

Ecological causes of multilevel covariance between size and first-year survival in a wild bird population

Sandra Bouwhuis^{1,2*}, Oscar Vedder^{1,2}, Colin J. Garroway¹ and Ben C. Sheldon¹

¹Department of Zoology, Edward Grey Institute, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; and

²Institute of Avian Research, An der Vogelwarte 21, 26386 Wilhelmshaven, Germany

Summary

1. Estimates of selection in natural populations are frequent but our understanding of ecological causes of selection, and causes of variation in the direction, strength and form of selection is limited.

2. Here, we apply a multilevel framework to partition effects of great tit fledging mass on first-year survival to hierarchical levels and quantify their ecological dependence using a data set spanning 51 years.

3. We show that estimates of the effect of fledging mass on first-year survival decline three-fold from year- to brood- to individual level, so that estimates of selection depend strongly on the level at which they are calculated.

4. We identify variables related to summer and winter food availability as underlying higher-level effects of fledging mass on first-year survival and show experimentally that brood-level effects originate early in development. Further, we show that predation and conspecific density modulate individual-level effects of fledging mass on first-year survival.

5. These analyses demonstrate how correlations between traits, fitness and environment influence estimates of selection and show how partitioning trait effects between levels of selection and environmental factors is a promising approach to identify potential agents of selection.

Key-words: ecological causes of selection, fledging mass, great tit, local survival, multilevel selection, *Parus major*, recruitment

Introduction

Natural selection is the process by which phenotypic trait distributions change when an environmental agent causes phenotypic trait variation to be related to fitness variation among individuals (Endler 1986). One widely used approach to quantify natural selection consists of regressing standardized trait values on measures of relative fitness, in either a univariate or multivariate framework, to obtain effect sizes that can be compared between populations or species (Arnold & Wade 1984a,b). From the use of these methods, estimates of the direction, strength and form of selection in natural populations have become plentiful (e.g. Kingsolver *et al.* 2001, 2012; Siepielski, DiBattista & Carlson 2009; Siepielski *et al.* 2013).

Despite the large number of estimates of selection, our understanding of why estimates take the direction, strength and form that they do is rather limited (Endler 1986; Wade & Kalisz 1990; MacColl 2011). We generally

do not know which environmental agents cause the link between phenotypic trait and fitness variation or the extent to which our estimates of selection are generated from (other) environmental factors acting as independent sources of variation in both the trait and fitness (Fig. 1a, also see Kingsolver *et al.* 2012). Making this distinction, and understanding the ecology that underlies our selection estimates, is important if we are to understand spatiotemporal variation in selection (MacColl 2011), or predict responses to selection (Morrissey, Kruuk & Wilson 2010), especially in relation to environmental change (Chevin, Lande & Mace 2010). Investigations of the extent to which estimates of selection are affected by environmental sources of correlated effects, and studies aiming to pinpoint actual agents of selection which underlie a causal relationship between traits and fitness, are therefore timely.

Two main methods have been proposed to study environmental aspects of selection. First, one could experimentally manipulate environmental factors (preferably in combination with the phenotypic trait of interest) and observe how the manipulation affects the direction,

*Correspondence author. E-mail: sandra.bouwhuis@ifv-vogelwarte.de

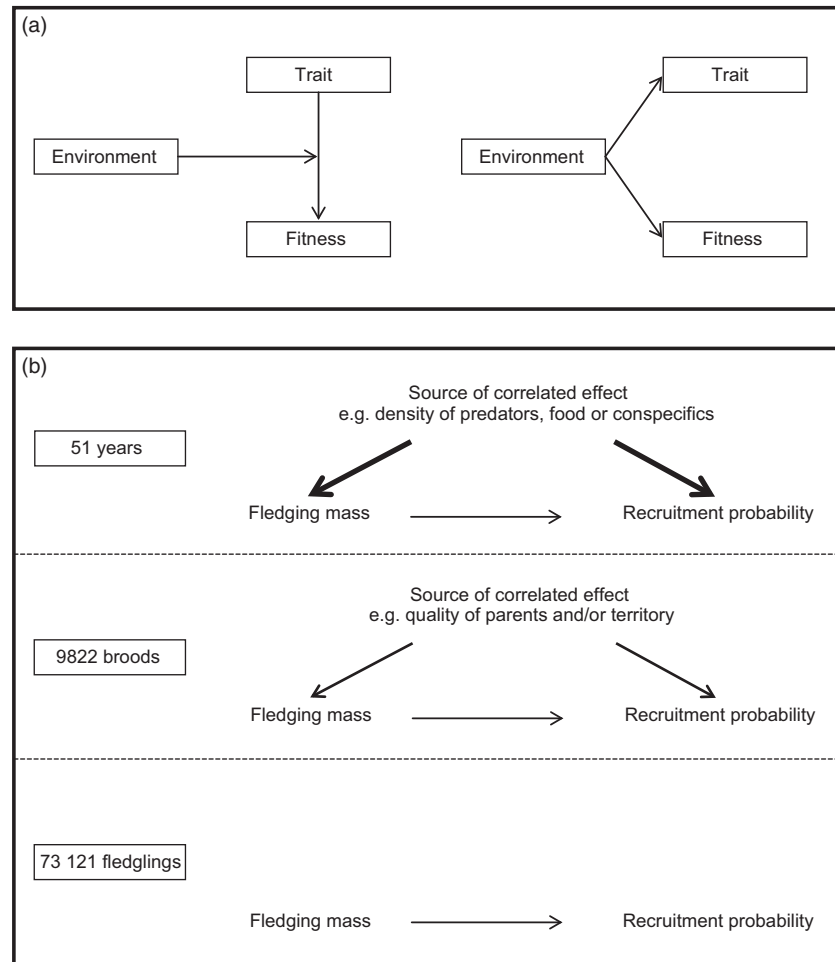


Fig. 1. Schematic overview of predicted origins (a) and sources (b) of the relationship between fledging mass and first-year survival (i.e. recruitment) probability at three hierarchical levels: between years ('year' level), between broods within years ('brood' level) and between offspring within broods ('individual' level).

strength or form of selection on a trait (see Calsbeek & Smith 2007 for an example). Secondly, one could use natural spatial or temporal environmental variation and ask whether selection gradients or differentials differ depending on environmental factors, in populations where experimental manipulation is not desirable or feasible (Wade & Kalisz 1990; MacColl 2011). We advocate a variant of the latter approach and suggest adding environmental factors as independent variables to trait–fitness regressions. If an environmental factor affects the trait and fitness independently, then adding the factor will control for this effect and the adjusted parameter estimate for the effect of the phenotypic trait on fitness is unbiased with respect to the environmental factor. Alternatively, when the shape of the relationship between the trait and fitness depends on the environmental factor, as evidenced by a significant interaction between the environmental factor and trait, this provides the best non-experimental support for the environmental factor's role as an agent of selection.

An advantage of the correlative regression approach described above is that it can easily be combined with a multilevel analysis of selection (Goodnight, Schwartz & Stevens 1992; Goodnight 2012). This will often be important because natural populations are generally characterized by a multilevel structure in which individuals are

found in subunits or different contexts, which may result in different selective processes occurring in different parts of the population (Heisler & Damuth 1987). Within groups of water striders *Aquarius remigis*, for example, male aggression is associated with increased mating success and selected for, while active female dispersal to less aggressive groups causes a negative between-group association between male aggression and mating success (Eldakar *et al.* 2010). As another example, predators may (i) select the lightest individuals in a group [because light individuals may be more sensitive to starvation and take more risks while foraging, or less able to escape predation due to being in worse condition (Lima 1998)] but be randomly attracted to groups, (ii) be attracted to on average lighter groups due to a greater expression of conspicuous behaviour of these groups but randomly take individuals from such groups or (iii) be attracted to on average lighter groups and non-randomly take individuals from such groups. Besides resulting in a different direction or form of selection, as in these examples, selective processes at different levels may also be driven by different agents of selection, or be differentially affected by environmental sources of correlated effects. To our knowledge, no study has yet addressed multiple environmental aspects of selection in a multilevel framework.

Here, we present a case study of a well-established correlation between a measure of size (quantified as body mass at fledging) and first-year survival (i.e. recruitment) probability in a passerine bird species, the great tit *Parus major*. Great tit fledging mass is a phenotypically plastic trait, sensitive to changes in brood size (e.g. Tinbergen & Boerlijst 1990) and breeding density (e.g. Garant *et al.* 2004) that has frequently been reported to be under strong positive selection (Perrins 1965; Garnett 1981; Tinbergen & Boerlijst 1990; Lindén, Gustafsson & Pärt 1992; Both, Visser & Verboven 1999; Visser & Verboven 1999; Monrós, Belda & Barba 2002; Garant *et al.* 2004; Royle *et al.* 2012). Variation in fledging mass occurs at many levels, but we focus here on three levels common to many phenotypic traits measured in natural populations: (i) between years (year level), (ii) between broods within years (brood level) and (iii) between offspring within broods (individual level) (Fig. 1b). In the absence of different environmental sources of correlated effects, or different selection pressures between levels, we would expect the slope of the mass–fitness relationship to be identical at each of these levels (Tinbergen & Boerlijst 1990; also see Fig. 2 in Goodnight, Schwartz & Stevens 1992). Indeed, the single study that we are aware of to have previously compared the correlation between fledging mass and first-year survival on two of these three levels found estimates of the mass–survival correlations at the brood- and individual level to be indistinguishable (Tinbergen & Boerlijst 1990). This finding was explained by assuming the trait–fitness relationship to be causal, that is, by assuming that fledging mass *per se* affects survival and that any brood-level effect is an indirect consequence of lighter individuals being over-represented in on average lighter broods (Tinbergen & Boerlijst 1990).

We agree that within broods, the scope for correlated effects of environmental variables on fledging mass and survival is smallest and the slope of a mass–fitness relation at this level will therefore be closest to representing the causal survival benefit that one extra gram of fledging mass conveys for an individual, independent of its other natal characteristics. Between broods, there is more scope for correlated effects of environmental variables on fledging mass and survival. Experienced parents may, for example, be able to fledge heavy chicks, which have good survival probabilities not just because they are heavy (i.e. as predicted by the individual-level slope) but also because they are, for instance, guided to better post-fledging (e.g. parasite-poor, predator-poor or food-rich) environments. As pointed out with the predation example above, however, the causal part of the mass–fitness slope may also be larger on the between-brood level if there is an additional effect of average brood mass on predator attraction, for example because light broods produce more begging noises and are easier to locate. On the other hand, at the year level, any correlation between fledging mass and first-year survival with a slope different from the one on the brood- or individual level is most likely entirely due

to correlated responses to environmental variables, such as predation pressure, food availability or the density of conspecifics, because competition for first-year survival will not occur between individuals from different years. Besides the correlation between fledging mass and first-year survival probability having seldom been compared between the levels described here, agents of selection on great tit fledging mass have also rarely been identified (Vedder, Bouwhuis & Sheldon 2014), although food availability and predator density have been suggested as likely candidates (Garnett 1981; Adriaensen *et al.* 1998).

In this study, we employ two approaches. First, we analyse a large longitudinal data set spanning 51 years, using a within-subject centring technique, implemented in a mixed model framework (van de Pol & Wright 2009) to distinguish, and directly compare, estimated effects of traits and environmental factors on fitness. Secondly, we make use of a reciprocal cross-fostering experiment to test whether the brood-level correlation between fledging mass and first-year survival probability can be attributed to the environment experienced before or after 14 days of offspring age. The former will include all causal mass and parental or environmental effects that could operate through the egg, or early parental care, whereas the latter will relate to direct effects of the nest-location on post-fledging survival, as well as effects of post-fledging parental care. Identifying the period during which a correlation arises will aid the identification of environmental factors underlying it. Combined, our analyses demonstrate how correlations between traits, fitness and environment influence estimates of selection and show how partitioning trait effects between levels of selection and environmental factors is a promising approach to identify agents of selection.

Materials and methods

STUDY POPULATION AND DATA COLLECTION

The great tit is a small passerine bird abundant in European woodlands and gardens. As a hole-nester, it readily accepts nest boxes for breeding, which allows monitoring of the whole breeding population if an excess of nest boxes is provided (Perrins 1979). The data we analyse here come from a long-term study population in the c. 385 ha mixed deciduous woodland of Wytham Woods, Oxfordshire, UK. Every breeding season, nest boxes were checked at least weekly to obtain information on clutch initiation date, clutch size, hatching date, parental identity and breeding success (Perrins 1965). Chicks were weighed to the nearest 0.1 g and ringed with individually numbered metal rings when the oldest chick in the nest was 15 days of age, which is when mass has usually approached an asymptote and can therefore be classified as fledging mass (van Balen 1973). First-year survival probability in relation to fledging mass was assessed as local recruitment, determined from observing locally hatched birds as parents in subsequent years. Such local recruitment underestimates actual survival, due to emigration of offspring from the woods, but emigration from Wytham was shown to be

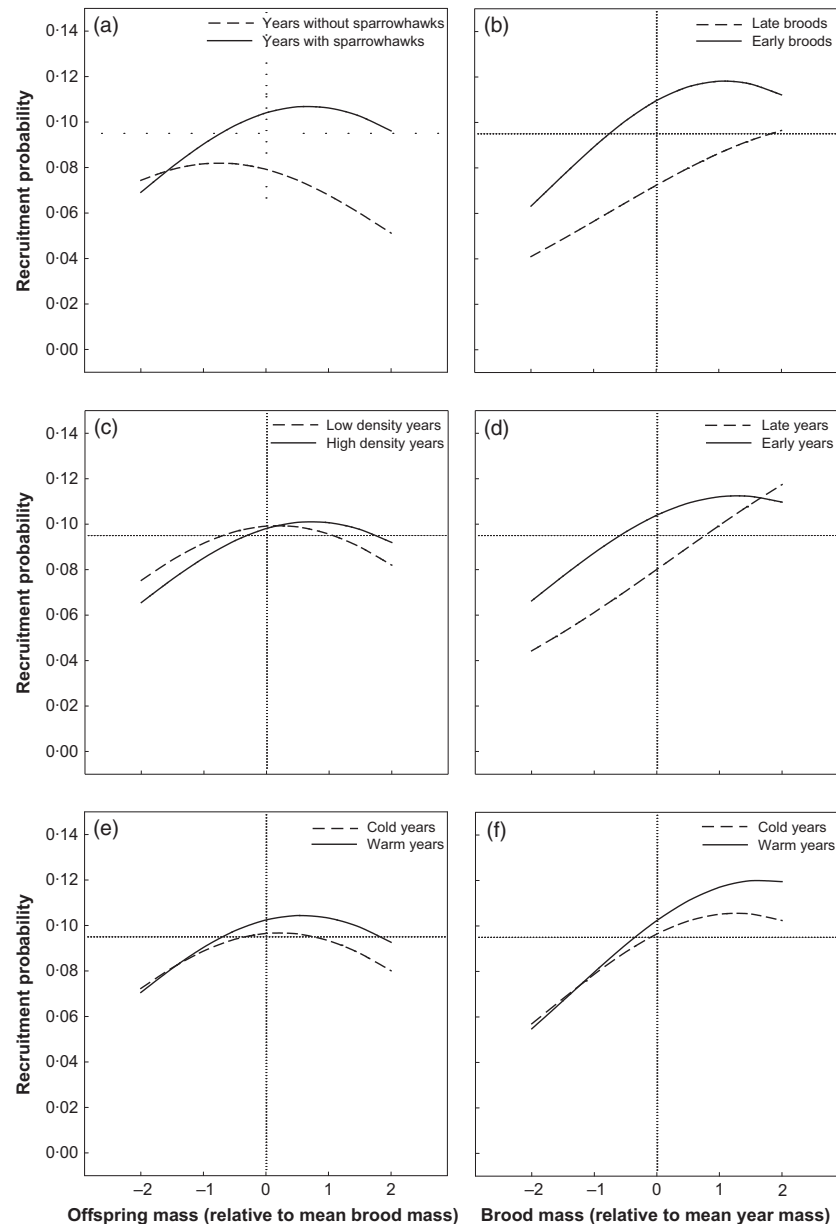


Fig. 2. Relationship between fledging mass (in g) and first-year survival probability at the level of offspring within broods (a, c, e) and broods within years (b, d, f), in relation to period (sparrowhawk) (a), breeding density (c), average winter temperature (e, f), and synchrony with the caterpillar food peak at the brood (b) and year (d) levels. Dotted lines represent average first-year survival probability over the study period (9.5%) and average offspring and brood fledging mass.

independent of offspring sex and fledging mass (Verhulst, Perrins & Riddington 1997; C.J. Garroway, C.A. Hinde & B.C. Sheldon, in prep) and is therefore unlikely to bias our results. Note that by quantifying fitness as local recruitment, we implicitly assume fledging mass to be an offspring trait. Fledging mass is, however, likely to reflect aspects of both the offspring and parental phenotype (e.g. 13 and 14% of variance in average nestling mass is explained by male and female identity, Browne *et al.* 2007), in which case the overall strength of selection on this trait will depend on both offspring and parental survival and the covariance between brood mean mass and future reproductive success. Similarly, we study fledging mass as an isolated trait, while we acknowledge that indirect genetic effects may also underlie evolutionarily relevant variation in fledging mass and survival and knowledge of such effects would be required to accurately predict genetic responses to selection on fledging mass.

For selection analyses, data analysed here come from 73 121 offspring hatched in 9822 nests in the 51 years between 1960 and

2010. The overall first-year survival probability (assessed to 2012) in this period was 0.095. As ecological variables potentially underlying correlations between fledging mass and first-year survival probability, at the year level, we tested: (i) period (sparrowhawk), a binary variable splitting the years 1960–1972 and 1973–2010, in which various ecological variables may also have differed, but which are particularly characterized by the absence and presence of a key predator of tits, the Eurasian sparrowhawk *Accipiter nisus* (Gosler, Greenwood & Perrins 1995; Vedder, Bouwhuis & Sheldon 2014), respectively; (ii) density, the overall population density measured as the number of nests per hectare in which at least one great tit chick reached the age of 15 days (Garant *et al.* 2004); (iii) beech mast, a three-category class variable describing the amount (none, little or much) of beech *Fagus sylvatica* mast available as food in winter (Perdeck, Visser & van Balen 2009; Grøtan *et al.* 2009); (iv) winter temperature, the average temperature in January and February after fledging (MacColl 2011); and (v) synchrony, the average synchrony of the

breeding population with the timing of the peak of its main food source, measured as the difference between the half-fall date of winter moth *Operophtera brumata* caterpillars and the average great tit clutch initiation date (Charmantier *et al.* 2008). The choice of these variables was governed by (i) evidence suggesting that they are related to variation in first-year survival in this, or other populations of great tits and (ii) availability of data for the majority of years over which this study was carried out. At the brood level, we tested (i) territory size, measured as the area of a Thiessen polygon (a geometric construct that places boundaries midway between adjacent neighbours), capped at 1 ha (Wilkin *et al.* 2006), (ii) oaks, the number of oaks with a larger than 70 cm diameter at breast height within a 75 m radius of the nest box (Wilkin, Perrins & Sheldon 2007) and (iii) nest synchrony, the difference between clutch initiation date and the population-level timing of the caterpillar half-fall date (Charmantier *et al.* 2008).

In 2011, 940 chicks that hatched in 115 broods were involved in a reciprocal cross-fostering experiment. When the oldest chick in a nest was 14 days of age, it and all its siblings were transferred to a nest matched for clutch initiation date (± 1 day) and brood size (± 1 chick), while the chicks in the foster nest were transferred to the initial nest. The next day, in their foster nests, all chicks were weighed and ringed according to standard protocol. The overall first-year survival probability to 2012 was 0.056. This experiment was performed to test a different hypothesis (relating to genetic and environmental effects on dispersal: C.J. Garroway, C.A. Hinde & B.C. Sheldon, in prep), and so the weighing of chicks was not performed on the day of cross-fostering, as would have been ideal for our purpose. However, as this is the age at which mass has usually approached an asymptote (van Balen 1973), we assume that foster parents could not alter chick weight substantially within a single day. This enabled us to test whether the brood-level correlation between fledging mass and first-year survival probability can be attributed to the environment experienced before or after 14 days of age.

STATISTICAL ANALYSES

Our long-term data were analysed in three steps. First, we constructed an individual-level generalized linear mixed model assuming a binomial error distribution in which we analysed effects of fledging mass on first-year survival (yes/no). Non-independence of offspring raised in the same brood, and of broods raised in the same year, was taken into account by fitting hierarchical random effects of year and brood. For fledging mass, we constructed three variables, which together sum to individual offspring mass: year_mass (annual average mass in g), brood_mass (as a deviation in g from year_mass) and ind_mass (as a deviation in g from the actual average brood mass). For example, for an individual fledgling of mass 19.5, hatched in a brood of average mass 18.6 in a year characterized by an average fledging mass of 18.5, the values for year_mass, brood_mass and ind_mass were 18.5, 0.1 and 0.9, respectively. Simultaneously fitting these three variables, as well as their squared variables (i.e. six mass variables in total), separates linear and quadratic fledging mass effects on different levels and allows direct comparison of parameter estimates (Tinbergen & Boerlijst 1990; van de Pol & Wright 2009).

Secondly, to this initial model, we added the annual and brood-level ecological variables described above as main effects, running a model for each variable separately. This allowed us to

quantify how each variable affected the parameter estimates of the six fledging mass variables. If the ecological variable were to affect both fledging mass and recruitment probability in the same direction, we would expect its addition to the model to reduce the parameter estimate of the effect of fledging mass on recruitment probability. If, on the other hand, the ecological variable was to affect fledging mass and recruitment probability in different directions, we would expect its addition to the model to increase the parameter estimate of the effect of fledging mass on recruitment probability. In both cases, the adjusted parameter estimate for the effect of fledging mass on recruitment probability is unbiased with respect to the environmental factor.

Thirdly, to the eight models constructed in the second step, we additionally added the interactions of the annual and brood-level ecological variables with all six fledging mass variables to test how the correlation between fledging mass and fitness itself depended on ecological variation. Each of the eight full models was simplified by backward stepwise removal of non-significant terms, where significance ($P < 0.05$, two-tailed) was assessed using the Wald statistic, which approximates the chi-square distribution.

Note that we chose to run our models for each variable separately. This was because models with six mass variables, eight environmental variables and 48 interaction terms get overly complex and data hungry, while causal effects will only be correctly reflected in estimates of selection if all sources of covariance are adequately measured and modelled, which cannot actually be achieved in studies of natural populations (e.g. Morrissey, Kruuk & Wilson 2010).

To analyse the experimental data, we fitted a brood-level generalized linear model with the proportion of recruited offspring as the dependent variable, the number of offspring in the brood as the denominator, and the average mass of both the natal brood of rearing (natal_mass, i.e. the actual mass of the focal brood) and foster brood of fledging (foster_mass, i.e. the mass of the fledglings taken out of the foster nest and moved to the natal box of the focal brood) as linear and quadratic explanatory variables. When only natal_mass turned out to predict first-year survival probability (see Results), we fitted an additional individual-level model with linear and quadratic effects of natal_brood_mass (in average mass in g) and ind_mass (as a deviation in g from natal_brood_mass) to test for between- and within-brood effects of fledging mass on first-year survival probability.

As most studies on selection report standardized selection differentials to allow comparison between populations or species, we wanted to quantify effects of multilevel covariance between fledging mass and recruitment probability on these measures. Standardized selection differentials for fledging mass were estimated using univariate linear regressions of relative fitness (i.e. first-year survival divided by the average first-year survival) on standardized values of fledging mass (Arnold & Wade 1984a,b). For the overall analysis, relative fitness and standardized fledging mass were calculated using pooled data on all individuals measured in all 51 years. For the brood-level analysis, relative fitness and standardized fledging mass were calculated using year-specific averages and standard deviations, while for the individual-level analysis, relative fitness was still obtained using the year-specific average, but fledging mass was standardized using brood-specific averages and the average brood-level standard deviation for each given year. All models were run in MLWIN 2.02 (Rasbash *et al.* 2005).

Results

MULTILEVEL COVARIANCE

A relationship between fledging mass and first-year survival probability was found at all three levels analysed: between years, between broods within years and between offspring within broods. The year-level (binomial) effect was 3.0 times stronger than the brood-level effect, which itself was 2.6 times stronger than the individual-level effect (Table 1).

Between years, only beech mast reduced the parameter estimate of the relationship between fledging mass and first-year survival probability (by 18%, Table 2), because mean fledging mass and recruitment probability were both elevated in years with a large winter beech mast. Beech mast is thus a considerable contributor to the overall relationship between fledging mass and survival. Period, that is, sparrowhawk presence, had the opposite effect of increasing the parameter estimate for the correlation between fledging mass and first-year survival probability by 16% (Table 2), because mean fledging mass was lower, but first-year survival probability higher, in the period in which sparrowhawks were present.

Population density, period (i.e. sparrowhawk presence), winter temperature and synchrony with the food peak all affected the strength of the size–fitness relationship at lower levels. In the 38-year period in which sparrowhawks were present, the individual-level relationship between fledging mass and first-year survival was positive, while it was negative in the 13 years when sparrowhawks were absent from the population (Table 3; Fig. 2a) (Table S1, Supporting information). In high density years and in years in which fledging was followed by a warm winter, the individual-level relationship between fledging mass and first-year survival probability was greater (i.e. selection was stronger; Table 3; Fig. 2c,e). In milder winters, the brood-level relationship between fledging mass and first-year survival probability was also stronger (Fig. 2f). In years in which the great tit population as a whole bred well in advance of the peak in caterpillar abundance, the brood-level effect of fledging mass on first-year survival was stabilizing at high mass, while it was directional in years in which the birds lagged behind their food peak (Fig. 2d).

Between broods within years, synchrony with the caterpillar half-fall date explained part of the relationship between fledging mass and first-year survival probability (Table 2, nest_mass² column), both when characterized at

Table 1. The effect of variation in great tit fledging mass on first-year survival probability at three levels: between years, between broods within years and between offspring within broods. Shown are parameter estimates obtained from a single model with standard errors, chi-square values and *P* values. Main parameter estimates and standard errors (SEs) were obtained from models assuming a binomial error distribution, but estimates in parentheses were obtained from the same model assuming a normal error distribution. Random effects are marked by (*r*)

Parameter	Est.	SE	χ^2_1	<i>P</i>
Year_mass	0.527 (0.04677)	0.162 (0.01411)	10.570	0.001
Year_mass ²	0.000 (0.00000)	0.000 (0.00000)	0.000	1.000
Brood_mass	0.176 (0.01229)	0.015 (0.00117)	145.644	<0.001
Brood_mass ²	−0.063 (−0.00137)	0.008 (0.00045)	57.421	<0.001
Ind_mass	0.069 (0.00455)	0.016 (0.00122)	18.765	<0.001
Ind_mass ²	−0.063 (−0.00220)	0.011 (0.00056)	30.627	<0.001
Year (<i>r</i>)	0.178 (0.00134)	0.038 (0.00028)	—	—
Brood (<i>r</i>)	0.294 (0.00265)	0.026 (0.00020)	—	—

Table 2. Percentage change in parameter estimates of the effect of fledging mass on first-year survival probability on three different levels when adding year- or brood-level environmental parameters as fixed effects (but not in interaction with fledging mass) to the model presented in Table 1. Because parameter estimates were derived in models assuming a binomial error distribution, for interpretability the percentage change in parameter estimates is additionally given as obtained from models assuming a normal error distribution in parentheses

Parameter	Year mass	Brood mass	Brood mass ²	Ind mass	Ind mass ²
Year level					
Period (sparrowhawk)	16 (16)	1 (0)	0 (0)	0 (0)	0 (0)
Beech mast	−18 (−20)	1 (0)	0 (0)	0 (0)	0 (0)
Winter temperature	2 (2)	0 (0)	0 (2)	0 (1)	0 (0)
Breeding density	−1 (−1)	0 (0)	0 (0)	0 (0)	0 (0)
Synchrony with food	−5 (−12)	4 (−1)	−14 (−26)	−3 (−9)	8 (−2)
Brood level					
Territory size	−5 (−4)	−2 (0)	−2 (1)	19 (24)	0 (0)
Number of large oaks	0 (0)	2 (1)	−2 (−2)	0 (−1)	0 (1)
Synchrony with food	1 (−8)	2 (−9)	−25 (−53)	−6 (−11)	6 (−7)

the population level (14%) and when characterized at the level of individual nests (25%). Synchrony at the brood level also affected the form of the relationship between fledging mass and first-year survival probability between broods, with brood-level effects of fledging mass on first-year survival being stabilizing in early, but directional in late, broods within years (Fig. 2b).

CROSS-FOSTERING EXPERIMENT

Brood-level analysis revealed that the proportion of fledglings recruited was predicted by the fledging mass obtained in the natal nest of rearing, rather than by the mass of the original chicks reared in the foster nest of fledging (foster_mass \pm SE: 0.042 ± 0.123 , $\chi^2_1 = 0.116$, $P = 0.733$; foster_mass² \pm SE: -0.008 ± 0.028 , $\chi^2_1 = 0.083$, $P = 0.773$). The natal nest mass effect was linear (natal_mass \pm SE: 0.348 ± 0.150 , $\chi^2_1 = 5.349$, $P = 0.021$), rather than quadratic (natal_mass² \pm SE: -0.143 ± 0.140 , $\chi^2_1 = 1.037$, $P = 0.309$). Hence, this experiment suggests that the brood-level relationship between fledging mass and recruitment is generated early in development and does not result from post-fledging effects.

Individual-level analysis also showed that there was evidence for a brood-level relationship between fledging mass and first-year survival probability (natal_brood_mass \pm SE: 0.358 ± 0.163 , $\chi^2_1 = 4.845$, $P = 0.028$), but did not reveal a statistically significant effect of fledging mass on survival within broods in this sample (ind_mass \pm SE: 0.216 ± 0.181 , $\chi^2_1 = 1.422$, $P = 0.223$; ind_mass² \pm SE: -0.046 ± 0.160 , $\chi^2_1 = 0.084$, $P = 0.772$). However, parameter estimates from this model did not differ significantly from those obtained from analyses on the 51-year data set.

SELECTION ANALYSES

Overall, the standardized directional selection differential (S') for fledging mass was estimated at 0.223 ± 0.011 across the 51 years of our study, while the nonlinear selection differential (c') was estimated to be stabilizing at -0.060 ± 0.006 . Using within-year relative fitness and standardized fledging mass reduced these differentials to 0.209 ± 0.013 (−6%) and -0.054 ± 0.007 (−10%), respectively. Using within-year relative fitness, but standardizing fledging mass within broods, the directional selection differential was estimated at 0.069 ± 0.012 (−67%), while the stabilizing selection differential was estimated at -0.070 ± 0.005 (+30%). Hence estimates of selection intensity are highly dependent on the extent to which the hierarchical structure of the data is considered.

Discussion

Despite a wealth of estimates of the direction, strength and form of phenotypic selection in natural populations, many questions remain regarding its consequences for evolutionary trajectories within those populations. Among

these questions are those regarding (i) the very limited evidence for responses to selection (Kingsolver *et al.* 2012), (ii) the nature of the agents of selection (MacColl 2011) and (iii) the causes of variation in the strength and form of selection (MacColl 2011; Kingsolver *et al.* 2012). Here, analysing a long-term data set spanning 51 years of breeding in a small passerine bird species, the great tit, we used a hierarchical partitioning approach to address these questions for a trait known to have a heritable basis and often found to be under strong selection: fledging mass.

We found a relationship between fledging mass and first-year survival probability at all three hierarchical levels we investigated: between years, between broods within years and between fledglings within broods. In the absence of different environmental sources of correlated effects or different selection pressures on each of these levels, we would expect the three slopes of the mass–fitness relationship to be identical (Tinbergen & Boerlijst 1990; also see Fig. 2 in Goodnight, Schwartz & Stevens 1992). However, we found the strength of these relationships (estimated from a model with binomial errors) to decrease by approximately a factor three with each level, which suggests that while the individual-level slope may be closest to representing the causal survival benefit that one extra gram of fledging mass *per se* conveys for an individual, extra causal mass effects on survival may exist on the brood level, and much of the mass–fitness relationship will to a considerable extent result from correlated effects of other factors at the brood- and year level (Tinbergen & Boerlijst 1990; Fig. 3). At the year level, we identified the size of the beech mast in the winter following fledging as a factor underlying correlated responses of fledging mass and survival, accounting for 18% of the mass effect on survival. In years in which fledging was followed by a large beech mast, both average fledging mass and recruitment probability were high (see Fig. S1a, Supporting information). For first-year survival probability, this is likely to be a direct effect of an extra, dependable food source in winter. For fledging mass, a direct effect of food is unlikely, due to the obvious time lag (beech mast is produced in the autumn, long after birds have fledged), but also because the biomass of the main caterpillar food source in spring is not related to the size of the beech mast crop in the following winter (Perdeck, Visser & van Balen 2009). An interesting possibility is that great tits, such as American and Eurasian red squirrels (*Tamiasciurus hudsonicus* and *Sciurus vulgaris*; Boutin *et al.* 2006), show anticipatory reproductive investment in years in which the beech mast will be large (which they can potentially predict from the flowering of beeches during the parental provisioning period), for example because costs of reproduction are easier to bear in such years or because the expected return on investment is higher.

At the brood level, the relationship between fledging mass and first-year survival probability was partly explained as a correlated effect of seasonal differences in reproductive performance. In this case, however, it was

Table 3. Effects of environmental parameters (left part of table: year level; right part of table: nest level) on first-year survival probability, either as main effects or in interaction with fledging mass. Shown are parameter estimates with standard errors (SEs) for statistically significant terms only, as estimated in the single minimal adequate model for each environmental parameter. Random effects are marked by (*r*)

Environmental parameter	Period (sparrowhawk)			Density			Beech mast			Winter temperature			Synchrony			Territory size			Oaks			Nest synchrony		
	Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE		Est.	SE	
Year _{mass}	0.609	0.163		0.520	0.183		0.433	0.133		0.537	0.161		0.523	0.187		0.499	0.165		0.529	0.162		0.532	0.188	
Year _{mass} ²	0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000	
Nest _{mass}	0.196	0.014		0.176	0.015		0.177	0.015		0.084	0.043		0.786	0.130		0.172	0.015		0.179	0.015		0.391	0.077	
Nest _{mass} ²	-0.063	0.008		-0.063	0.008		-0.063	0.008		-0.063	0.008		0.090	0.070		-0.062	0.009		-0.062	0.008		0.045	0.030	
Ind _{mass}	-0.100	0.042		-0.046	0.050		0.069	0.016		-0.037	0.048		0.067	0.019		0.082	0.017		0.069	0.016		0.065	0.019	
Ind _{mass} ²	-0.067	0.011		-0.064	0.011		-0.063	0.012		-0.065	0.011		-0.069	0.013		-0.063	0.012		-0.063	0.011		-0.067	0.013	
Env	0.300	0.144		-0.034	0.331		0.159	0.127		0.027	0.036		0.041	0.018		0.332	0.064		-0.001	0.001		0.038	0.003	
Env × year _{mass}	-	-		-	-		-	-		-	-		-	-		-	-		-	-		-	-	
Env × year _{mass} ²	-	-		-	-		-	-		-	-		-	-		-	-		-	-		-	-	
Env × nest _{mass}	-	-		-	-		-	-		0.022	0.009		-0.018	0.004		-	-		-	-		-0.006	0.002	
Env × nest _{mass} ²	-	-		-	-		-	-		-	-		-0.004	0.002		-	-		-	-		-0.003	0.001	
Env × ind _{mass}	0.190	0.045		0.197	0.082		-	-		0.019	0.008		-	-		-	-		-	-		-	-	
Env × ind _{mass} ²	-	-		-	-		-	-		-	-		-	-		-	-		-	-		-	-	
Year (<i>r</i>)	0.165	0.035		0.177	0.038		0.107	0.024		0.176	0.037		0.186	0.047		0.168	0.038		0.177	0.038		0.188	0.047	
Nest (<i>r</i>)	0.253	0.025		0.296	0.026		0.308	0.026		0.285	0.026		0.315	0.032		0.273	0.026		0.297	0.026		0.275	0.031	

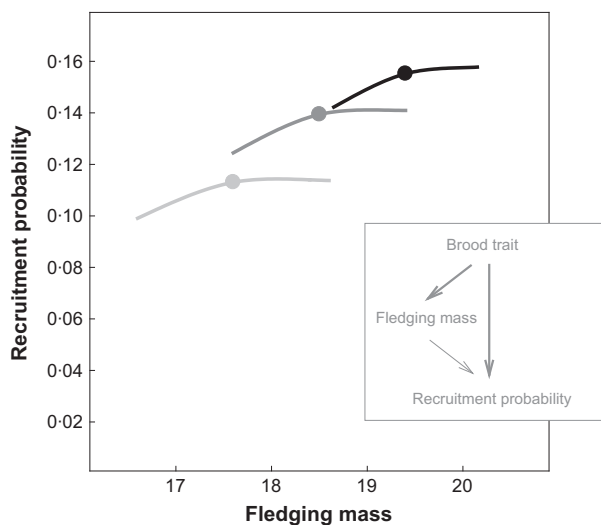


Fig. 3. The predicted relationship between great tit fledging mass and first-year survival probability at the brood and individual levels in a year with a mean fledging mass of 18.5 g. The figure illustrates model predictions for three hypothetical broods with an average fledging mass of 17.6, 18.5 and 19.4 g (i.e. average ± 1 SD of brood mass), while offspring within these broods have a range of 2.0, 1.8 and 1.6 g, respectively (i.e. average ± 1 SD offspring mass per average brood mass). The inset shows that because the brood-level relationship is 2.6 times stronger than the individual-level relationship, this relationship mostly results from a correlated response to a brood-level trait.

the nonlinear aspect of the relationship that was affected, because years in which the great tit population as a whole initiated nests early in relation to the caterpillar half-fall date, as well as individual broods that were initiated early (see Fig. S1c, Supporting information), were characterized by intermediate fledging mass, but high first-year survival probability, in comparison with late, or intermediately timed, years and broods. When taking into account the population- and nest-level synchrony of clutch initiation with the caterpillar half-fall date, estimates of the effect of mass on survival at the brood level were 14% and 25% smaller, respectively. Since the earliest breeders are generally the individuals of highest quality (Verhulst & Nilsson 2008), parental quality that is transferred to offspring may explain the brood-level effect of synchrony on recruitment probability. Alternatively, early broods may provide fledglings with good food conditions just after fledging, or with more time to develop flight skills before predators exert their strongest selection pressure, as well as providing high nutrition during earlier phases of growth. The year-level effect is less strong and explanations of this effect are less obvious.

Our analyses of the long-term data set did not help us to identify factors underlying the linear aspect of the brood-level relationship between fledging mass and first-year survival probability in addition to that of the individual-level relationship. However, our cross-fostering experiment revealed that this relationship can be attributed to causal mass effects and the environment

experienced before fledging, since the brood-level mean fledging mass obtained before being cross-fostered at the age of 14 days was related to first-year survival probability, while the average fledging mass of the chicks originally raised by the foster parents was not. Post-fledging, brood-level parental quality effects, such as parents raising a heavy brood leading fledged offspring to high-quality or low-risk feeding areas, thereby improving their survival prospects, can therefore be discounted, while pre-fledging brood-level effects (other than territory size or number of oaks within a 75 m radius of the nest box) should be considered in future work.

Partitioning the relationship between fledging mass and first-year survival probability to different organizational levels allows quantification of its effect on estimates of selection differentials. To date, selection analyses in wild populations have usually been performed in two ways. First, overall estimates of selection differentials have been obtained using overall relative fitness and phenotypic trait values standardized with the overall mean and variance of the trait in all individuals. Secondly, year-specific estimates of selection differentials have been obtained using year-standardized phenotypic trait values and year-specific relative fitness. In the latter case, statistical significance of the resulting estimates of selection differentials is sometimes tested in additional models where a random effect is included to account for the statistical non-independence of offspring reared in the same breeding attempt (e.g. Kruuk, Merilä & Sheldon 2001; Charmantier *et al.* 2004; Garant *et al.* 2004). Our finding that the relationship between fledging mass and first-year survival probability varies by a (binomial) factor three between levels, however, implies that estimates of selection differentials on fledging mass may, in absence of evidence for between-group causal effects of the trait on fitness, be most appropriately made at the within-group level. In our case, the difference is substantial. While our estimate of the overall directional selection differential was only 6% higher than our estimate controlling for the between-year relationship between fledging mass and recruitment probability, this estimate itself was 67% higher than our estimate controlling for the brood-level relationship as well. Note that we do not argue that, in our case, competition for survival in relation to fledging mass occurs only between fledglings raised in the same brood. This may be the case, since analyses of data from our population suggest that selection on fledging mass predominantly takes place in the first few weeks after fledging (S. Bouwhuis & B. C. Sheldon, unpublished results), when fledglings rely on post-fledging parental care and roam the woods in family flocks (Verhulst & Hut 1996) and when the main source of mortality is predation (Naef-Daenzer, Widmer & Nuber 2001). However, we argue that in the absence of evidence for causal brood mass effects on survival, the within-level selection differential will most accurately reflect the extent to which fledging mass variation *per se* is the cause of fitness variation between individuals and

be least sensitive to environmental contributions to the phenotype-fitness covariance overwhelming the evolutionarily relevant covariance (also see Morrissey, Kruuk & Wilson 2010). Either way, this result shows that failing to take into account the phenotypic structuring within a population can result in greatly inflated estimates of selection differentials, as may also be the case for estimates of heritability (Price & Schluter 1991). We therefore suggest that explanations for the mismatch between observed and expected evolutionary dynamics in natural populations should also consider the possibility that selection estimates are inflated, in addition to other factors which have often been concerned with the quantitative genetic parameter estimation (Brooks & Endler 2001; Merilä, Sheldon & Kruuk 2001; Kruuk *et al.* 2002).

In our multilevel approach of adding environmental factors as independent variables to trait-fitness regressions, a significant interaction between the environmental factor and trait would provide the best non-experimental support for the environmental factor's role as an agent of selection. In our case, we found individual-level selection to be stronger in a period in which sparrowhawks were present in the population, in years when the population-wide breeding density was high, and in years in which the breeding season was followed by a mild winter. In each case, we found the lightest fledglings to suffer most from reduced recruitment probability. The first result suggests that sparrowhawk predation is a likely agent of selection on great tit fledging mass, which is confirmed by our recent finding that lighter great tit fledglings are more likely to be found as prey remains in sparrowhawk pellets in the vicinity of sparrowhawk nests (Vedder, Bouwhuis & Sheldon 2014). This finding is not in agreement with the idea that heavy individuals may lack the agility and speed to escape predation (Witter & Cuthill 1993), but rather suggests that light fledglings are least well developed for flight ability and escape, or beg loudest and are therefore easiest for predators to locate. Our second result that breeding density and winter temperature affect the strength of individual-level selection suggests that the density of conspecifics, which compete for clumped and limited food sources over winter, may be an additional agent of selection, if a higher breeding density and a milder winter both result in an increase in the number of competitors via increased production and reduced mortality, respectively. Indeed, great tit fledging mass (which is related to adult mass, Perrins & McCleery 2001) may confer an advantage in competition for food (e.g. Lange & Leimar 2004).

Overall, our analyses show that correlated responses of traits and fitness to environmental factors at different organizational levels can make up a large component of estimates of selection, such that analyses of microevolution are best performed using estimates of selection at the level at which causality is most likely to occur. Our analyses also identify environmental factors at a range of levels as underlying substantial parts of the non-causal

covariance between size and fitness in a wild bird population. In cases where experimental manipulation of traits is difficult, partitioning correlations between levels of selection and environmental factors is a very promising approach to understanding the causes of selection.

Acknowledgements

We are grateful to the many people who collected data over the course of this study, as well as the many funding agencies, too numerous to list here, that have funded the long-term research. We would like to thank Fanie Pelletier, Julien Martin and Thomas Reed for their helpful comments and constructive criticism that improved the paper. Recent data collection was funded by an ERC Advanced Investigator grant (AdG 250164) to BCS, and we gratefully acknowledge the role of the EGI Social Networks group in collecting these data. SB and OV were funded by NWO Rubicon grants.

Data accessibility

Data are archived at the Dryad Digital Repository <http://dx.doi.org/10.5061/dryad.dd815> (Bouwhuis *et al.* 2014).

References

- Adriaenssen, F., Dhondt, A.A., van Dongen, S., Lens, L. & Matthysen, E. (1998) Stabilizing selection on blue tit fledgling mass in the presence of sparrowhawks. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 1011–1016.
- Arnold, S.J. & Wade, M.J. (1984a) On the measurement of natural and sexual selection – theory. *Evolution*, **38**, 709–719.
- Arnold, S.J. & Wade, M.J. (1984b) On the measurement of natural and sexual selection – applications. *Evolution*, **38**, 720–734.
- van Balen, J.H. (1973) A comparative study of the breeding ecology of the great tit, *Parus major*, in different habitats. *Ardea*, **61**, 1–93.
- Both, C., Visser, M.E. & Verboven, N. (1999) Density-dependent recruitment rates in great tits: the importance of being heavier. *Proceedings of the Royal Society B-Biological Sciences*, **266**, 465–469.
- Boutin, S., Wauters, L.A., McAdam, A.G., Humphries, M.M., Tosi, G. & Dhondt, A.A. (2006) Anticipatory reproduction and population growth in seed predators. *Science*, **314**, 1928–1930.
- Bouwhuis, S., Vedder, O., Garroway, C.J. & Sheldon, B.C. (2014) Ecological causes of multi-level covariance between size and first-year survival in a wild bird population. *Dryad Digital Repository*, <http://dx.doi.org/10.5061/dryad.dd815>
- Brooks, R. & Endler, J.A. (2001) Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). *Evolution*, **55**, 1002–1015.
- Browne, W.J., McCleery, R.H., Sheldon, B.C. & Pettifor, R.A. (2007) Using cross-classified multivariate mixed response models with application to life history traits in great tits (*Parus major*). *Statistical Modelling*, **7**, 217–238.
- Calsbeek, R. & Smith, T.B. (2007) Probing the adaptive landscape using experimental islands: density-dependent natural selection on lizard body size. *Evolution*, **61**, 1052–1061.
- Charmanier, A., Kruuk, L.E.B., Blondel, J. & Lambrechts, M.M. (2004) Testing for microevolution in body size in three blue tit populations. *Journal of Evolutionary Biology*, **17**, 732–743.
- Charmanier, A., McCleery, R.H., Cole, L.R., Perrins, C.M., Kruuk, L.E.B. & Sheldon, B.C. (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, **320**, 800–803.
- Chevin, L.M., Lande, R. & Mace, G.M. (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *Plos Biology*, **8**, e1000357.
- Eldakar, O.T., Wilson, D.S., Dlugos, M.J. & Pepper, J.W. (2010) The role of multilevel selection in the evolution of sexual conflict in the water strider *Aquarius remigis*. *Evolution*, **64**, 3183–3189.
- Endler, J.A. (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton.

- Garant, D., Kruuk, L.E.B., McCleery, R.H. & Sheldon, B.C. (2004) Evolution in a changing environment: a case study with great tit fledging mass. *American Naturalist*, **164**, E115–E129.
- Garnett, M.C. (1981) Body size, its heritability and influence on juvenile survival among great tits, *Parus major*. *Ibis*, **123**, 31–41.
- Goodnight, C.J. (2012) On multilevel selection and kin selection: contextual analysis meets direct fitness. *Evolution*, **67**, 1539–1548.
- Goodnight, C.J., Schwartz, J.M. & Stevens, L. (1992) Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *American Naturalist*, **140**, 743–761.
- Gosler, A.G., Greenwood, J.J.D. & Perrins, C.M. (1995) Predation risk and the cost of being fat. *Nature*, **377**, 621–623.
- Grotan, V., Saether, B.E., Engen, S., van Balen, J.H., Perdeck, A.C. & Visser, M.E. (2009) Spatial and temporal variation in the relative contribution of density dependence, climate variation and migration to fluctuations in the size of great tit populations. *Journal of Animal Ecology*, **78**, 447–459.
- Heisler, I.L. & Damuth, J. (1987) A method for analyzing selection in hierarchically structured populations. *American Naturalist*, **130**, 582–602.
- Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E. et al. (2001) The strength of phenotypic selection in natural populations. *American Naturalist*, **157**, 245–261.
- Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.
- Kruuk, L.E.B., Merilä, J. & Sheldon, B.C. (2001) Phenotypic selection on a heritable size trait revisited. *American Naturalist*, **158**, 557–571.
- Kruuk, L.E.B., Slate, J., Pemberton, J.M., Brotherstone, S., Guinness, F. & Clutton-Brock, T.H. (2002) Antler size in red deer: heritability and selection but no evolution. *Evolution*, **56**, 1683–1695.
- Lange, H. & Leimar, O. (2004) Social stability and daily body mass gain in great tits. *Behavioral Ecology*, **15**, 549–554.
- Lima, S.L. (1998) Stress and decision making under the risk of predation: recent developments from behavioral, reproductive, and ecological perspectives. *Advances in the Study of Behavior*, **27**, 215–290.
- Lindén, M., Gustafsson, L. & Pärt, T. (1992) Selection on fledging mass in the collared flycatcher and the great tit. *Ecology*, **73**, 336–343.
- MacColl, A.D.C. (2011) The ecological causes of evolution. *Trends in Ecology and Evolution*, **26**, 514–522.
- Merilä, J., Sheldon, B.C. & Kruuk, L.E.B. (2001) Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112**, 199–222.
- Monrós, J.S., Belda, E.J. & Barba, E. (2002) Post-fledging survival of individual great tits: the effect of hatching date and fledging mass. *Oikos*, **99**, 481–488.
- Morrissey, M.B., Kruuk, L.E.B. & Wilson, A.J. (2010) The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, **23**, 2277–2288.
- Naef-Daenzer, B., Widmer, F. & Nuber, M. (2001) Differential post-fledging survival of great and coal tits in relation to their condition and fledging date. *Journal of Animal Ecology*, **70**, 730–738.
- Perdeck, A.C., Visser, M.E. & van Balen, J.H. (2009) Great tit *Parus major* survival, and the beech-crop cycle. *Ardea*, **88**, 99–108.
- Perrins, C.M. (1965) Population fluctuations and clutch-size in the great tit, *Parus major* L. *Journal of Animal Ecology*, **34**, 601–647.
- Perrins, C.M. (1979) *British Tits*. Collins, London.
- Perrins, C.M. & McCleery, R.H. (2001) The effect of fledging mass on the lives of great tits *Parus major*. *Ardea*, **89**, 135–142.
- van de Pol, M. & Wright, J. (2009) A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour*, **77**, 753–758.
- Price, T. & Schluter, D. (1991) On the low heritability of life-history traits. *Evolution*, **45**, 853–861.
- Rasbash, J., Steele, F., Browne, W. & Prosser, B. (2005) *A User's Guide to MlwiN Version 2.0*. Institute of Education, London.
- Royle, N.J., Pike, T.W., Heeb, P., Richner, H. & Koelliker, M. (2012) Offspring social network structure predicts fitness in families. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 4914–4922.
- Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, **12**, 1261–1276.
- Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., DiBattista, J.D. & Carlson, S.M. (2013) The spatial patterns of directional phenotypic selection. *Ecology Letters*, **16**, 1382–1392.
- Tinbergen, J.M. & Boerlijst, M.C. (1990) Nestling weight and survival in individual great tits (*Parus major*). *Journal of Animal Ecology*, **59**, 1113–1127.
- Vedder, O., Bouwhuis, S. & Sheldon, B.C. (2014) The contribution of an avian top predator to selection in prey species. *Journal of Animal Ecology*, **83**, 99–106.
- Verhulst, S. & Hut, R.A. (1996) Post-fledging care, multiple breeding and the costs of reproduction in the great tit. *Animal Behaviour*, **51**, 957–966.
- Verhulst, S. & Nilsson, J.-A. (2008) The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **363**, 399–410.
- Verhulst, S., Perrins, C.M. & Riddington, R. (1997) Natal dispersal of great tits in a patchy environment. *Ecology*, **78**, 864–872.
- Visser, M.E. & Verboven, N. (1999) Long-term fitness effects of fledging date in great tits. *Oikos*, **85**, 445–450.
- Wade, M.J. & Kalisz, S. (1990) The causes of natural selection. *Evolution*, **44**, 1947–1955.
- Wilkin, T.A., Perrins, C.M. & Sheldon, B.C. (2007) The use of GIS in estimating spatial variation in habitat quality: a case study of laydate in the great tit *Parus major*. *Ibis*, **149**, 110–118.
- Wilkin, T.A., Garant, D., Gosler, A.G. & Sheldon, B.C. (2006) Density effects on life-history traits in a wild population of the great tit *Parus major*: analyses of long-term data with GIS techniques. *Journal of Animal Ecology*, **75**, 604–615.
- Witter, M.S. & Cuthill, I.C. (1993) The ecological costs of avian fat storage. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **340**, 73–92.

Received 10 October 2013; accepted 19 June 2014
Handling Editor: Fanie Pelletier

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. Notes on the relationships between fledging mass and first-year survival in relation to beech mast, sparrowhawk presence and synchrony with a food source.

Fig. S1. The year- and brood-level relationships between fledging mass and survival to recruitment are partly due to correlated effects on both of (a) beech mast, (b) period (sparrowhawk presence), and (c) synchrony with the caterpillar half-fall date.

Table S1. The effect of variation in great tit fledging mass on first-year survival probability at three levels: between years, between broods within years, and between offspring within broods in 13 years in which sparrowhawks were absent from the population.