appears to be highly polygenic[43;44], although the finding that inbred individuals are on average smaller does suggest some directional dominance[45–48] which would also generate skew in the opposite direction to selection. Two other studies have looked at the distribution of breeding values (indirectly through estimating the skew of breeding values estimated in a Gaussian model) and while one also found little evidence of skew[30], the other found skew in the opposite direction to selection[29]. Lack of genetic skew would also be a consequence of selection acting on an environmentally correlated trait, rather than acting directly on size[49;50] (discussed further below). More widespread assessments of the prevalence of genetic skew are needed to assess the generality of these results.

Environmental skew has received little attention from theoreticians, with most studies assuming that environmental effects are normally distributed[11;12;14]. There are, however, several biological processes that are known to induce environmental skew. As far as we are aware, these processes are all predicted to generate negative environmental skew, which fits with our

general observation of negative skew in juvenile body size across species (Figure 2). For example, asymmetric competition, when larger individuals have a disproportionate negative competitive effect on others, can drive negative skew[51–54]. Blue tits have moderate levels of hatching asynchrony (hatching spread is approximately 2 days; see[55] for distribution across bird species) which is expected to generate asymmetries in competitive ability[56] and therefore skew at the within-nest level. However, the dominant source of phenotypic skew is at the between-nest level (contribution to phenotypic skew relates to standardised skew and variance) and so if asymmetric competition was the main driver of phenotypic skew, it would require parental ability to be driven by asymmetric adult competition, perhaps through differences in condition and/or territory quality. An alternative explanation is that (some) chicks have yet to reach their asymptotic size by the time of measurement and so variation in their size at this time is driven by variation in growth rate and asymptotic size. If variation in growth rate is largely at the between-nest level and variation in the asymptote is largely genetic, as has been suggested in great tits[57], then the non-linearity of growth functions could result in skew that is primarily environmental in origin (see[58] for a related result). This skew would be expected to disappear further into development as all chicks reach their asymptotic size, but due to the strong selective disappearance of small chicks this may not necessarily manifest itself (see below).

The strong, negative environmental skew led the PO-regression in all traits to be convex. This occurs because the long tail of small individuals are primarily small because of environmental factors and so resemble their parents less than larger individuals. Most discussions of the linearity of the PO-regression focus on how, in combination with a non-linear fitness function, a non-linear PO-regression leads the breeder's equation to be inaccurate, through generating a covariance between the residuals from a linear fitness function and the linear PO-regression[3;14] (see also Figure S18). This 'spurious response to selection'[14] will be largest when the non-linear fitness function and the PO-regression either have the same non-linear shape (e.g. both concave) causing a positive covariance between residuals, leading the breeder's equation to under-estimate the response to selection or opposite shapes (e.g. one concave and one convex), creating negative covariance between residuals and so over-estimation of selection response. Skew generates quite predictable and simple non-linearity in the PO-regression (Figure 1), and so generally accelerating or decelerating fitness functions will be more likely to generate a spurious response to selection, as is seen with wing length (Figure S18).

We additionally show that the selective disappearance of small individuals alongside a non-linear PO-regression leads to h^2 estimates that are biased towards the slope of the surviving large individuals. This selection bias is particularly striking in estimates from PO-regression (approx 65% increase in h^2 for mass and head-bill length; Figure 5) but importantly also occurs in animal models applied to pedigrees where information about the genetic variance comes primarily from parent-offspring comparisons (e.g. typical bird pedigrees without cross-fostering), although to a lesser degree (23% increase in animal models compared to a 61% increase in PO-regression; Table 1). This bias occurs because both PO-regression and the animal model assume that the relationship between offspring and parental phenotypes is linear, and so assumes the missing parent-offspring comparisons would follow the same slope. It is worth noting that we simulated closed populations and so a higher relatedness structure than in most wild bird populations, which are characterised by low recruitment and high immigration. Thus, our simulations likely underestimated the possible bias in animal models. We also

demonstrated that cross fostering eliminated this bias in animal models. This occurs because cross-fostering shifts the majority of the information for estimating V_A from parent-offspring comparisons, to sibling comparisons, and sibling comparisons are made before selection whilst parent-offspring comparisons are made after.

Previous work in this system has shown that selection differentially eliminates negative environmental, but not genetic, deviations for mass over the course of development[59]. This was interpreted as mass being an environmentally correlated target of selection rather than the true target (i.e. no causal relationship between size and survival)[49]. However, incorporating skew into our models challenges this interpretation as, under our model, size is the true target of selection. As the long tail of small individuals are small for environmental reasons, the selective disappearance of these individuals drives the observed decrease in environmental variance and skew though ontogeny. Given the selective disappearance previously observed was prior to the measurements analysed here[59] it seems likely that the environmental skew we observe is an underestimation of the true skew, meaning we are likely underestimating the true non-linearity of the PO-regression. Multivariate methods would account for this selective disappearance[60], however, these proved too complex to implement in this instance.

Given the consistent negative environmental skew we see across the four traits, and the conserved nature of negative phenotypic skew in juvenile (but not adult) size across bird species, we believe a concave PO-regression for juvenile size traits might be a general finding. As found here, juvenile body size is also generally under strong viability selection across taxa[31]. Together, this suggests that previous heritability estimates of juvenile size are likely to have been systematically over-estimated, especially as a large proportion are based on PO-regressions[61]. Indeed, tarsus length heritability estimates from PO-regressions have been shown to be consistently larger than those from animal models[61]. Juvenile size is a hallmark trait of evolutionary stasis, whereby traits that should respond to selection in the wild appear not to. Although these results do not fully explain this stasis, they do show that the predicted response to selection may be being substantially overestimated in traits with non-Gaussian phenotypic distributions.

Lande-Arnold regression is by far the most common method for estimating β[5;33;62] and is known to be unbiased in the presence of phenotypic skew only if the fitness function is linear or quadratic *and* this quadratic term is modelled[22]. Although the estimated survival functions deviated from a quadratic for all traits, estimates of β were close to those that would have been obtained under Lande-Arnold regression including the quadratic term (β_2) for all traits, and without the quadratic term (β_1) for three traits. The near equivalence of these different estimates seems at odds with the conclusions of Bonamour *et al.*[17], who demonstrate that selection gradients approximated with Lande-Arnold regression are biased in the presence of phenotypic skew. However, Bonamour *et al.* only modelled the linear term in the Lande-Arnold regression (β_1) whilst assuming a quadratic fitness function - had the quadratic term also been included, the linear term in the Lande-Arnold regression (β_2) would have been unbiased ([22],[3] Chapter 29), in correspondence with our wing length results (β_1 underestimated β, but β_2 did not). However, there is no reason to believe including a quadratic term in a Lande-Arnold regression will generally result in a good approximation of β. Indeed, Morrissey & Sakrejda[5] compared β with that approximated from a quadratic Lande-Arnold regression and found quite large proportional differences (approx. 30%), although small differences in absolute terms.

We therefore urge caution in assuming that our results are a general statement about the accuracy of Lande-Arnold regression under non-normality.

Quantitative genetics uses two main frameworks to predict how traits will respond to selection. Here we demonstrate how both of these frameworks are affected by skew at the environmental and genetic levels. Genetic skew can lead both the breeder's equation and Lande's gradient equation to be inaccurate. Although little or no genetic skew has been found in the few studies that have tried to quantify it, it remains unknown to what extent this is a generality, and will be highly dependent on the genetic architecture of specific traits. In the absence of genetic skew, the gradient equation presents an accurate prediction of selection response[11], although environmental skew provides challenges to the accurate estimation of both β and V_A. Whilst the breeder's equation may provide a more intuitive way of thinking about selection response, the extensions to this framework that allow for non-linearity[12] are complex and computationally expensive. We therefore recommend a focus on the gradient equation (and its extensions[11]) in wild systems, where fitness functions are highly likely to be non-linear and trait distributions are commonly skewed.

Methods

This study was preregistered (see https://osf.io/7qyp4/). We have highlighted in the following sections where our methods deviate from those planned.

Meta-analysis of Skew

We collected raw data on juvenile and adult tarsus length from several sources: we used a mailing list to request data, we searched the dryad repository for 'tarsus', we emailed groups with known long-term avian datasets that were not represented in these sources and included any tarsus length data that we otherwise encountered. When datasets from different studies of the same population overlapped in time, we use the largest single dataset available. Datasets were taken from [44;63–100].

Sample standardised skew was estimated from raw data z as

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(z_i-\hat{\mu})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\hat{\mu})^2\right]^{3/2}}\frac{\sqrt{n(n-1)}}{n-2}$$

with sampling variance as

$$\frac{6n(n-1)}{(n-2)(n+1)(n+3)}$$

where n is sample size and $\hat{\mu}$ the estimate of the trait mean [101].

Using these data, we ran a random-effect meta-analytic model in MCMCglmm with age (juvenile or adult) as a fixed factor and random effects of species and study. Models were run for 65000 iterations, with a burnin of 15000 and a thinning intervals of 50. The priors for the random-effect variances were scaled (by 100) $F_{1,1}$ and the prior for the residual variance was inverse-gamma with a shape and scale of 0.001. The fixed effects had a diffuse normal prior (mean=0, variance=10^{10}).

Study population

We used data from a nest-box population of blue tits (*Cyanistes caeruleus*), on the Dalmeny estate, Edinburgh, United Kingdom, collected from 2011 to 2018, with 253 nest-boxes over two sites. Detailed methods are described in [59;102]. Briefly, all nests were visited regularly until the discovery of the first egg, and then daily for egg cross-fostering, when eggs were weighed. From 2011-2013 and 2016-2018 a partial egg cross-fostering design was used to enable additive genetic and nest-of-rearing effects on offspring size to be separated [59]. In 2014-2015 a mixture of full and partial cross-fostering was used as part of a separate experiment. Full details of cross-fostering can be found in [103]. After egg laying was complete, nests were left undisturbed for 11 days and then checked daily for hatching. At hatching (day 0), all chicks were uniquely marked (within a nest). The chicks had blood samples taken at day 3 and were given a unique metal ring at day 9. At day 15, chick's tarsus, wing and head-bill lengths were measured and they were weighed. For the morphometric measurements, one chick from each nest was measured twice in order to account for measurement error [59]. From day 10, adults were

caught at the nest in order to identify them; blood samples and morphometric measurements were taken and the birds were uniquely ringed. At the end of the season we checked all nests and recorded any dead chicks left in the nest. From this we could infer which chicks fledged. Chicks were considered recruited if they were recaptured as breeders in subsequent years. Permission to monitor, catch and ring the birds was given by Scottish Natural Heritage and the British Trust for Ornithology and permission to take blood samples was granted by the UK Government's Home Office. All permission and licenses were granted to JDH.

Social parentage was assigned through catching parents at the nest. When no female was caught, the social female was assigned a dummy mother identity. When no male was caught, the social father was assigned as the genetic sire with the largest proportion of paternity in a nest, either a male caught at a different nest that year, or an unsampled male assigned a dummy identity.

For the assignment of genetic parentage and chick sex, genotypes were obtained using blood and tissue samples from adults and chicks. Genotyping and pedigree reconstruction largely followed protocols outlined in[59] and[102]. However, adults not caught in the focal year but that were known to be alive (because they were caught in subsequent years and were aged 2 years or over) were allowed to be parents of chicks in the focal year. The distance between the nest-of-origin of the chicks and the nest at which these candidate parents were caught in the subsequent year was fitted as a covariate. Mothers were allowed to be polygamous when (half) sib-ships were assigned to chicks with unknown fathers (see Supplementary Methods). When assigning chick sex, we used morphological sexing of recruits over molecular sexing from chicks (sexing didn't match for 5 chicks).

For our analysis we included data on chick size measured at day 15 post-hatching, collected on this project from 2011-2018, and additionally chick recruitment data from 2019 and 2020. We included all nests for which hatching date was known. Although similar morphological data was collected in 2010, we excluded all records from this year as egg size was not measured. Egg size was used to account for nest-of-origin effects in our models (see below). We also excluded data from an additional two nests where egg size was not measured, from chicks for which molecular sexing was not successful (n=20 chicks) and where we did not have one of the day 15 measurements (n=11 chicks). In total, we had records of 5123 day 15 chicks in 715 nests, with 642 chicks repeatedly measured.

Statistical analysis

All models were run in a Bayesian framework. From all models posterior means and 95% credible intervals are presented. A p-value for the fixed effects and covariances in these models was approximated (pMCMC) as two times the smaller number of iterations where the parameter value is either less than zero or greater than zero[104]. We use a threshold of 0.005 to refer to results as significant and those between 0.05 and 0.005 as suggestive[105].

Decomposing phenotypic skew using hierarchical models

We modelled the four traits (tarsus length, head-bill length, mass and wing length) measured at day 15 using linear mixed effects models with sex (2 level factor), year (8 level factor), time of day (continuous - hours from midnight) and egg size (continuous) as fixed. Additive genetic

and nest-of-rearing effects were modelled as random. Because we have repeated measurements of tarsus, wing and head-bill lengths, we additionally modelled measurement error effects in these traits, by including bird identity effects, which are equivalent to the residuals in a model without repeat measures, and the residuals are measurement error effects[59]. In contrast to past analyses[59;102], we do not model nest-of-origin effects but rather include egg size as a covariate to account for these effects (see[59] and Supplementary materials). As estimating skew-t distributed random effects (see below) is parameter heavy, including a covariate rather than a random effect is preferable, especially as nest-of-origin effects are very small for these traits[59;102].

Skew due to the fixed effects was obtained by multiplying the fixed effect design matrix by the fixed effects and estimating the parameters for the skew-t distribution of the resulting variable. These were used when calculating the non-linear parent offspring regression and when plotting the sample skew. This method assumes that the joint distribution of the covariates is equal to the empirical distribution we observe. In combination with a diffuse prior on the fixed effects, this assumption probably leads to a small inflation in the estimated (absolute) skew. Time of day was excluded from this estimate as any skew induced by this is due to our sampling design rather than being biologically relevant.

In order to estimate skew in in the random effects, we fitted random effects with skew-t distributions. The residuals for the repeat measured traits were treated as Gaussian as these represent measurement errors. As with the normal distribution, the skew-t distribution[106–109] has a location ξ and scale ω parameter, but also parameters δ and ν which modify the skew and tailness, respectively. The distribution converges on a normal distribution when $\delta = 0$ and ν approaches infinity. As δ moves away from 0 and ν decreases the (absolute) skew in a variable increases, with the sign of δ signifying the direction of the skew. The skew-t distribution is unbounded and readily allows for considerable amounts of positive and negative skew. The reasons for the use of this distribution are further discussed in the supplementary materials. Our approach to modelling the additive genetic effects is to extend standard quantitative genetic models by allowing the base population breeding values to have a skew-t distribution, with normally distributed Mendelian sampling deviations in the descendants (with variance $\omega^2(1-\overline{F})/2$ where $\overline{F}$ is the average inbreeding coefficient of the individual's parents). This assumes that inheritance occurs under the Gaussian descendants infinitesimal model[35;110]; i.e. the Mendelian sampling deviations are normally distributed within families, and any genetic skew results from selection. In practice, however, the Mendelian sampling deviations are largely confounded with residual effects in our data because there are few parent-offspring comparisons (due to high migration and low recruitment) and so inferences are probably quite robust to any violation of the Gaussian descendants assumption. Initially we tried to fit this model in an animal model framework, but due to poor mixing we chose to approximate the model using a dam-sire model. This model discards information about the Mendelian-sampling deviations and subsumes them in the residual effects which then come from a mixture distribution[111]. Given there is little information in our data about the Mendelian-sampling deviations the dam-sire and animal models are expected to give almost identical answers (see Supplementary Materials). Although this method allows us to directly estimate skew in breeding values, when the environmental residuals are skew-t, as assumed here, the mixture distribution does not have standard from. Here, we approximate the mixture distribution as skew-t and although we cannot derive the full distribution of the environmental residuals we are able to obtain their

variance and skew. These models provided little evidence for genetic skew in any trait and so we reverted to an animal model with normally distributed breeding values - the animal model approach having the advantage that the environmental residual skew can then be directly estimated. The dam and sire effects were modelled in a multi-membership model where the two sets of effects were constrained to having the same skew-t distribution.

Initially, we intended to model chick mass over ontogeny in a multivariate framework (see preregistration), as in previous studies of this population[59;102]. However, implementing the required multivariate skew-t models proved too challenging. Since there is strong directional selection on chick body mass throughout ontogeny[59;102], our estimates of skew at day 15 are likely underestimates as the univariate analysis used will fail to account for selective disappearance prior to day 15[59;102]. We also planned to have a global box-cox parameter in case there was a single transformation that would make everything linear and additive. However, given the problems we had with implementing more complex models, we chose not to include this additional complexity.

It should also be noted that estimates from these skew-t models seem to be more sensitive to unmodelled heteroskedasticity than standard Gaussian mixed effects models, even when skew exists, and this can lead to biased fixed effect and variance estimates. This led us to fit a reduced set of fixed effects compared with previous analyses[59;102] and outlined in our pre-registration (see Supplementary materials). To partly address this issue we also ran equivalent Gaussian models for all skew-t models, and present the results in the Supplementary materials. There were small differences the between models but the results remain qualitatively the same (see SM; Figure S17, Tables S4-15).

These models were run using Stan (version 2.21.0)[112] using the cmdstanr package (Stan Development Team, 2019) in R (version 4). Four chains were run for each model with a warmup of 4000 iterations and 6000 iterations post-warmup, with the exception of the dam-sire wing length model which was run with a warmup of 5000 iterations and 10000 iterations post-warmup. Convergence of individual chains was visually assessed, as well as ensuring that the Gelman–Rubin diagnostic (R-hat) across chains was less than 1.1[113]. We used diffuse normal priors for fixed effects (mean=0 and standard deviation=100), half-Cauchy priors (mean=0 and standard deviation=10) for standard deviations and uniform priors from -1 to 1 for δ and 4 to 40 on ν. The choice of priors is discussed further in the Supplementary materials.

Non-Linear Parent-Offspring Regression

The PO-regression function is defined as $E[z_o|z]$ where z_o is the phenotype of offspring from a parent with phenotype z. Assuming random mating and environmental values in the offspring (e_o) are independent of parental phenotypes this becomes $\frac{1}{2}E[g|z] + \frac{1}{2}E[g] + E[e_o]$ under the Gaussian descendants assumption, where g is breeding value. Have θ_g be the parameters of the breeding value distribution and θ_e the parameters of the environmental distribution. Then,

$$E[g|z] = \frac{\int (z-e)p(z-e|\theta_g)p(e|\theta_e)de}{\int p(z-e|\theta_g)p(e|\theta_e)de} \quad (1)$$

The integrals have to be evaluated numerically, which is time consuming, and so the regression function was evaluated at the posterior mean of the parameters from the skew-t animal models to give $E[z_o|z]$ for each trait (Figure 5). Also, note that in the presence of pre-breeding survival selection, the term $\frac{1}{2}E[g]$ in the intercept of the regression function should be replaced by $\frac{1}{2}(E[g] + \Delta g)$ where Δg is the change in mean breeding value due to selection such that $E[g] + \Delta g$ is the expected breeding value of the other parent[14].

Selection on chick body mass

Given that we were not able to model chick body mass in a multivariate framework, we did not model survival throughout ontogeny as originally planned (see preregistration), but rather modelled survival from day 15 to fledging and fledging to recruitment. We modelled this as an event history in a probit regression (binomial error distribution and probit link function) including a quadratic effect of chick size at day 15 on both events, allowing us to model the stabilising component of selection. These models accounted for measurement error in tarsus, head-bill and wing lengths, using the repeated measurements of these traits. Originally we planned to correct our measurements for time of day effects (see preregistration). However, these effects proved to be very small and for most traits non-significant (see Supplementary Results). We therefore decided not to add this extra complexity into our models.

Sex, day of hatching within the nest, year, clutch size, male presence, nest hatch date were also included as fixed effects. All fixed effects were allowed to differ between the two events. Finally we modelled the 2x2 covariance matrix of nest-of-rearing effects. This model was run using Stan. Four chains were run for each model with 5000 iterations and a warmup of 2500 iterations with a thinning interval of 10. Convergence of chains was assessed as above. Diffuse priors for fixed effects (mean=0 and standard deviation=100), half-Cauchy priors for all standard deviations (mean=0 and standard deviation=10) and LKJ priors on correlations with shape=2[114] were used.

The Individual Relative Fitness Function

Partitioning the linear predictors for each survival event (1: day 15 to fledging, 2: fledging to recruitment) into a part due to the trait and a part due to remaining terms (denoted η), and assuming that the distribution of $\eta^{(1)}$ and $\eta^{(2)}$ are bivariate normal conditional on the trait z, then the absolute fitness function has the form:

$$W(z) = \quad F_{MVN}\left(\mathbf{s}|\mathbf{\Sigma}\right) \tag{2}$$

where F_{MVN} is the multivariate normal cumulative density function in which the first argument is the quantile to be evaluated and the second argument is the (co)variance of the variates (the means are zero and are therefore not given). For event i

$$s^{(i)} = E[\eta^{(i)}] + \frac{COV(\eta^{(i)}, z)}{\mu_2}(z - \mu) + \beta^{(i)}z + \frac{1}{2}\gamma^{(i)}z^2 \tag{3}$$

where $\beta^{(i)}$ and $\frac{1}{2}\gamma^{(i)}$ are the linear and quadratic effect of the trait on event i, μ is the trait mean and μ_i the i^{th} central moment of the phenotypic distribution.

$$\Sigma^{(i,j)} = COV(\eta^{(i)}, \eta^{(j)}) - \frac{COV(\eta^{(i)}, z)COV(\eta^{(j)}, z)}{\mu_2} + COV(u^{(i)}, u^{(j)}) + \delta^{(i,j)} \quad (4)$$

where $u^{(i)}$ are the nest effects for event i and $\delta^{(i,j)} = 1$ when $i = j$ and represents the residual variance.

The partial derivative of $W(z)$ with respect to z is given by

$$\begin{aligned}\frac{\partial W(z)}{\partial z} = \; & f_N\left(s^{(1|2)}|\mathbf{\Sigma}^{(1|2)}\right)\left(\frac{COV(\eta^{(1)},z)^2}{\mu_2} + \beta^{(1)} + \gamma^{(1)}z - \frac{\mathbf{\Sigma}^{(1,2)}}{\mathbf{\Sigma}^{(2)}}\left(\frac{COV(\eta^{(2)},z)^2}{\mu_2} + \beta^{(2)} + \gamma^{(2)}z\right)\right) \\ & F_N\left(s^{(2)}|\mathbf{\Sigma}^{(2)}\right) + f_N\left(s^{(2)}|\mathbf{\Sigma}^{(2)}\right)\left(\frac{COV(\eta^{(2)},z)^2}{\mu_2} + \beta^{(2)} + \gamma^{(2)}z\right)F_N\left(s^{(1|2)}|\mathbf{\Sigma}^{(1|2)}\right)\end{aligned} \quad (5)$$

where f_N and F_N are the density and cumulative density functions for a centred normal distribution, and

$$s^{(1|2)} = s^{(1)} - \frac{\mathbf{\Sigma}^{(1,2)}}{\mathbf{\Sigma}^{(2)}}s^{(2)} \qquad \mathbf{\Sigma}^{(1|2)} = \mathbf{\Sigma}^{(1)} - \frac{(\mathbf{\Sigma}^{(1,2)})^2}{\mathbf{\Sigma}^{(2)}} \quad (6)$$

Solving Equation 5 to find the stationary point(s), and therefore the optimal trait value, is difficult. Instead we evaluated the derivative of Equation 5 at the minimum and maximum observed trait value and assessed whether the derivative at the minimum is positive and negative at the maximum. This condition implies an optimal trait value within the range of observed trait values.

Selection Gradients

The Lande-Arnold method[22] for estimating the selection gradient is only robust to phenotypic skew if the fitness function is quadratic and both the mean-centered trait value and its square are fitted in the regression[3;22]. We therefore computed three selection gradients. Using the notation in[34], we calculated our best estimate of it[115],

$$\beta = E\left[\frac{\partial w(z)}{\partial z}\right] = \int \frac{\partial w(z)}{\partial z}p(z)dz \approx \frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial w(z)}{\partial z}\right|_{z_i} \quad (7)$$

where $p(z)$ is the probability density function for z, $w(z)$ is the relative fitness function obtained by dividing $W(z)$ by mean fitness ($E[W] = \int W(z)p(z)dz$) and z_i are the observed trait values. Put simply, we calculated the mean partial derivative of individual fitness function (from Equation 5) across our observed phenotypic distributions, divided by mean fitness.

The linear selection differential is defined as

$$S = \int zw(z)p(z)dz - \mu \approx \frac{1}{n}\sum_{i=1}^{n} z_i w(z_i) - \hat{\mu} \quad (8)$$

and the quadratic selection differential as

$$C = \int (z-\mu)^2 p(z) w(z) dz - \mu_2 \approx \frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{\mu})^2 w(z_i) - \hat{\mu}_2 \tag{9}$$

From these we can calculate the expected linear regression coefficient from the Lande-Arnold method when only the linear term was fitted:

$$\hat{\beta}_1 = \frac{\hat{S}}{\hat{\mu}_2} \tag{10}$$

and the linear regression coefficient from the Lande-Arnold method when both the linear and quadratic term are fitted (Eq. 29.28a from[3]):

$$\hat{\beta}_2 = \frac{(\hat{\mu}_4 - \hat{\mu}_2^2)\hat{S} - \hat{\mu}_3\hat{C}}{\hat{\mu}_2(\hat{\mu}_4 - \hat{\mu}_2^2) - \hat{\mu}_3^2} \tag{11}$$

Selection cannot operate on between-sex differences in trait values (the average fitness of the two sexes is constrained to be equal) and we assume that selection does not operate on between-year differences in trait values (which might occur if juvenile size impacts on adult survival). We therefore estimated each β as the average of each sex by year combination (Figure 4 e-h), calculated across the posterior distribution of the survival model.

Response to Selection

The extension of Lande's gradient equation to a non-normal distribution of genetic effects is (combining Equations 26 and 42 from[11]):

$$\Delta\mu = \sum_{j=1}^{\infty} K^{j+1}(g)\frac{1}{j!}\int \frac{\partial^j w(z)}{\partial z^j} p(z) dz \tag{12}$$

where $K^j(x)$ denotes the j^{th} cumulant of x, which up to the third cumulant (skew) is

$$\Delta\mu = V_A E\left[\frac{\partial w(z)}{\partial z}\right] + \frac{S_A}{2} E\left[\frac{\partial^2 w(z)}{\partial z^2}\right] \tag{13}$$

where S_A is the skew in the additive genetic effects. When the distribution of additive genetic values is normal and/or the fitness function is linear, Equation 12 reduces to Lande's gradient equation

$$\Delta\mu = V_A E\left[\frac{\partial w(z)}{\partial z}\right] = V_A\beta \tag{14}$$

since all cumulants > 2 of the genetic distribution are zero.

Heritability

We compared how well our inferred non-linear PO-regression (Equation 1) performed at predicting offspring phenotype compared to linear single-parent mid-offspring regression. Using the 182 individuals (118 male and 64 female) that were measured as chicks at day 15 and survived to produce offspring that were also measured at day 15, we fitted a weighted (by family size) regression with our inferred non-linear PO-regression fitted as an offset. We then compared the fit of this model to an identical model but where the raw parental phenotype was also fitted as a covariate with a free parameter.

We then compared estimates of the heritability before and after selection (h_b^2 and h_a^2, respectively). The heritability can be defined as the regression coefficient of a linear mid-PO-regression, and can be calculated before selection

$$h_b^2 = \ 2\frac{COV(z_o,z)}{\mu_2} = \frac{V_A}{V_P} \tag{15}$$

or after selection

$$h_a^2 = \ 2\frac{E[w(z)z_o z]-E[w(z)z_o]E[w(z)z]}{E[w(z)z^2]-E[w(z)z]^2} \tag{16}$$

The posterior distribution of h_b^2 was evaluated directly, but the i^{th} posterior sample of h_a^2 was obtained by simulating 10^4 values of z and z_o using the parameters sampled at the i^{th} iteration of the trait model, calculating expected fitness for each sampled z using the parameters sampled at the i^{th} iteration of the fitness model, and then evaluating the relevant expectations.

Simulations

To test how different sampling designs and standard estimation procedures (PO-regression and Gaussian animal model) impact estimates of heritability in the presence of skew and selection, we simulated data according to the posterior mean of the parameters from our skew-t quantitative genetic and selection models for mass. A closed population with 1000 breeding pairs was simulated over three generations, with 10 measured full-sib offspring per pair. Four scenarios were simulated: either nests were not cross-fostered or they were paired and five offspring reciprocally crossed, and the random effects were either skew t-distributed (with ω, δ and ν parameters set to their posterior means) or they were normally distributed but with matching variance. The probability of a chick recruiting to be a parent was obtained by applying the estimated survival model for chick mass to the simulated phenotype. Each of the four scenarios were simulated 2000 times and for each data set the heritability was estimated directly using PO-regression and as the estimate of the additive genetic variance over the sum of all variances estimated from a Gaussian animal model fitted in ASReml-R[116].

Data availability

All data and code can be found at https://doi.org/10.5281/zenodo.5794316.

Acknowledgements

We thank our many field assistants for help with data collection, Shinchi Nakagawa, Anders Moller, Diego Santiago-Alarcon, Roger Jovani, Sergi Sales and Nuria Rodriguez for providing raw data, Eryn McFarlane, Julie Gauzere and Ed Ivimey-Cook for helpful discussions and two anonymous reviewers for their comments on the manuscript. This work was funded by Natural Environment Research Council (NE/P000924/1) and Royal Society Fellowship to JDH, and supported by Lord Rosebery and Dalmeny estate.

Author contributions

JLP and JDH conceived and designed the project. JLP, HEL, CET and JDH generated the data. JLP and JDH analysed the data and wrote the paper. All authors have read and approved the paper.

Competing interests

The authors declare no competing interests

References

[1] Lush, J. L. *Animal Breeding Plans* (Iowa State College Press, Ames, Iowa, 1937).

[2] Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334 (1976).

[3] Walsh, B. & Lynch, M. *Evolution and Selection of Quantitative Traits* (Oxford University Press, Oxford, UK, 2018).

[4] Schluter, D. Estimating the Form of Natural Selection on a Quantitative Trait. *Evolution* **42**, 849–861 (1988).

[5] Morrissey, M. B. & Sakrejda, K. Unification Of Regression-Based Methods For The Analysis Of Natural Selection. *Evolution* **67**, 2094–2100 (2013).

[6] Falconer, D. & Mackay, T. F. *Introduction to Quantitative Genetics* (Longman, New York, 1996), 4th edn.

[7] Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Inc., Sunderland, MA, 1998).

[8] Roff, D. A. *Evolutionary quantitative genetics* (Springer Science & Business Media, 2012).

[9] Nishida, A. & Abe, T. Non-Linear Heritability and Asymmetrical Selection Responses caused by Skewed Distribution of Breeding Value in Selected Population. *Japanese Journal of Zootechnical Science* **51**, 495–500 (1980).

[10] Barton, N. H. & Turelli, M. Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetical Research* **49**, 157–173 (1987).

[11] Turelli, M. & Barton, N. H. Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* **138**, 913–941 (1994).

[12] Gimelfarb, A. & Willis, J. H. Linearity versus nonlinearity of offspring-parent regression: An experimental study of Drosophila melanogaster. *Genetics* **138**, 343–352 (1994).

[13] Rice, S. H. *Evolutionary theory: mathematical and conceptual foundations* (Sinauer Associates, 2004).

[14] Heywood, J. S. An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. *Evolution* **59**, 2287–2298 (2005).

[15] Jones, A. G., Bürger, R., Arnold, S. J., Hohenlohe, P. A. & Uyeda, J. C. The effects of stochastic and episodic movement of the optimum on the evolution of the G-matrix and the response of the trait mean to selection. *Journal of Evolutionary Biology* **25**, 2210–2231 (2012).

[16] Urban, M. C., Bürger, R. & Bolnick, D. I. Asymmetric selection and the evolution of extraordinary defences. *Nature Communications* **4** (2013).

[17] Bonamour, S., Teplitsky, C., Charmantier, A., Crochet, P. A. & Chevin, L. M. Selection on skewed characters and the paradox of stasis. *Evolution* **71**, 2703–2713 (2017).

[18] Jacquard, A. Heritability: one word, three concepts. *Biometrics* 465–477 (1983).

[19] Charlesworth, B. The Heritability of Fitness. In Bradbury, J. & Andersson, M. (eds.) *Sexual Selection: Testing the Alternatives*, 21–40 (John Wiley & Sons Limited, 1987).

[20] Nishida, A. & Abe, T. Distribution of Genetic and Environmental Effects and Linearity of Heritability. *Canadian Journal of Genetics and Cytology* **16**, 3–10 (1974).

[21] Robertson, A. The non-linearity of offspring-parent regression. In Pollak, E., Kempthorne, . & Bailey Jr, T. B. (eds.) *Proceedings of the International Conference on Quantitative Genetics*, 297–304 (Iowa State University Press, Ames, 1977).

[22] Lande, R. & Arnold, S. J. . The Measurement of Selection on Correlated Characters. *Evolution* **37**, 1210–1226 (1983).

[23] Beardsley, J. P., Bratton, R. & Salisbury, G. The Curvilinearity of Heritability of Butterfat Production. *Journal of Dairy Science* **33**, 93–97 (1950).

[24] Nishida, A. Some Characteristics of Parent-Offspring Regression in Body-Weight of Mus-Musculus at Different Ages. *Canadian Journal of Genetics and Cytology* **14**, 293–303 (1972).

[25] Mäki-Tanila, A. *The Validity of the Heritability Concept in Quantitative Genetics*. Ph.D. thesis, University of Edinburgh (1982).

[26] Gifford, D. R. & Barker, J. S. The nonlinearity of offspring-parent regression for total sternopleural bristle number of Drosophila melanogaster. *Theoretical and Applied Genetics* **82**, 217–220 (1991).

[27] Koerhuis, A. N. Non-normality of egg production distributions in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability. *Livestock Production Science* **45**, 69–85 (1996).

[28] Mbaga, S. H. & Hill, W. G. Linear versus nonlinear offspring-parent regression in unselected random-bred mice. *Journal of Animal Breeding and Genetics* **114**, 299–307 (1997). URL http://www.ncbi.nlm.nih.gov/pubmed/21395825.

[29] McGuigan, K., Van Homrigh, A. & Blows, M. W. Genetic analysis of female preference functions as function-valued traits. *The American Naturalist* **172**, 194–202 (2008).

[30] Reid, J. M. *et al.* Immigration counter-acts local micro-evolution of a major fitness component: Migration-selection balance in free-living song sparrows. *Evolution Letters* **5**, 48–60 (2021).

[31] Rollinson, N. & Rowe, L. Persistent directional selection on body size and a resolution to the paradox of stasis. *Evolution* **69**, 2441–2451 (2015).

[32] Merilä, J., Sheldon, B. & Kruuk, L. Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112**, 199–222 (2001).

[33] Kingsolver, J. G. *et al.* The strength of phenotypic selection in natural populations. *The American Naturalist* **157**, 245–261 (2001).

[34] Geyer, C. J. & Shaw, R. G. Commentary on Lande-Arnold analysis. technical report no. 670 (2008).

[35] Turelli, M. Commentary: Fisher's infinitesimal model: A story for the ages. *Theoretical Population Biology* **118**, 46–49 (2017).

[36] Bulmer, M. G. *The Mathematical Theory of Quantitative Genetics* (Oxford University Press, Oxford,UK, 1980).

[37] Turelli, M. & Barton, N. H. Dynamics of polygenic characters under selection. *Theoretical Population Biology* **38**, 1–57 (1990).

[38] Zeng, Z. B. Genotypic distribution at the limits to natural and artificial selection with mutation. *Theoretical Population Biology* **32**, 90–113 (1987).

[39] Keightley, P. D. & Hill, W. G. Directional selection and variation in finite populations. *Genetics* **117**, 573–582 (1987).

[40] Reed, D. R., Lawler, M. P. & Tordoff, M. G. Reduced body weight is a common effect of gene knockout in mice. *BMC Genetics* **9** (2008).

[41] Fisher, R. A., Immer, F. R. & Tedin, O. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics* **17**, 107–124 (1932).

[42] Gimelfarb, A. Offspring-parent genotypic regression: how linear is it? *Biometrics* **42**, 67–71 (1986).

[43] Santure, A. W. *et al.* Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology* **24**, 6148–6162 (2015). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13452.

[44] Silva, C. N. S. *et al.* Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity* **119**, 197–205 (2017). URL https://www.nature.com/articles/hdy201729.

[45] Becker, P. J., Hegelbach, J., Keller, L. F. & Postma, E. Phenotype-associated inbreeding biases estimates of inbreeding depression in a wild bird population. *Journal of Evolutionary Biology* **29**, 35–46 (2016).

[46] Huisman, J., Kruuk, L. E., Ellisa, P. A., Clutton-Brock, T. & Pemberton, J. M. Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 3585–3590 (2016).

[47] Pemberton, J. M., Ellis, P. E., Pilkington, J. G. & Bérénos, C. Inbreeding depression by environment interactions in a free-living mammal population. *Heredity* **118**, 64–77 (2017).

[48] Hajduk, G. K. *et al.* Inbreeding, inbreeding depression, and infidelity in a cooperatively breeding bird*. *Evolution* **72**, 1500–1514 (2018).

[49] Rausher, M. D. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* **46**, 616–626 (1992).

[50] Alatalo, R. V., Gustafsson, L. & Lundberg, A. Phenotypic Selection on Heritable Size Traits: Environmental Variance and Genetic Response. *The American Naturalist* **135**, 464–471 (1990). URL https://www.journals.uchicago.edu/doi/10.1086/285056.

[51] Koyama, H. Intraspecific competition among higher plants. VIII. Frequency distribution of individual plant weight as affected by the interaction between plants. *J Inst Polytech Osaka Cy University* **7**, 73–94 (1956).

[52] Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford University Press, Oxford, UK, 1997).

[53] Weiner, J. Asymmetric competition in plant populations (1990).

[54] Bassar, R. D. *et al.* The effects of asymmetric competition on the life history of Trinidadian guppies. *Ecology Letters* **19**, 268–278 (2016).

[55] Muller, M. & Groothuis, T. G. Within-clutch variation in yolk testosterone as an adaptive maternal effect to modulate avian sibling competition: evidence from a comparative study. *The American Naturalist* **181**, 125–136 (2013).

[56] Nilsson, J.-A. & Svensson, M. Sibling competition affects nestling growth strategies in marsh tits. *Journal of Animal Ecology* 825–836 (1996).

[57] Gebhardt-Henrich, S. & Van Noordwijk, A. The genetical ecology of nestling growth in the great tit. environmental influences on the expression of genetic variances during growth. *Functional ecology* 469–476 (1994).

[58] Gebhardt-Henrich, S. Heritability of growth curve parameters and heritability of final size: a simulation study. *Growth, development, and aging: GDA* **56**, 23–33 (1992).

[59] Hadfield, J. D., Heap, E. A., Bayer, F., Mittell, E. A. & Crouch, N. M. A. Disentangling genetic and prenatal sources of familial resemblance across ontogeny in a wild passerine. *Evolution* **67**, 2701–13 (2013).

[60] Hadfield, J. D. Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B: Biological Sciences* **275**, 723–734 (2008).

[61] Postma, E. Four decades of estimating heritabilities in wild vertebrate populations: Improved methods, more data, better estimates? In Charmantier, A., Garant, D. & Kruuk, L. E. B. (eds.) *Quantitative Genetics in the Wild*, 16–33 (Oxford University Press, Oxford, 2014).

[62] Dingemanse, N. J., Araya-Ajoy, Y. G. & Westneat, D. F. Most published selection gradients are underestimated: Why this is and how to fix it. *Evolution* 1–13 (2021).

[63] Arct, A., Drobniak, S., Mellinger, S., Gustafsson, L. & Cichon, M. Data from: Parental genetic similarity and offspring performance in blue tits in relation to brood size manipulation (2020). URL https://doi.org/10.5061/dryad.v6r0758.

[64] Bebbington, K. *et al.* Data from: Consequences of sibling rivalry vary across life in a passerine bird (2016). URL https://doi.org/10.5061/dryad.12np0.

[65] Bebbington, K. *et al.* Data from: Telomere length reveals cumulative individual and transgenerational inbreeding effects in a passerine bird (2016). URL https://doi.org/10.5061/dryad.52fp4.

[66] Becker, P. J. J. *et al.* Data from: Mother-offspring and nest mate resemblance but no heritability in early-life telomere length in white-throated dippers (2015). URL https://doi.org/10.5061/dryad.b2v37.

[67] Berzins, L. L., Gilchrist, H. G. & Burness, G. Data from: No assortative mating based on size in black guillemots breeding in the Canadian Arctic (2015). URL https://doi.org/10.5061/dryad.1bm5t.

[68] Caizergues, A. E., Gregoire, A. & Charmantier, A. Data from: Urban versus forest ecotypes are not explained by divergent reproductive selection (2018). URL https://doi.org/10.5061/dryad.tv45802.

[69] Camacho, C., Canal, D. & Potti, J. Data from: Nonrandom dispersal drives phenotypic divergence within a bird population (2014). URL https://doi.org/10.5061/dryad.h22n9.

[70] Class, B. & Brommer, J. Data from: Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations? (2020). URL https://doi.org/10.5061/dryad.zpc866t6d.

[71] Cornell, A., Gibson, K. F. & Williams, T. D. Data from: Physiological maturity at a critical life-history transition and flight ability at fledging (2017). URL https://doi.org/10.5061/dryad.c2n66.

[72] Cox, A. R., Robertson, R. J., Lendvai, A. Z., Everitt, K. & Bonier, F. Data from: Rainy

springs linked to poor nestling growth in a declining avian aerial insectivore (Tachycineta bicolor) (2019). URL https://doi.org/10.5061/dryad.7m41jd8.

[73] DeSimone, J. G., Clotfelter, E. D., Black, E. C. & Knutie, S. A. Data from: Avoidance, tolerance, and resistance to ectoparasites in nestling and adult tree swallows (2017). URL https://doi.org/10.5061/dryad.9bb60.

[74] Dubuc-Messier, G. *et al.* Data from: Gene flow does not prevent personality and morphological differentiation between two blue tit populations (2018). URL https://doi.org/10.5061/dryad.31tc3s8.

[75] Grunst, M. L., Raap, T., Grunst, A. S., Pinxten, R. & Eens, M. Data from: Artificial light at night does not affect telomere shortening in a developing free-living songbird: a field experiment (2019). URL https://doi.org/10.5061/dryad.8216g63.

[76] Husby, A., Schielzeth, H., Forstmeier, W., Gustafsson, L. & Qvarnström, A. Data from: Sex chromosome linked genetic variance and the evolution of sexual dimorphism of quantitative traits (2012). URL https://doi.org/10.5061/dryad.451n7.

[77] Ihle, M. *et al.* Data from: Rearing Success Does Not Improve With Apparent Pair Coordination in Offspring Provisioning (2019). URL https://zenodo.org/record/3459642.

[78] Jacob, S. *et al.* Data from: Microbiome affects egg carotenoid investment, nestling development and adult oxidative costs of reproduction in Great tits (2015). URL https://doi.org/10.5061/dryad.9n741.

[79] Krause, E. T., Krüger, O. & Schielzeth, H. Data from: Long-term effects of early nutrition and environmental matching on developmental and personality traits in zebra finches (2018). URL https://doi.org/10.5061/dryad.6j700.

[80] Krist, M., Janča, M., Edme, A. & Dzuro, R. Data from: Are prenatal maternal resources more important in competitive than in benign postnatal environments? (2016). URL https://doi.org/10.5061/dryad.823f0.

[81] Krist, M., Remeš, V., Uvírová, L., Nádvorník, P. & Bureš, S. Data from: Egg size and offspring performance in the collared flycatcher (Ficedula albicollis): a within-clutch approach (2010). URL https://doi.org/10.5061/dryad.1758.

[82] Kvalnes, T. *et al.* Data from: Offspring fitness and the optimal propagule size in a fluctuating environment (2018). URL https://doi.org/10.5061/dryad.m74c7m9.

[83] Kvalnes, T. *et al.* Data from: Reversal of response to artificial selection on body size in a wild passerine (2017). URL https://doi.org/10.5061/dryad.v50r8.

[84] Moiron, M. *et al.* Data from: Functional relations between body mass and risk-taking behavior in wild great tits (2018). URL https://doi.org/10.5061/dryad.14cn58v.

[85] Nishida, Y. & Takagi, M. Data from: Song performance is a condition-dependent dynamic trait honestly indicating the quality of paternal care in the Bull-headed Shrike (2018). URL https://doi.org/10.5061/dryad.c84f7c4.

[86] Nord, A. & Nilsson, J.-A. Data from: Incubation temperature affects growth and energy metabolism in blue tit nestlings (2011). URL https://doi.org/10.5061/dryad.jb314.

[87] Pap, P. L. *et al.* Data from: Selection on multiple sexual signals in two Central- and Eastern-European populations of the barn swallow (2019). URL https://doi.org/10.5061/dryad.64p7k2f.

[88] Perrier, C., Delahaie, B. & Charmantier, A. Data from: Heritability estimates from genome wide relatedness matrices in wild populations: application to a passerine, using a small sample size (2018). URL https://doi.org/10.5061/dryad.k6r1mk8.

[89] Podofillini, S. *et al.* Data from: Benefits of extra food to reproduction depend on maternal condition (2020). URL https://doi.org/10.5061/dryad.5db0168.

[90] Poissant, J., Morrissey, M. B., Gosler, A. G., Slate, J. & Sheldon, B. C. Data from: Multivariate selection and intersexual genetic constraints in a wild bird population (2016). URL https://doi.org/10.5061/dryad.qt745.

[91] Poorboy, D. *et al.* Data from: Experimental cross-fostering of eggs reveals effects of territory quality on reproductive allocation (2018). URL https://doi.org/10.5061/dryad.h8v8157.

[92] Rioux Paquette, S., Pelletier, F., Garant, D. & Bélisle, M. Data from: Severe recent decrease of adult body mass in a declining insectivorous bird population (2014). URL https://doi.org/10.5061/dryad.67t23.

[93] Sakaluk, S. K. *et al.* Data from: Genetic and environmental variation in condition, cutaneous immunity, and haematocrit in house wrens (2014). URL https://doi.org/10.5061/dryad.jk2m0.

[94] Simpson, R. K. & McGraw, K. J. Data from: Multiple signaling in a variable environment: expression of song and color traits as a function of ambient sound and light (2017). URL https://doi.org/10.5061/dryad.1j81k.

[95] Song, Z. *et al.* Data from: Silver spoon effects of hatching order in an asynchronous hatching bird (2018). URL https://doi.org/10.5061/dryad.184c1dj.

[96] Torres, R., Chin, E., Rampton, R. & Williams, T. D. Data from: Are there synergistic or antagonistic effects of multiple maternally-derived egg components (antibodies and testosterone) on offspring phenotype? (2019). URL https://doi.org/10.5061/dryad.j348s75.

[97] Vermeulen, A., Müller, W. & Eens, M. Data from: Vitally important – does early innate immunity predict recruitment and adult innate immunity? (2016). URL https://doi.org/10.5061/dryad.p0s3g.

[98] Weber, B. M. *et al.* Data from: Pre- and post-natal effects of experimentally manipulated maternal corticosterone on growth, stress reactivity, and survival of nestling house wrens (2019). URL https://doi.org/10.5061/dryad.16049f4.

[99] Santiago-Alarcon, D. & Parker, P. G. Sexual Size Dimorphism and Morphological

Evidence Supporting the Recognition of two Subspecies in the Galápagos Dove. *The Condor* **109**, 132–141 (2007).

[100] Santos, E. S. A. & Nakagawa, S. Breeding Biology and Variable Mating System of a Population of Introduced Dunnocks (Prunella modularis) in New Zealand. *PLOS ONE* **8**, e69329 (2013).

[101] Joanes, D. N. & Gill, C. A. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47**, 183–189 (1998). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9884.00122.

[102] Thomson, C. E. *et al.* Selection on parental performance opposes selection for larger body size in a wild population of blue tits. *Evolution* **71**, 716–732 (2017).

[103] Thomson, C. E. & Hadfield, J. D. No evidence for sibling or parent–offspring coadaptation in a wild population of blue tits, despite high power. *Evolution* **73**, 28–41 (2019).

[104] Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* **59**, 390–412 (2008).

[105] Benjamin, D. J. *et al.* Redefine statistical significance. *Nature Human Behaviour* **2**, 6–10 (2018).

[106] Branco, M. D. & Dey, D. K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113 (2001).

[107] Azzalini, A. & Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **65**, 367–389 (2003).

[108] Arellano-Valle, R. B. & Azzalini, A. On the Unification of Families of Skew-normal Distributions. *Scandinavian Journal of Statistics* **33**, 561–574 (2006).

[109] Azzalini, A. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs (Cambridge University Press, 2013).

[110] Barton, N. H., Etheridge, A. M. & Véber, A. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* **118**, 50–73 (2017).

[111] Quaas, R. L. & Pollak, E. J. Mixed model methodology for farm and ranch beef cattle testing programs. *journal of Animal Science* **51**, 1277–1287 (1980).

[112] Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **76** (2017).

[113] Gelman, A., Rubin, D. B. *et al.* Inference from iterative simulation using multiple sequences. *Statistical science* **7**, 457–472 (1992).

[114] Lewandowski, D., Kurowicka, D. & Joe, H. Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis* **100**, 1989–2001 (2009).

[115] Janzen, F. J. & Stern, H. S. Logistic regression for empirical studies of multivariate selection. *Evolution* **52**, 1564–1571 (1998).

[116] Butler, D., Cullis, B. R., Gilmour, A. R., Gogel, B. J. & Thompson, R. *ASReml-R Reference Manual Version 4* (VSN International Ltd, Hemel Hempstead, UK, 2017).

Figures and Tables

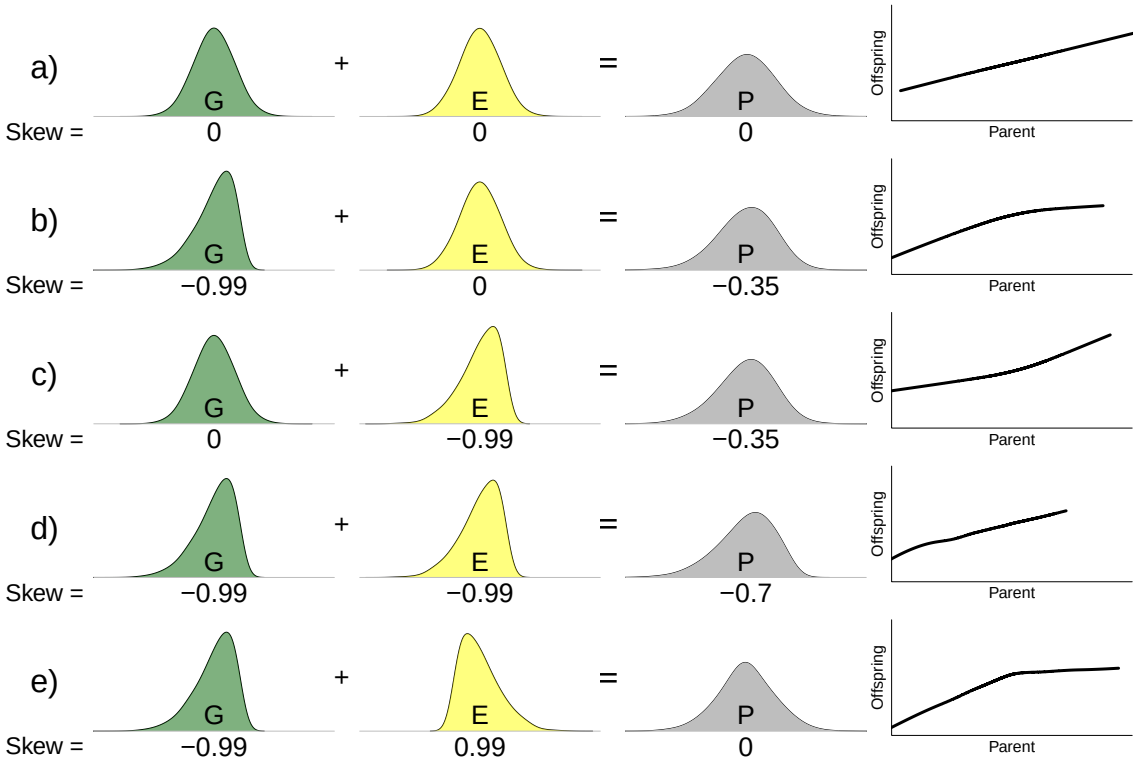

Figure 1: The effects of different distributions of breeding values (G) and environmental values (E) on the distribution of phenotypes (P) and the shape of the PO-regression. When both genetic and environmental values are normally distributed (a), as typically assumed, there is a linear PO-regression. Negative genetic (b) and environmental (c) skew affect the shape of the parent-offspring relationship in opposite directions, whilst inducing the same phenotypic skew. If genetic and environmental distributions are skewed in the same direction (d) their effects on the parent-offspring relationship can cancel each other out, giving a linear parent-offspring relationship, despite considerable phenotypic skew. If genetic and environmental are skewed in opposite directions (e), although they may can cancel each other out at the phenotypic level, they induce a highly non-linear parent-offspring relationship. 1-5) are all simulated with a heritability (V_A/V_P) of 0.5.

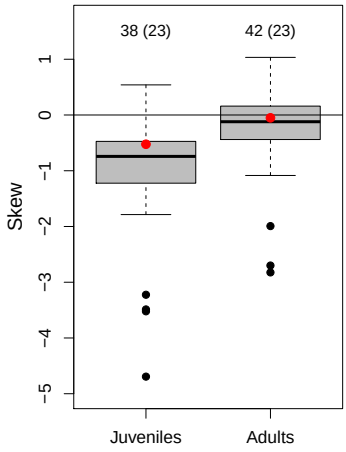

Figure 2: Skew in the distribution of avian tarsus lengths across different species, measured as the coefficient of skew. In the boxplots, the center line shows the median; box limits show upper and lower quartiles; whiskers show 1.5x interquartile range; points show outliers. Numbers above the plots show the number of estimates, and species in parenthesis. The red points show the skew in our blue tit data.

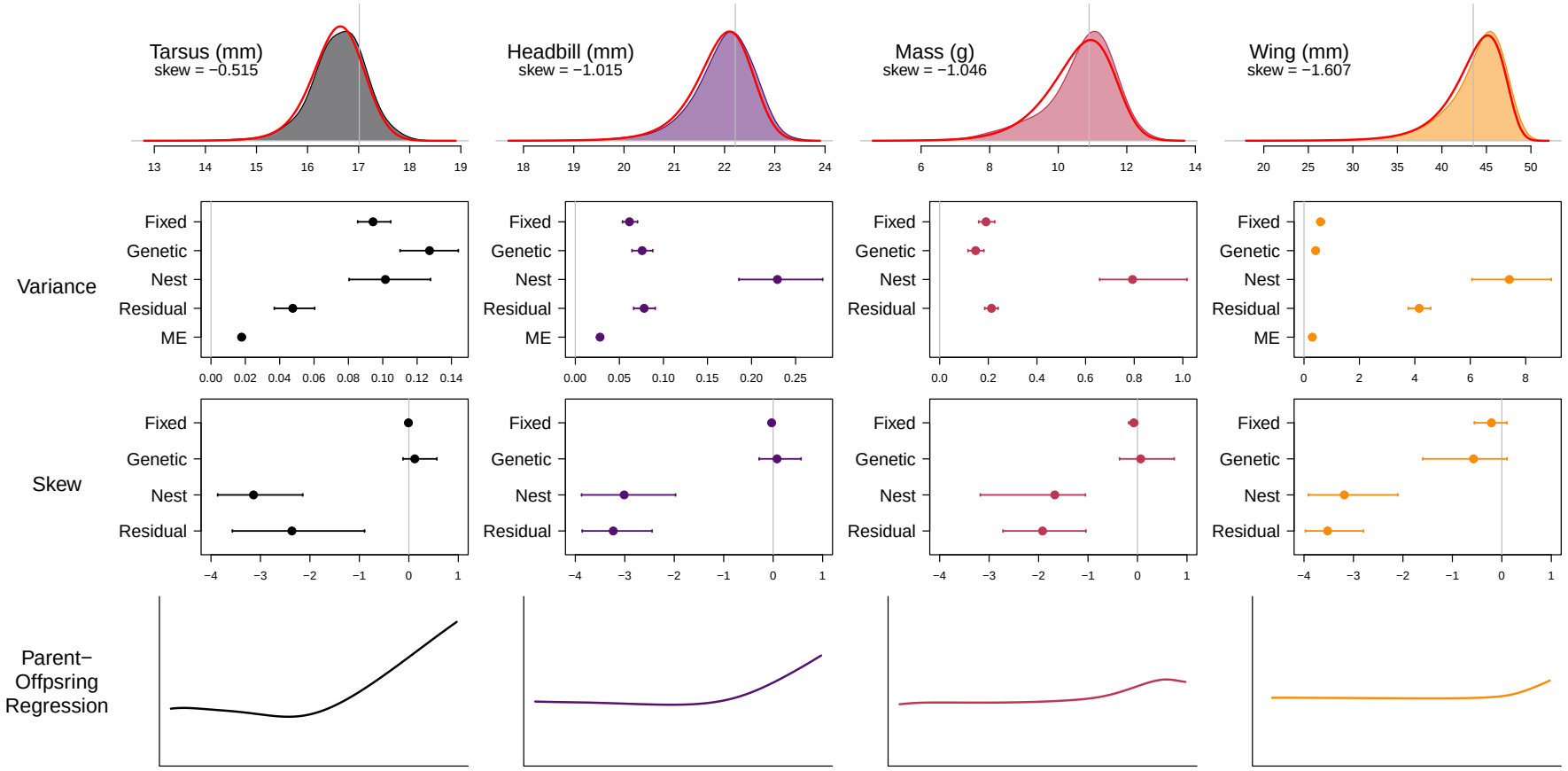

Figure 3: Decomposition of variance and skew in juvenile body size traits in blue tits. Top plots shows the phenotypic distribution of the traits, with the red line showing the distribution predicted from the skew models. The middle rows show the variance and skew (top and bottom, respectively) for each component for all four traits, with all model estimates coming from the skew-t animal model, except the genetic skew which was estimated in the skew-t dam-sire model (see methods). ME stands for measurement error. The bottom row shows the predicted shape of the PO-regression based on the model estimates.

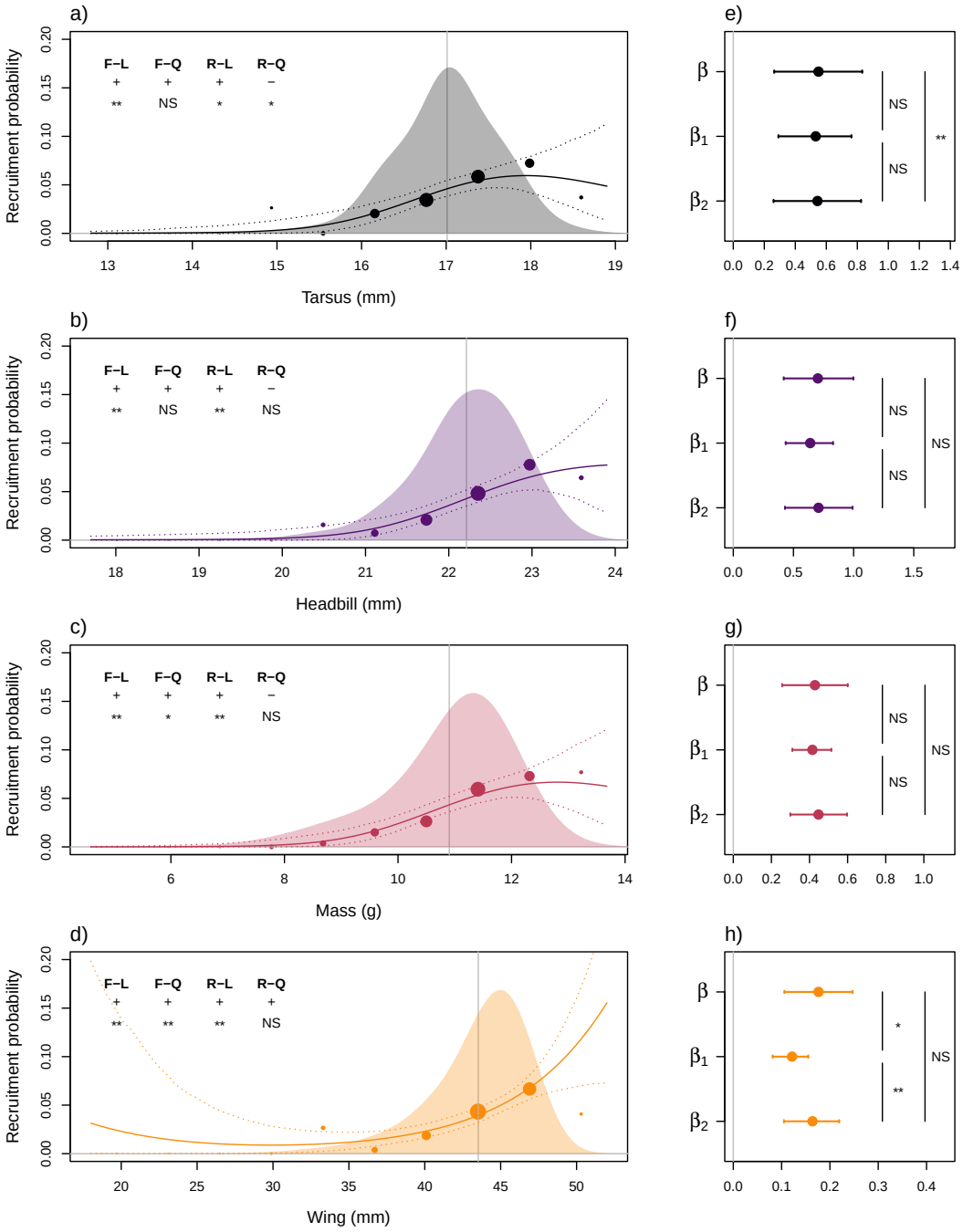

Figure 4: Average (over years and sexes) fitness functions (a-d) and selection gradients (e-h) for tarsus length, head-bill length, mass and wing length, respectively, from day 15 to recruitment. In plots a-d, solid lines show the posterior mean fitness functions, dotted lines show the 95% credible intervals, and points show the average survival of measured individuals from day 15 to recruitment in equally spaced intervals. The size of the points is proportional to the square root of the sample size. The phenotypic distribution of the traits is shown, with the grey vertical line showing the phenotypic mean. The direction and significance of the effect of the trait on fitness is also shown, 'F' and 'R' are survival from day 15 to fledging and from fledgling to recruitment respectively, and 'L' and 'Q' and linear and quadratic effects. In plots e-h, β refers to the selection gradient derived from this fitness function, β_1 and β_2 refer to the approximations from the Lande-Arnold regression excluding and including a quadratic term, respectively. In all plots 'NS' indicates $p > 0.05$, '*' indicates $0.05 > p > 0.005$ and '**' $p < 0.005$.

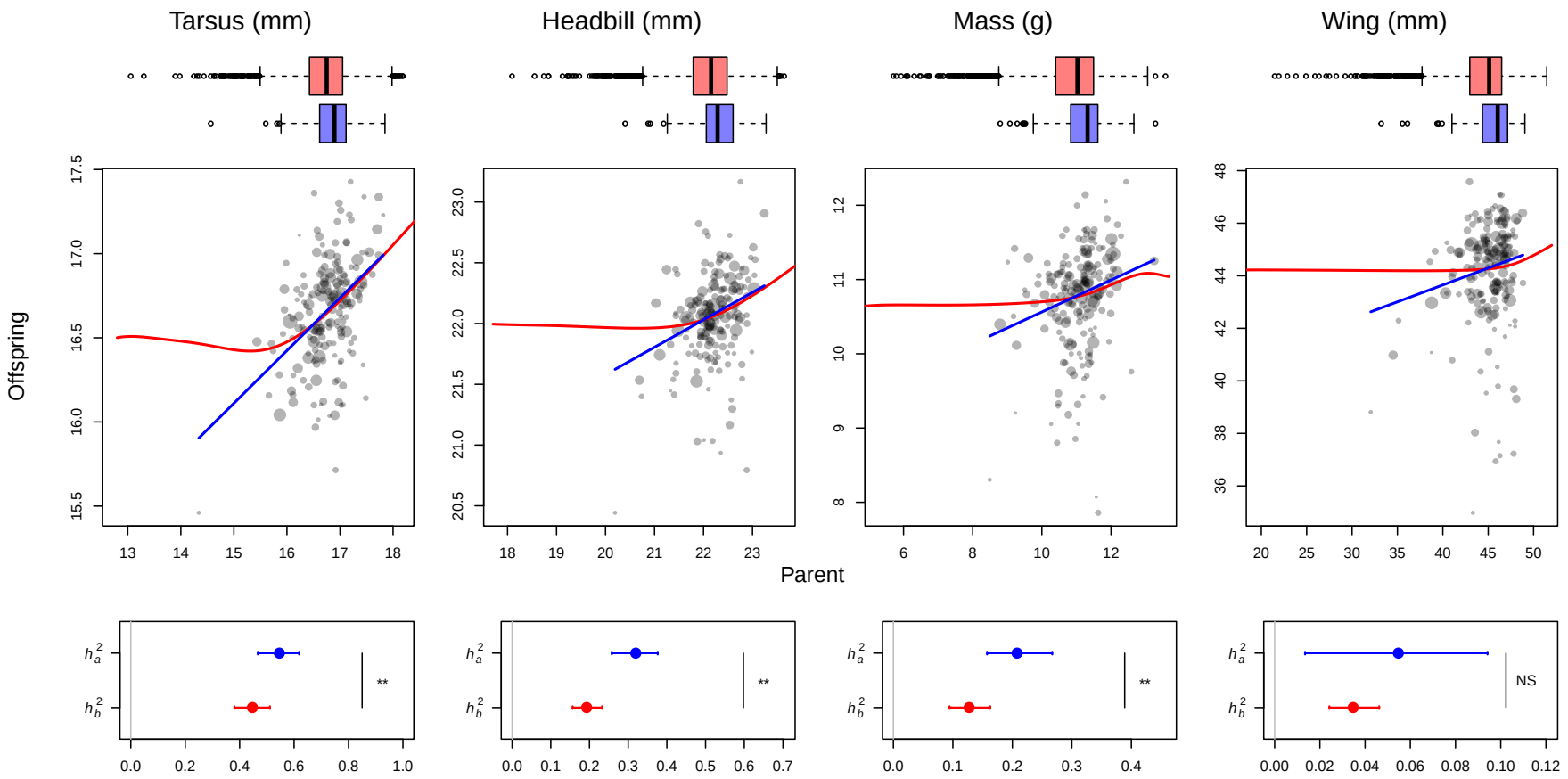

Figure 5: PO-regressions for four body size traits. Top panels show distribution of all chicks (red) and those that survived to recruit (blue), representing the distribution of potential parents before and after selection, respectively. Scatter plots show mid-offspring versus single parental traits. Values are corrected for year, sex and time of day at which they were measured, and the size of the points is proportional to the square root of the family size. The red line is the predicted non-linear PO-regression based on the posterior means of the parameters from the skew-t quantitative genetic model and the blue line is the fit of a weighted (by family size) linear regression to the actual data. These are not corrected for measurement error. Lower panels show the comparison between heritabilities calculated before (h_b^2) and after (h_a^2) selection, calculated across the posterior distribution of the skew-t animal model trait models. In these lower plots all heritabilities account for measurement error. In all plots 'NS' indicates $p > 0.05$, '*' indicates $0.05 > p > 0.005$ and '**' $p < 0.005$.

Table 1: Estimates (mean± SE) of heritability and additive genetic variance from PO-regression and Gaussian animal models (AM) across 2000 simulated data sets. Three-generation simulations were set up with either no cross-fostering (N) or with nests paired and half of each nest's offspring reciprocally crossed (X). Phenotypes were simulated according to the model estimated for chick mass exactly (skewed) or as Gaussian with matching variance. The probability of a chick recruiting to be a parent was obtained by applying the estimated survival model for chick mass.

	Simulated	Gaussian		Skewed	
		N	X	N	X
h^2 PO	0.138	0.139±0.001	0.140±0.001	0.223±0.001	0.224±0.001
h^2 AM	0.138	0.137±0.001	0.135±0.000	0.170±0.001	0.141±0.000
V_A AM	0.148	0.146±0.001	0.144±0.000	0.181±0.001	0.150±0.000