

Multiple aspects of plasticity in clutch size vary among populations of a globally distributed songbird

David F. Westneat^{1*}, Veronika Bókony², Terry Burke^{3,4}, Olivier Chastel⁵, Henrik Jensen⁶, Thomas Kvalnes⁶, Ádám Z. Lendvai^{5,7}, András Liker^{2,4}, Douglas Mock⁸, Julia Schroeder^{4,†}, P. L. Schwagmeyer⁸, Gabriele Sorci⁹ and Ian R. K. Stewart^{1,‡}

¹Department of Biology and Center for Ecology, Evolution, and Behavior, University of Kentucky, 101 Morgan Building, Lexington, KY 40506-0225, USA; ²Department of Limnology, University of Pannonia, Pf. 158, Veszprém, H-8201, Hungary; ³Department of Genetics, University of Nottingham, NG7 2RD, UK; ⁴Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK; ⁵Centre d'Etudes Biologiques de Chizé, Centre Nationale de la Recherche Scientifique, Villiers-en-Bois, Beauvoir sur Niort F-79360, France; ⁶Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim NO-7491, Norway; ⁷Department of Biology, Virginia Tech, 4102 Derring Hall, Blacksburg, VA 24060, USA; ⁸Department of Biology, 730 Van Vleet Oval, Norman, OK 73019, USA; ⁹Biogéosciences, CNRS UMR 6282, Université de Bourgogne, 6 Boulevard Gabriel, Dijon F-21000, France

Summary

1. Plasticity in life-history characteristics can influence many ecological and evolutionary phenomena, including how invading organisms cope with novel conditions in new locations or how environmental change affects organisms in native locations. Variation in reaction norm attributes is a critical element to understanding plasticity in life history, yet we know relatively little about the ways in which reaction norms vary within and among populations.

2. We amassed data on clutch size from marked females in eight populations of house sparrows (*Passer domesticus*) from North America and Europe. We exploited repeated measures of clutch size to assess both the extent of within-individual phenotypic plasticity and among-individual variation and to test alternative hypotheses about the underlying causes of reaction norm shape, particularly the decline in clutch size with date.

3. Across all populations, females of this multibrooded species altered their clutch size with respect to date, attempt order, and the interaction of date and order, producing a reaction norm in multidimensional environmental space. The reaction norm fits that predicted by a model in which optimal clutch size is driven by a decline with date hatched in the ability of offspring to recruit. Our results do not fit those predicted for other proposed causes of a seasonal decline in clutch size.

4. We also found significant differences between populations in response to date and the date by attempt order interaction. We tested the prediction that the relationship with date should be increasingly negative as breeding season becomes shorter but found steeper declines in clutch size with date in populations with longer seasons, contrary to the prediction. Populations also differed in the level of among-individual variation in reaction norm intercept, but we found no evidence of among-individual variation in reaction norm slope.

5. We show that complex reaction norms in life-history characters exhibit within- and among-population variance. The nature of this variance is only partially consistent with current life-history theory and stimulates expansions of such theory to accommodate complexities in adaptive life history.

*Correspondence author. E-mail: biodfw@uky.edu

†Current address: Department of Biology, University of Delaware, Newark, DE 19716, USA

‡Current address: Evolutionary Biology, Max Planck Institute for Ornithology, Eberhard-Gwinner-Strasse, Seewiesen 82319, Germany

Key-words: adaptation, clutch size, life history, phenology of breeding, phenotypic plasticity, trade-offs, within-individual variance

Introduction

Many organisms exhibit labile traits – behavioural, physiological or morphological characters that differ in expression during an individual's lifetime. Labile traits are examples of phenotypic plasticity, broadly defined as the effect of an environmental factor on the phenotypic expression of a genotype (Stearns 1989; Scheiner 1993). Because individual organisms can express labile traits multiple times, the trait values can be described by a function relating phenotype to environment (a reaction norm; Woltereck 1909; Bradshaw 1965; Gomulkiewicz & Kirkpatrick 1992; Nussey, Wilson & Brommer 2007). For linear reaction norms, both the elevation (or intercept; the individual's mean phenotype in the average environment) and the slope (the change in the individual's phenotype in different environments) could vary among individuals. Because the parameters of a reaction norm have often been hypothesized to have evolved through natural selection (e.g. Gotthard & Nylin 1995; Pigliucci 2001; Ghalambor, Angeloni & Carroll 2010), the nature of variation in reaction norm parameters is essential for both selection and, to the extent that this variation has a genetic basis, the evolutionary response to selection (e.g. Lande 2009). Despite a growing number of studies measuring reaction norm parameters in natural populations (e.g. Dingemanse & Wolf 2013), many important aspects of reaction norms and the way they vary within and among populations remain poorly understood (e.g. Husby *et al.* 2010).

Consider the number of eggs laid by birds per breeding attempt, or clutch size, which is an important component of life history that varies substantially both between and within species (e.g. Lack 1947; Martin 1987; Crick, Gibbons & Magrath 1993; Bennett & Owens 2002). There is growing evidence that this trait is phenotypically plastic – individuals within a population produce different clutch sizes in different breeding attempts (van Noordwijk 1989; Postma & van Noordwijk 2005; Westneat, Stewart & Hatch 2009; Husby *et al.* 2010) or in response to manipulated environmental conditions (e.g. Boutin 1990; Nager, Rueger & van Noordwijk 1997; Clifford & Anderson 2001). This plasticity appears adaptive because females usually alter the size of their clutch in the direction expected to maximize lifetime reproductive success (Pettifor, Perrins & McCleery 2001).

One possible example of adaptive plasticity is that clutch size in many bird species declines with clutch initiation date (Klomp 1970; Drent & Daan 1980; Murphy 1986; Brommer, Pietiäinen & Kokko 2002). At least 10 hypotheses exist to explain such declines (Decker, Conway & Fontaine 2012), but tests of predictions are made difficult by confusion over whether the decline is produced by differences between individuals or plasticity within

individuals. For example, the decline could be driven by individuals of high-quality breeding early and producing more eggs, or it could reflect phenotypic plasticity if individuals adjust clutch size in response to time of season (Drent & Daan 1980; Verhulst & Nilsson 2008). Plasticity could be adaptive because season affects either the benefits of producing a given clutch size or the costs of that clutch size to the parents. Many individual factors could be sensitive to season and affect one or the other of these two fitness components (Decker, Conway & Fontaine 2012), including seasonal declines in food supply for eggs or nestlings (Lack 1947) or in juvenile survival (Drent & Daan 1980; Rowe, Ludwig & Schluter 1994), or seasonal increases in the costs of care to the parents' survival or condition (Siikamäki, Hovi & Rätti 1994, Decker, Conway & Fontaine 2012).

The hypotheses that declines in clutch size arise from differences in quality or from plasticity can be distinguished in species that produce multiple breeding attempts within a season. The replicate measures of clutch size allow estimates of both within- and among-individual variation and their covariates. In addition, comparing reaction norm variation within and among populations of such species offers several opportunities for better understanding the nature of phenotypic plasticity in general, and the underlying causes of clutch size variation in particular.

In general, one might predict that differences between populations in reaction norm attributes (either intercept or slope) would arise if there were both differences between populations in ecological conditions and variation within populations in individual responses to those conditions (e.g. Scheiner 1993; Nussey, Wilson & Brommer 2007). Specifically, the hypothesized causes of seasonal declines in clutch size make different predictions about reaction norm shape in species that produce multiple clutches in a season, and comparisons among populations might further distinguish among these hypotheses. For example, if the seasonal decline arises because parents are matching clutch size to food supply, then the reaction norm should decline with date, possibly as a quadratic (food supply likely first increases and then decreases with date), and populations with different season lengths should show different quadratic relationships. If date affects the costs to parents of reproducing, then clutch size should decline with date within individuals. An additional, independent decline with number of previous attempts might be expected under the assumption that prior breeding also increases the costs of reproduction for the current attempt (e.g. Williams 2005). If parental costs are associated with the end of the breeding season, then populations with shorter seasons should also show a stronger decline with date. A final hypothesis is that the

decline with date results solely from a decline in offspring fitness with the date they are hatched. Rowe, Ludwig and Schluter (1994) made several predictions about how this time horizon hypothesis would influence plasticity in clutch size with respect to date. One prediction was that where juvenile survival is more negatively affected by date, then date should have a stronger effect on clutch size. However, Brommer, Pietiäinen and Kokko (2002) showed that in Ural owls (*Strix uralensis*), the clutch size–date relationship was inversely related to the date–juvenile recruitment rate relationship, opposite to the prediction.

Rowe *et al.* also predicted that independently of the decline with date, clutch size in species that breed multiple times a season was expected to increase with attempt order within the season and decline with a date by attempt order interaction (their Fig. A1). As far as we are aware, no other hypothesized factor except the time horizon hypothesis should predict this type of reaction norm. Westneat, Stewart and Hatch (2009) tested these predictions using a long-term data set from a single population of house sparrows (*Passer domesticus*) and found that clutch size indeed declined with date, increased with attempt order within a season after controlling for date and declined more strongly with date as attempt order increased. These results supported the time horizon hypothesis, but we do not know if this pattern is present in all populations of house sparrows nor how differences between populations might influence it.

Here, we present an among-population analysis of within- and among-individual variance in clutch size reaction norms, using data collected from eight populations of house sparrows. The house sparrow has a nearly global distribution, with long-established populations in western Asia and throughout Europe and variably established introduced populations in North America, South America, Africa and Australia. The species also breeds from the equator to just south of the Arctic circle and just north of the Antarctic Circle in both mainland and island locales (Anderson 2006). This wide distribution and the habit of producing multiple clutches per season offer an unusual opportunity to compare multiple populations to assess several predictions about the evolution of reaction norms in general and the forces affecting plasticity of clutch size in particular. First, to be under selection, reaction norms must exhibit within-population variation in slope (e.g. Postma & van Noordwijk 2005; Nussey, Wilson & Brommer 2007), and the extent of plasticity should affect fitness. Secondly, if selection on plasticity is driven by general features of the life history of the organism, then given the results of Westneat, Stewart and Hatch (2009), house sparrows would seem to be influenced by the time horizon hypothesis. If the time horizon is a general feature of their life history, then the multidimensional reaction norm exhibited in the Kentucky population should be present in all populations. Thirdly, if plasticity is driven by ecological conditions, populations differing in ecology should exhibit predictably different reaction norms, possibly in both elevation and slope.

In particular, because the work of Rowe, Ludwig and Schluter (1994) was focused on the idea that if juvenile fitness and thus the reproductive value of an egg declines as lay date approaches the end of the season, clutch sizes in populations with shorter seasons should display a steeper decline with date and a stronger interaction with attempt order. The comparative analysis also allows tests of the alternative hypotheses for the decline in clutch size with date. Finally, the distribution of house sparrows world-wide is a mix of populations that are native or introduced as well as insular or continental. We also asked whether the magnitude of among-individual variation differed among populations, which might suggest differences in genetic structure or the presence of additional environmental variables affecting the plasticity of clutch size.

Methods

DATA SET

We analysed data on clutch size from individually marked female house sparrows from eight multiyear studies distributed in North America and Europe (Table 1): Chizé (France), Helgeland (Norway), Hoedic (France), Kentucky (USA), Lundy Island (UK), Nottingham (UK), Oklahoma (USA) and Veszprém (Hungary). At all sites, females were individually marked with either a numbered metal band, a unique combination of colour bands on their legs, or both. Females were captured using seed-baited traps (Kentucky: Westneat, Stewart & Hatch 2009), mist-nets (Helgeland: Jensen *et al.* 2008; Veszprém: Bókonyi *et al.* 2008) or some combination of methods (Chizé: Chastel & Kersten 2002; Lendvai & Chastel 2010; Hoedic: C. Bichet, D.J. Penn, Y. Moodley, L. Dunoyer, E. Cellier-Holzem, M. Belvalette, A. Grégoire, S. Garnier & G. Sorci, unpublished; Lundy Island: Cleasby *et al.* 2010; Nottingham: Burke 1984; Oklahoma: Schwagmeyer, Mock & Parker 2002). In seven of the eight populations, females bred in artificial nest boxes, but in the Helgeland population, females nested in cavities and crevices in farm buildings and other man-made structures. In all populations, nests were checked at least once per week beginning in early spring continuing all summer. Clutch completion and final clutch size were indicated by the same number of eggs on two successive checks or information (such as the timing of hatching) that indicated a check occurred in the middle of incubation (sparrows incubate for 10–11 days following clutch completion, Anderson 2006). The Hoedic population was followed through the first two attempts but not through the end of breeding, so it was removed from some analyses. For seven of the eight populations, females were assigned to nest attempts based on them being observed repeatedly entering or standing on the nest box during incubation or nestling provisioning, or being captured in the box. In some cases in each population, the female was not observed at a particular clutch, but was assigned that clutch based on the fact that she nested there during the previous or subsequent attempt (or both), and there were no unusual disruptions or long intervals between attempts to suggest there had been a change in ownership. In the Helgeland population, females were assigned to nests by use of microsatellite analysis of DNA collected from females and nestlings (Jensen *et al.* 2008). Clutches at Helgeland that failed to hatch therefore could not be assigned except if they were at the same site and timed in between two other attempts by the same female.

Table 1. Summary statistics on location, study duration, timing of breeding and key variables relevant to analysis of clutch size for eight populations of house sparrows. Means \pm 1 SD and ranges given as appropriate

Variable	Population							
	Chizé	Helgeland	Hoedic ^a	Kentucky	Lundy	Nottingham ^b	Oklahoma	Veszprém
Latitude	46N 8' 50"	66N 32' 16"	47N 20' 24"	38N 6' 42"	51N 9' 53"	52N 56' 51"	35N 23' 00"	47N 05' 32"
Longitude	0W 25' 29"	12E 50' 58"	2W 52' 42"	84W 29' 51"	4W 39' 44"	1W 05' 45"	97W 45' 00"	17E 53' 43"
Status	Native	Native	Native	Introduced	Native ^c	Native	Introduced	Native
Location	Mainland	Island	Island	Mainland	Island	Mainland ^d	Mainland	Mainland
Years	2005–2007	2003–2008	2008–2010	1992–2011	2000–2012	1980–1982	2000–2006	2005–2009
Clutches	250	286	81	2125	1220	172	527	168
Clutch size	4.9 \pm 0.8	5.0 \pm 1.0	4.5 \pm 0.8	4.9 \pm 0.9	4.2 \pm 0.8	4.2 \pm 0.7	4.4 \pm 0.8	4.7 \pm 0.8
Date of first egg (Julian)	138 \pm 24	154 \pm 25	139 \pm 19	141 \pm 37	155 \pm 31	150 \pm 29	139 \pm 31	142 \pm 30
Attempts per female per season	(96–199)	(107–212)	(112–176)	(60–220)	(88–224)	(102–222)	(80–207)	(90–210)
Attempts per female per season	1.6 \pm 0.6	1.8 \pm 0.7	1.5 \pm 0.5	2.7 \pm 1.2	2.3 \pm 0.9	1.2 \pm 0.5	1.9 \pm 0.9	2.2 \pm 0.9
Attempts per female per season	(1–3)	(1–4)	(1–2)	(1–6)	(1–6)	(1–3)	(1–4)	(1–4)
Total attempts known per female	3.0 \pm 1.5	3.7 \pm 2.7	2.4 \pm 1.4	7.3 \pm 5.2	6.4 \pm 3.9	1.2 \pm 0.8	4.7 \pm 3.5	2.8 \pm 1.9
Total attempts known per female	(1–6)	(1–11)	(1–6)	(1–26)	(1–17)	(1–4)	(1–14)	(1–9)
Years female bred	1.7 \pm 0.7	1.9 \pm 1.3	1.4 \pm 0.6	2.3 \pm 1.3	2.5 \pm 1.3	1.1 \pm 0.4	1.9 \pm 1.1	1.3 \pm 0.7
Years female bred	(1–3)	(1–5)	(1–3)	(1–6)	(1–6)	(1–2)	(1–5)	(1–4)

^aHoedic population was not followed through the end of breeding in most years.

^bNottingham study occurred on two sites located 40 km apart in Nottinghamshire, England. Coordinates are the midpoint between them.

^cThe Lundy Island population nearly went extinct in 2000, and new birds were introduced. However, for the purposes of this analysis, it was still considered a native population, as the introduced birds were from the nearby mainland, and the island previously supported house sparrows.

^dWhile Britain is an island, the Nottingham study was embedded in a much larger population and so is considered more mainland in key features such as dispersal.

All clutches were assigned an attempt order, referring to their position in the series of clutches produced by a known female in that season. For nearly all clutches in all populations, we also determined the date that the first egg of the clutch was laid. For clutches checked during laying, this was deduced from the fact that house sparrows lay 1 egg per day (Anderson 2006). For successful clutches that were not checked until laying had been completed, we inferred the date the first egg was laid based upon the date they hatched, assuming an 11-day incubation period between the laying of the penultimate egg and hatching (Anderson 2006). For clutches that never hatched but had been checked at least twice during incubation, we estimated the date of first egg as the midpoint of the period between the earliest and latest possible date of first egg. Breeding season was estimated as the range in days between the first egg date of the first clutch and that of the last clutch in that population in that season.

STATISTICAL ANALYSIS

Each clutch laid by each female was considered an observation, and in total, we amassed data on 4871 breeding attempts, with 42 attempts omitted because of unknown clutch sizes due to gaps in egg laying or nest failure before a final clutch size could be determined. We analysed the variation in clutch size using linear mixed-effects models with restricted maximum-likelihood estimation and a Gaussian error structure. We checked the distribution of residual error in the global model and found it to fit well with expectation of normality. Over 80% of the residuals behaved as expected, with only a slight excess of small residuals (Fig. S1).

The initial model of the i th clutch (Y_{ijk}) included three random intercept effects representing the hierarchical nature of the data, and so is useful to understand at what levels key parameters affecting variation in clutch size may act. The three random intercepts included effects of the g th population (pop_{0g}), the k th year nested within population ($year_{0kg}$) and the j th individual nested within population (ind_{0jg}):

$$Y_{ijk} = (\beta_0 + pop_{0g} + year_{0kg} + ind_{0jg}) + e_{0ijk} \quad \text{eqn 1}$$

where e_{0ijk} is the residual clutch size (within-individual) for the i th clutch, and pop_{0g} , $year_{0kg}$ and ind_{0jg} are all normally distributed with mean 0 and variance σ^2 , σ_{pop}^2 , σ_{year}^2 and σ_{ind}^2 , respectively. The statistical significance of a random effect was tested using the likelihood ratio test (LRT) in which twice the difference in log likelihood (-2dLL) between a model (fitted by restricted maximum likelihood) with the random effect and a model lacking that term is distributed as a chi-square with 1 d.f. (Pinheiro & Bates 2000). We also tested whether the within-population individual variance differed between populations by estimating the among-individual variance for each population (e.g. $\sigma_{ind,g}^2$). This required estimating among-individual variance for all eight populations (or seven when we omitted Hoedic from the analysis) and so was tested using an LRT with 7 (or 6) d.f.

We tested for the effect of clutch initiation date and attempt order on the variation in clutch size using 4829 clutches with full information). Although clutch initiation date could be considered a second phenotype (e.g. Husby *et al.* 2010), we treated it as an environmental variable because after the first clutch of the season, subsequent clutch initiation dates are affected by an array of new variables that complicate a multivariate analysis. We created two variables for each of these factors allowing us to assess

between- and within-individual differences in initiation date and attempt order (e.g. van de Pol & Wright 2009). To assess among-individual variance in both variables, we calculated the mean clutch initiation date and mean attempt order over all clutches produced by each female. We centred these values (so that intercepts would be estimated at the mean) by subtracting the yearly population mean date and attempt order from each value (mean-centred-between, or B). To assess within-individual variance, we subtracted the female's mean date and attempt order from that of each of her clutches (mean-centred-within, or W). These variables allowed us to assess clutch size plasticity in response to clutch initiation date, attempt order, and their interaction, and control for between-individual biases in both variables, as depicted in the following model:

$$\begin{aligned} Y_{ijk} = & \beta_0 + pop_{0g} + year_{0kg} + ind_{0jg} + \beta_1 \times dateB_{jkg} \\ & + \beta_2 \times attemptB_{jkg} + \beta_3 \times dateW_{ijk} + \beta_4 \times attemptW_{ijk} \\ & + \beta_{12} \times dateB_{jkg} \times attemptB_{jkg} + \beta_{13} \times dateB_{jkg} \times dateW_{ijk} \\ & + \beta_{24} \times attemptB_{jkg} \times attemptW_{ijk} + \beta_{34} \times dateW_{ijk} \\ & \times attemptW_{ijk} + \beta_{33} \times dateW_{ijk} \times dateW_{ijk} + e_{0ijk} \end{aligned} \quad \text{eqn 2}$$

The variables of within-female date ($dateW$) and within-female attempt order ($attemptW$) were strongly positively correlated ($r = 0.87$). This raised concerns that collinearity would produce spurious results in the analysis. We assessed this by running maximum-likelihood models omitting one or the other of those variables and comparing AIC values. Removing either within-female date or within-female attempt produced substantially larger AIC values ($\Delta\text{AIC} > 2$), indicating that including both improved model fit despite the strong correlation between them. We proceeded to analyse both terms together with their interactions, but we also checked AIC values of models lacking the within-individual date or attempt order to ensure their contribution was not influenced by this correlation.

To test the idea that the reaction norm might differ between populations and to assess how, model 2 was modified to include population as a fixed effect, where both the main effect of population as well as interaction terms with all three within-individual terms (within-individual mean-centred date, attempt order and their interaction) were added. To test specific differences between populations in characteristics (e.g. season length), we altered model 2 by adding season length (determined as the difference in days between the lay date of the earliest nest and that of the last nest for each population-season) and its interactions with within-individual date and attempt order as fixed effects.

All analyses were conducted in SAS 9.2 (SAS Institute 2008). Fixed effects were tested using F tests with denominator degrees of freedom estimated with the Kenward–Roger method, which adjust degrees of freedom to control for repeated measures within the appropriate level of replication (individual within population and year, year within population and population). Code used for the analysis in Table 3 is provided in the Supplementary Material.

Results

PATTERNS OF VARIANCE

The 4829 clutches averaged 4.6 eggs, and the variance in clutch size over the entire data set was 0.85. Clutch size,

dates of first egg and number of attempts per season and per female varied among the eight populations (Table 1). Model 1, which partitioned the variance in clutch size into that among populations, among years within populations, and among individuals within populations, revealed significant variance of all three random effects (Table 2).

REACTION NORM FOR CLUTCH SIZE

Linear mixed model analysis confirmed the basic multidimensional reaction norm reported for Kentucky sparrows by Westneat, Stewart and Hatch (2009). In the full data set of all eight populations, clutch size declined linearly within females with date, also showed a negative quadratic with date, independently increased with attempt order, and the decline in date was more negative as attempt order increased (Table 3). We assessed whether these patterns were driven by the KY data by testing the model without them. However, the results were qualitatively the same as those from the full analysis (i.e. all

Table 2. Partitioning of variance in clutch size into components for 4829 clutches in eight populations of house sparrows from a REML mixed model with three random intercepts and no fixed effects (Model 1 in main text)

Variance component	Estimate \pm SE	Likelihood ratio test*	<i>P</i> -value
Population	0.09 \pm 0.05	54.7	<0.0001
Year (within population)	0.01 \pm 0.005	26.1	<0.0001
Individual (within population)	0.11 \pm 0.01	155.8	<0.0001
Residual (within individuals)	0.64 \pm 0.02	–	–

*Calculated from the fit of the complete model (log likelihood = 12207.5) and a model lacking the focal term (d.f. = 1).

effects were in the same direction and remained statistically significant; results not shown).

We also controlled for individual differences within populations in mean number of attempts and dates of clutch initiation. Females with a larger average attempt order (more attempts per season) produced larger clutches (*attemptB*, Table 3), and those with a later average date of clutch initiation produced smaller clutches (*dateB*, Table 3). Individual females with a later average date of clutch initiation also showed a steeper decline in clutch size as within-individual date progressed (*dateB* by *dateW*; Table 3).

POPULATION DIFFERENCES IN THE CLUTCH SIZE REACTION NORM

Populations differed in several elements of the reaction norm. Because the Hoedic population was not fully studied through the entire breeding season, we omitted it from most of these analyses. Populations had significantly different mean clutch sizes ($F_{6,1801} = 49.0$, $P < 0.0001$, Table S1) and differed significantly in the relationship between date and clutch size within females (Table S1, Fig. 1; *dateW* by population interaction: $F_{6,3020} = 2.6$, $P < 0.02$). In five of the seven populations, clutch size decreased significantly with date within females (Table S1). Both Helgeland and Nottingham exhibited a non-significant positive association between clutch size and date within females. Population identity did not affect the magnitude of the effect of attempt order ($F_{6,3283} = 0.9$, $P = 0.47$, Table S1), which was positive in six of seven populations. The interaction between date and attempt order was negative in all but one population (Chizé), but did not significantly differ among populations ($F_{6,4568} = 1.6$, $P = 0.15$). The nonlinear effect of within-individual variation in date was negative in five of seven populations (positive in Helgeland and Veszprém) but

Table 3. Results of REML linear mixed model of clutch size from 4829 clutches produced by 1512 female house sparrows from eight populations (Model 2 in text)

Factor	Effect \pm SE	F-Statistic	d.f.	<i>P</i> -value
Global intercept (β_0)	4.7 \pm 0.1	–	–	–
<i>AttemptB</i>	0.25 \pm 0.03	51.8	1, 1086	<0.0001
<i>DateB</i>	–0.007 \pm 0.001	48.0	1, 2074	<0.0001
<i>DateB</i> X <i>AttemptB</i>	–0.001 \pm 0.001	0.8	1, 2622	0.38
<i>AttemptW</i>	0.12 \pm 0.03	13.6	1, 3817	0.0002
<i>DateW</i>	–0.009 \pm 0.0008	111.0	1, 3803	<0.0001
<i>DateW</i> X <i>AttemptW</i>	–0.003 \pm 0.0008	13.1	1, 4333	0.0003
<i>DateW</i> X <i>DateW</i>	–0.00009 \pm 0.00002	19.8	1, 4502	<0.0001
<i>DateB</i> X <i>DateW</i>	–0.0003 \pm 0.00004	53.8	1, 3568	<0.0001
<i>AttemptB</i> X <i>AttemptW</i>	0.008 \pm 0.03	0.08	1, 3561	0.78

Population, year within population and individual within population were included as random effects. Fixed effects were mean-centred between-individual date (*DateB*) and attempt number (*AttemptB*) and their interaction, mean-centred within-individual date and attempt (*DateW* and *AttemptW*, respectively) and their interaction, the quadratic of *DateW*, and interactions of within- and between-individual date and attempt order.

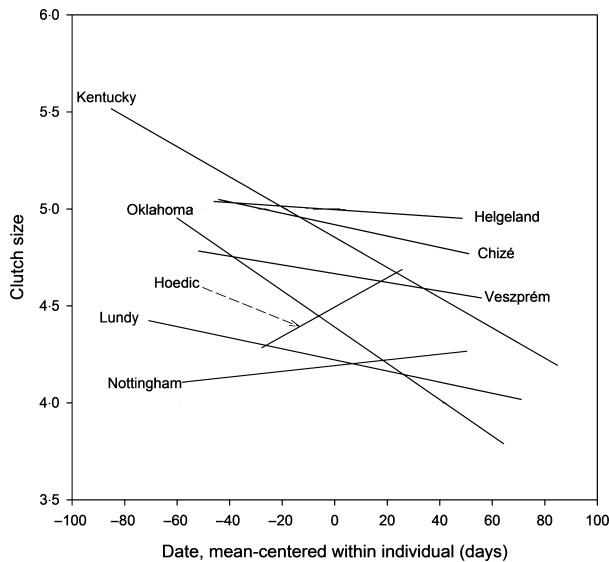


Fig. 1. Population average reaction norms of clutch size with respect to nest initiation date (mean-centred within-individual) for the eight populations of house sparrows studied. Lines are plotted over the range of dates experienced by individuals in each population. Slopes are unadjusted for the influence of other variables and estimated slopes from LMM analysis differ slightly (see Table S1).

also did not differ significantly among populations ($DateW^2$; $F_{6,4517} = 1.3$, $P = 0.26$).

Some of this variability among populations appeared linked to population attributes. In these analyses, we used all eight populations, included population identity as a random effect and the fixed effects as shown in Table 3 excepting the non-significant interaction terms, and tested separately population location (on an island or mainland) and population status (introduced or native). We found no difference in population mean clutch size between island and mainland populations (mainland–island = 0.1 ± 0.3 eggs, $F_{1,6.2} = 0.2$, $P = 0.70$), but mainland populations had a significantly more negative slope with respect to date than did island populations (mainland–island = -0.005 ± 0.002 eggs per day, $F_{1,3807} = 7.7$, $P = 0.005$). The two introduced populations had similar mean clutch sizes to the six native populations (introduced–native = 0.1 ± 0.3 eggs, $F_{1,5.8} = 0.2$, $P = 0.66$), but exhibited significantly more negative slope with respect to date (introduced–native = -0.005 ± 0.002 , $F_{1,3805} = 9.3$, $P = 0.002$). No other interaction terms including either location or status were significant.

We tested whether variation in the length of the breeding season influenced elements of the clutch size reaction norm. We altered model 2 in three ways and ran the modified model to test this idea. First, we dropped all fixed effects terms that were not significant. Secondly, we added season length for each year within each population as a continuous covariate. Finally, we added a season length by within-individual date interaction term. This analysis thus tested the impact of variation in season length,

combining both within- and between-population effects on the term whose impact varied significantly among populations.

We found no main effect of season length on clutch size (-0.001 ± 0.003 eggs per day, $F_{1,65.1} = 0.3$, $P = 0.62$). However, we found that the negative impact of within-individual date was significantly more negative as season length increased (Season length by date: -0.00009 ± 0.00002 eggs per day, $F_{1,3590} = 23.8$, $P < 0.0001$), a result that is hinted at in Fig. 1 (which shows only between-population differences in season length). While adding season length had little qualitative effect on most other terms (fixed and random) in the model shown in Table 3, two fixed terms were substantially different. The significantly negative effect of within-individual date entirely disappeared (0.003 ± 0.003 eggs per day, $F_{1,3612} = 1.2$, $P = 0.27$), suggesting that the previous main effect was driven by results arising from populations or years with longer seasons (See Table S1). Secondly, the significant negative effect of the interaction between within-individual date and attempt order became non-significant (-0.002 ± 0.003 eggs per day per attempt, $F_{1,4237} = 0.5$, $P = 0.47$), possibly because the season length–date interaction may have accounted for some of the variation explained by attempt order and date as season length affects the number of attempts that are possible.

AMONG-INDIVIDUAL VARIANCE IN REACTION NORM PARAMETERS

As was found in Westneat, Stewart and Hatch (2009), female identity explained a significant portion of the variance in clutch size (Table 2) and this persisted regardless of the fixed effects included in the models. In contrast to Westneat, Stewart and Hatch (2009), here, we controlled for potential biases due to individual females producing clutches at different times in the season or having different numbers of attempts. Both individual mean date of clutch initiation and individual mean attempt order had significant effects on clutch size (Table 3), but both had roughly similar effect sizes as within-female date and attempt order, suggesting that they may arise from the same underlying reaction norm. Despite this, female identity (random effect of individual in model 2) continued to have a significant effect on variance in clutch size, accounting for 14% (0.12) of the total variance (0.85) in clutch size.

We tested whether among-individual variance in clutch size differed between populations by comparing a modification of the model in Table 3 (population omitted as a random effect, included as a fixed effect along with a population by $dateW$ interaction) with one in which the random effect of female identity was split into separate estimates from each of the eight populations (Fig. 2). A LRT revealed that the model with the separate population estimates was a significantly better fit than one with a single estimate of among-female variation ($-2dLL = 23.5$,

d.f. = 7, $P = 0.0006$), indicating that populations differed in among-female variance in clutch size after controlling for within-individual plasticity and between-individual differences in timing. The model could not estimate variance in the Nottingham population (Fig. 2), possibly due to small samples sizes for both the number of individuals and the number of clutches per individual.

We also assessed potential among-individual variance in slopes with respect to both date and attempt order by adding a random slope term to the model. We found no evidence for differences in individual slope with respect to *dateW* (estimate = 0.00006 ± 0.00005 , $-2dLL = 2.6$, d.f. = 2, $P = 0.30$). We also found no evidence that the multidimensionality of the reaction norm due to the interaction between date and attempt order varied among individuals ($-2dLL = 0.3$, d.f. = 2, $P = 0.86$). By contrast, we found significant between-individual differences in slope with respect to within-individual attempt order (0.014 ± 0.006 , $-2dLL = 9.7$, d.f. = 2, $P = 0.01$), with the estimated covariance between slope and intercept slightly negative (-0.004 ± 0.007). Estimating the individual random slope term for each population did not significantly improve the fit of the model ($-2dLL = 6.8$, d.f. = 6, $P = 0.34$).

Discussion

Our results from multiple populations confirm that clutch size in house sparrows exhibits a complex form of phenotypic plasticity in which the reaction norm is a warped plane in multidimensional environmental space. This

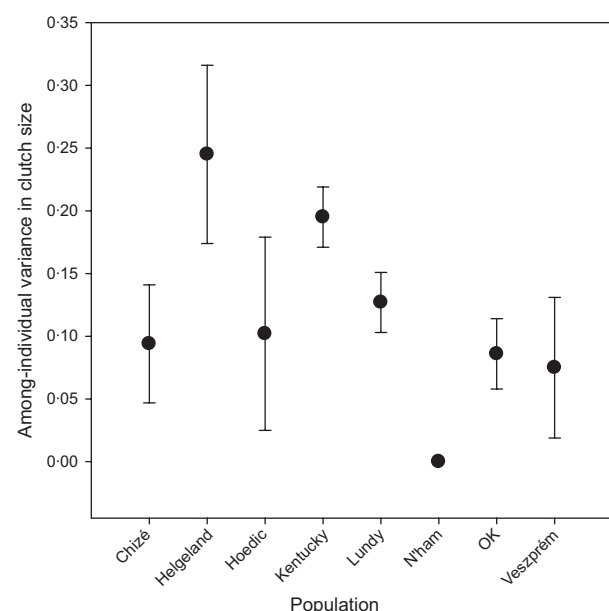


Fig. 2. Estimates of population-specific among-individual variance in clutch size from the full REML mixed model including both within- and among-individual fixed effects (date and attempt order and their interaction; Table 3). Error bars are estimated standard errors.

reaction norm varied among populations but not in the way predicted by any of the hypothesized forces thought to drive selection on the shape of the optimal reaction norm. The reaction norm is not as expected if seasonal declines in clutch size are driven by changing food resources, risk of predation or effects on parental residual reproductive value. The attributes of this reaction norm, with a decline in clutch size with date, an increase in clutch size with attempt order, and a more negative clutch size–date relationship with increasing attempt order, fitted that expected under the time horizon hypothesis (Rowe, Ludwig & Schluter 1994). However, season length did not have the expected positive effect on the date–clutch size relationship and was instead significantly negative. Our results also show that populations differed in the nature of individual variation in reaction norm parameters. This may impact understanding the evolution of reaction norms in general and the specific forces potentially acting on plasticity in clutch size in birds. We explore these implications in more detail.

COMPARATIVE ANALYSES OF REACTION NORMS

The clutch size reaction norm, as measured across eight populations of house sparrows, exhibits several intriguing properties. Populations differed in intercept (elevation), and there was significant repeatability of elevation within individuals across all populations after controlling for some potential biases due to plasticity. However, the level of repeatability in clutch size differed between populations. Because within-population, among-individual variation could reflect the level of underlying genetic variation (Nussey, Wilson & Brommer 2007), differences in repeatability could reflect differences in evolutionary potential. Between-individual differences in any phenotypic trait could arise from two sources: genetic variation or unaccounted environmental variation acting throughout an individual's reproductive life span. The characteristics of populations showing high among-individual variance support the latter hypothesis over the former for two main reasons. North American populations of house sparrows have lower levels of genetic diversity (as measured at presumably neutral microsatellite loci) than do European populations (Schrey *et al.* 2011). We tested whether introduced populations (the two from North America) had a different among-individual variance than did native populations and found no evidence for a difference (LRT, $\chi^2 = 0.1$, d.f. = 1, $P = 0.75$). Island populations also generally have lower levels of genetic variation than mainland populations in many species (e.g. Frankham 1997). However, we found no difference in the among-individual variation in clutch size between island and mainland populations (LRT, $\chi^2 = 0.8$, d.f. = 1, $P = 0.37$).

A second reason for suspecting that population-level differences in between-individual variation may be due to unaccounted environmental variation is that populations differed in several variables that contributed to

phenotypic plasticity in clutch size. For example, based on the number of years bred, populations might vary in their age distributions (Table 1). Clutch size increased with number of years a female was present in our data set (best estimate of female age that we have; model as in Table 3 with addition of female years and the Nottingham population omitted; effect = 0.10 ± 0.01 , $F_{1,3465} = 42.2$, $P < 0.0001$), but this did not eliminate the differences between populations in among-female variance ($-2dLL = 20.9$, d.f. = 5, $P < 0.001$). Nevertheless, females in some locations could be experiencing a wider range of other environmental conditions. Because we have not accounted for all environmental factors in our analysis, it seems likely that differences in environmental variance may be the cause of population differences in between-individual variance.

The evolution of plasticity requires variation in reaction norm slope, and our results suggest complexity in patterns of variation in slopes. An especially interesting element of the clutch size reaction norm in sparrows is the interaction between date and attempt order. In most populations, this is negative (although it is significantly so in only two, Table S1), meaning that as females have produced more prior clutches, there is a more negative effect of date. This interaction term produces the non-additive feature of the multidimensional reaction norm and is a unique prediction of the Rowe, Ludwig and Schluter (1994) model of optimal clutch size. To evolve, there must be within-population variation in this interaction, but we found no support for individuals differing in this parameter of the reaction norm. We strongly suspect this is because of statistical power rather than biology; while we have data on a total of 1512 females, for only 107 did we have three or more clutches (sufficient to measure slope with some residual variance) for each of three attempt orders. To reduce the impact of sampling variance on the residual variance, we would need even more clutches per attempt order, and consequently, sample size drops considerably. Power to detect among-individual variation in this parameter of the reaction norm is thus likely to be quite low (e.g. van de Pol 2012).

Our power to detect variance among individuals in univariate slopes is considerably greater, and our results from these analyses raise some interesting questions regarding the evolution of plasticity. First, we found that populations differed significantly in the average slope of clutch size with date. Divergence in slope is expected when plasticity is under different selection pressures in different populations. Divergent selection and an evolutionary response are only possible if there are individual, heritable, differences in slope, yet we failed to detect significant among-individual variation in this slope. We did uncover significant among-individual variation in slope with respect to attempt order, but found no evidence of among-population variance in slope. Thus, our results seem to indicate a paradoxical situation; for the parameter of the clutch size reaction norm (slope with respect to

date) that appears to have diverged between populations, there is no evidence of the individual variation within populations that is necessary for selection to lead to such divergence. Conversely, for the parameter of the reaction norm (slope with respect to attempt order) that exhibits the necessary variance among individuals, no divergence between populations has apparently occurred.

There are many possible explanations for this situation, but we focus on two that seem especially interesting. One is that slope with respect to attempt order is under little or no selection in all populations, thereby retaining individual variation and limiting divergence, whereas slope with respect to date is under strong stabilizing selection with some directional selection, and the combination has eliminated present sources of individual variation but also has led to divergence. Testing this would require data on selection and heritability (additive genetic variance) of clutch size from each population, which we do not have at present. Another hypothesis that deserves more attention in cases of plasticity is the action of additional, unaccounted environmental variables. Most studies of reaction norms assess slopes in only a single environmental axis (Pigliucci 2001; Brommer, Pietiäinen & Kokko 2002; Postma & van Noordwijk 2005; Nussey, Wilson & Brommer 2007). We explicitly analysed the clutch size reaction norm as a plane in bivariate environmental space. Westneat, Stewart and Hatch (2009) considered additional environmental variables (e.g. precipitation) but found no phenotypic association. It is conceivable, however, that clutch size responds to variables that are not captured by either date or attempt order. If these differ among populations in ways that generate both differences between individuals within populations and average effects that differ between populations, then multidimensional plasticity could explain the patterns of variation we observed. Testing this idea would require a more detailed understanding of the ecology of clutch size and within-individual plasticity to identify this unknown variable (or variables) and measurements of them within and among populations.

THE LIFE HISTORY OF CLUTCH SIZE

Our results raise new questions about life history and clutch size. We tested whether the differences between populations in reaction norm slopes with respect to date (Table 3, Fig. 1) might distinguish among various hypotheses for why clutch size declines with date. As found in Westneat, Stewart and Hatch (2009), house sparrows exhibit a complex reaction norm that is best explained by the time horizon hypothesis. We reasoned this support would be stronger if season length had a positive effect on the relationship between clutch size and date (less negative in longer seasons). We found the opposite; the steeper declines in clutch size occurred in populations with longer breeding seasons. This result raises questions about the assumption for the time horizon hypothesis that

offspring quality declines with date. Data on recruitment rate have been published for two of the populations we studied; in Helgeland, Ringsby *et al.* (2002) found that recruitment probability increased initially and then declined with hatch date (Husby *et al.* 2006). In Oklahoma, recruitment rate appears to be influenced by the quality of food items provided by parents, which increases with date; when food quality is held constant, recruitment declines with date (Schwagmeyer & Mock 2008). This is an intriguing coincidence given the relationships between clutch size and date in the two populations (Fig. 1), with Helgeland having a flat relationship and Oklahoma steeply negative, but we clearly need more information from all populations, including these two, to understand the relationship between juvenile survival and clutch size.

It is possible the time horizon applies to some populations more than others, but if so, why? One possibility is that the key relationship driving the time horizon effect is more complex than assumed in the Rowe, Ludwig and Schluter (1994) model. Brommer, Pietiäinen and Kokko (2002) noted that relatively subtle differences in the relationship of offspring fitness with date could have large effects on the outcome of the model. The model can be simplified to the following equation, which is the condition that must be met for the optimal time–clutch combination:

$$\frac{C'(t)}{C(t)} = \frac{V'(t)}{V(t)} \quad \text{eqn 3}$$

where $C(t)$ is the clutch size at time t and $V(t)$ is the recruitment probability at time t , and C' and V' are the rates of change in clutch size or recruitment probability with t . In our analysis of variation in breeding season length and clutch size, we assumed that $V(t) = 0$ at the end of the breeding season and that peak recruitment [$\max V(t)$] was the same in all populations; thus, season length would be collinear with $V(t)$. Brommer, Pietiäinen and Kokko (2002) noted in their study that differences in $\max V(t)$ alone could create differences in $C'(t)$. This could affect our results as well. However, Rowe, Ludwig and Schluter (1994) assumed that parents stop breeding when $V(t) = 0$. We suggest that this assumption may be invalid because other factors may influence when parents stop breeding, ensuring that $V(t)$ does not equal zero and altering the relationship between $V'(t)$ and season length. For example, if independent offspring must have time to acquire skills at finding food (e.g. seeds) before peak times of food stress (e.g. Loman 1982; Hochachka 1990), then it is possible the decline in offspring fitness over the breeding season depends on when this food stress occurs relative to the end of the breeding season. If the insect food fed to nestlings declines early such that parents cease breeding long before the period of food stress that juveniles might experience, then breeding date may have little impact on clutch size. The relative timing and rate of declines in insect food and periods of food stress could produce complex differences in reaction norms for clutch

size between populations. Alternatively but similarly, house sparrows moult once per year in autumn (Anderson 2006). Late-hatched juveniles might experience increased mortality if their moult is delayed into increasingly colder weather or if they have to moult quickly and produce poorer feathers as a result. As with the first hypothesis, differences between populations in when breeding ceases and when moult is optimal could create selection favouring a different decline in clutch size with date.

Other possibilities exist, although all of them would require more detailed information on the underlying basis of presumed seasonal declines in recruitment. Such data are not presently available for the populations in this study. Moreover, the details are likely to matter. Consider the Oklahoma population, the southernmost locale in our analysis. Females there started breeding early in the season, but also ceased breeding in early July (Table 1), possibly because the hot and dry conditions of late July and August in Oklahoma increase the costs of breeding for adults. Because of its southern location, however, there is a long delay between breeding and cold weather. Because of this delay, we would expect that late-hatched juveniles should be in less of a time crunch and so should have similar survival to early-hatched juveniles. The Rowe, Ludwig and Schluter (1994) model would thus predict that date would have little effect on clutch size in Oklahoma, but in fact, the Oklahoma birds have the steepest decline with date. This implies that either our presumption about the timing of stressful conditions for juveniles (e.g. when cold weather arrives) is incorrect, or that other processes linked to date are affecting clutch size.

Finally, clutch size is likely to be part of an integrated phenotype (Pigliucci 2003) that includes when a female begins breeding within each season and how many attempts she has. In our analyses, we treated attempt and date as environmental factors, but some of their variation is likely due to variation in female phenotype. For example, the timing of a female's first attempt of the season also exhibits phenotypic plasticity and shows among-individual variance in plasticity (e.g. Brommer *et al.* 2005; Brommer, Rattiste & Wilson 2007; Husby *et al.* 2010), which sometimes has a genetic basis (e.g. Charmantier *et al.* 2008). Earlier breeding in a multibrooded species such as the house sparrow is typically associated with more attempts and earlier initiation dates for each attempt (Anderson 2006). Both date and attempt order in house sparrows are themselves traits that could also be sensitive to environmental conditions. Thus, phenotypic integration of date of first breeding with clutch size is likely to exist. If so, then among-individual variation in plasticity associated with date of first breeding (e.g. spring temperature) could exist. Selection acting on the decision to start breeding may influence reaction norm shape for clutch size during later attempts. A looming challenge then is to assess the level of phenotypic integration within and among populations simultaneously with independent effects of environment on life-history traits such as clutch size.

Data Accessibility

Data available from the Dryad Digital Repository: <http://doi.org/10.5061/dryad.112cs>.

Acknowledgements

We thank the large number of field personnel across eight studies who helped contribute to this data set. We also thank the multiple agencies that supported this work, including the U.S. National Science Foundation (ÁZL, DFW, DM, IRKS and PLS), the Norwegian Research Council (HJ and TK), NERC (TB and JS), Hungarian Scientific Research Fund (Grants no. T047256, K72827, K84132, PD76862; AL, ÁZL and VB), Hungarian Scholarship Board [ÁZL, the CRBPA (OC)] and the French ANR (GS). The lead author also thanks the University of Kentucky for support during preparation of this manuscript and the Norwegian University of Science and Technology for hosting him during a sabbatical when plans for this paper took shape. We also appreciate the useful suggestions on the manuscript provided by three anonymous reviewers.

References

- Anderson, T.R. (2006) *The Ubiquitous House Sparrow: From Genes to Populations*. Oxford University Press, New York, New York, USA.
- Bennett, P.M. & Owens, I.P.F. (2002) *Evolutionary Ecology of Birds: Life Histories, Mating Systems, and Extinction*. Oxford University Press, Oxford, UK.
- Bókonyi, V., Liker, A., Lendvai, A.Z. & Kulcsár, A. (2008) Risk-taking and survival in the House Sparrow *Passer domesticus*: are plumage ornaments costly? *Ibis*, **150**, 139–151.
- Boutin, S. (1990) Food supplementation experiments with terrestrial vertebrates: patterns, problems, and the future. *Canadian Journal of Zoology*, **68**, 203–220.
- Bradshaw, A.D. (1965) Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, **13**, 115–155.
- Brommer, J.E., Pietiäinen, H. & Kokko, H. (2002) Cyclic variation in seasonal recruitment and the evolution of the seasonal decline in Ural owl clutch size. *Proceedings of the Royal Society of London B, Biological Sciences*, **269**, 647–654.
- Brommer, J.E., Rattiste, K. & Wilson, A.J. (2007) Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull *Larus canus*. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 687–693.
- Brommer, J.E., Merilä, J., Sheldon, B.C. & Gustafsson, L. (2005) Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution*, **59**, 1362–1371.
- Burke, T. (1984) *The ecological genetics of two populations of the house sparrow, Passer domesticus*. PhD thesis, University of Nottingham, Nottingham, UK.
- Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., 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.
- Chastel, O. & Kersten, M. (2002) Brood size and body condition in the House Sparrow *Passer domesticus*: the influence of brooding behavior. *Ibis*, **144**, 284–292.
- Cleasby, I., Nakagawa, S., Gillespie, D.O.S. & Burke, T. (2010) The influence of sex and body size on survival recruitment in the house sparrow. *Biological Journal of the Linnean Society*, **101**, 680–688.
- Clifford, L.D. & Anderson, D.J. (2001) Food limitation explains most clutch size variation in the Nazca booby. *Journal of Animal Ecology*, **70**, 539–545.
- Crick, H.Q.P., Gibbons, D.W. & Magrath, R.D. (1993) Seasonal changes in clutch sizes in British birds. *Journal of Animal Ecology*, **62**, 263–273.
- Decker, K.L., Conway, C.J. & Fontaine, J.J. (2012) Nest predation, food, and female age explain seasonal declines in clutch size. *Evolutionary Ecology*, **26**, 683–699.
- Dingemanse, N.J. & Wolf, M. (2013) Between-individual differences in behavioural plasticity: causes and consequences. *Animal Behaviour*, **85**, 1031–1039.
- Drent, R.H. & Daan, S. (1980) The prudent parent: energetic adjustments in avian breeding. *Ardea*, **68**, 225–252.
- Frankham, R. (1997) Do island populations have less genetic variation than mainland populations? *Heredity*, **78**, 311–327.
- Ghalambor, C.K., Angeloni, L.M. & Carroll, S.P. (2010) Behavior as phenotypic plasticity. *Evolutionary Behavioral Ecology* (eds D.F. Westneat & C.W. Fox), pp. 90–130. Oxford University Press, New York, New York, USA.
- Gomulkiewicz, R. & Kirkpatrick, M. (1992) Quantitative genetics and the evolution of reaction norms. *Evolution*, **46**, 390–411.
- Gotthard, K. & Nylin, S. (1995) Adaptive plasticity and plasticity as an adaptation: a selective review in morphology and life history. *Oikos*, **74**, 3–17.
- Hochachka, W.M. (1990) Seasonal decline in reproductive performance of song sparrows. *Ecology*, **71**, 1279–1288.
- Husby, A., Sæther, B.-E., Jensen, H. & Ringsby, T.H. (2006) Causes and consequences of adaptive seasonal sex ratio variation in house sparrows. *Journal of Animal Ecology*, **75**, 1128–1139.
- Husby, A., Nussey, D.H., Visser, M.E., Wilson, A.J., Sheldon, B.C. & Kruuk, L.E.B. (2010) Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution*, **64**, 2221–2237.
- Jensen, H., Steinsland, I., Ringsby, T.H. & Sæther, B.-E. (2008) Evolutionary dynamics of a sexual ornament in the house sparrow (*Passer domesticus*): the role of indirect selection within and between the sexes. *Evolution*, **62**, 1275–1293.
- Klomp, H. (1970) The determination of clutch-size in birds: a review. *Ardea*, **58**, 1–124.
- Lack, D. (1947) The significance of clutch size. *Ibis*, **89**, 302–352.
- Lande, R. (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, **22**, 1435–1448.
- Lendvai, A.Z. & Chastel, O. (2010) Natural variation in stress response is related to post-stress parental effort in male house sparrows. *Hormones and Behavior*, **58**, 936–942.
- Loman, J. (1982) A model of clutch size determination in birds. *Oecologia*, **52**, 253–257.
- Martin, T.E. (1987) Food as a limit on breeding birds: a life history perspective. *Annual Review of Ecology and Systematics*, **18**, 453–487.
- Murphy, M.T. (1986) Temporal components of reproductive variability in eastern kingbirds (*Tyrannus tyrannus*). *Ecology*, **67**, 1483–1492.
- Nager, R.G., Rueger, C. & van Noordwijk, A. (1997) Nutrient or energy limitation on egg formation: a feeding experiment in great tits. *Journal of Animal Ecology*, **66**, 495–507.
- van Noordwijk, A.J. (1989) Reaction norms in genetical ecology: studies of the great tit exemplify the combination of ecophysiology and quantitative genetics. *BioScience*, **39**, 453–458.
- Nussey, D.H., Wilson, A.J. & Brommer, J.E. (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, **20**, 831–844.
- Pettifor, R.A., Perrins, C.M. & McCleery, R.H. (2001) The individual optimization of fitness: variation in reproductive output, including clutch size, mean nestling mass and offspring recruitment, in manipulated broods of great tits *Parus major*. *Journal of Animal Ecology*, **70**, 62–79.
- Pigliucci, M. (2001) *Phenotypic Plasticity: Beyond Nature and Nurture*. The Johns Hopkins University Press, Baltimore, Maryland, USA.
- Pigliucci, M. (2003) Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecology Letters*, **6**, 265–272.
- Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, New York, USA.
- van de Pol, M. (2012) Quantifying individual variation in reaction norms: how study design influences the accuracy, precision, and power of random regression models. *Methods in Ecology and Evolution*, **3**, 268–280.
- van de Pol, M. & Wright, J. (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, **77**, 253–258.
- Postma, E. & van Noordwijk, A.J. (2005) Genetic variation in clutch size in natural populations of birds from a reaction norm perspective. *Ecology*, **86**, 2344–2357.
- Ringsby, T.H., Sæther, B.-E., Tufto, J., Jensen, H. & Solberg, E.J. (2002) Asynchronous spatiotemporal demography of a house sparrow metapopulation in a correlated environment. *Ecology*, **83**, 561–569.

- Rowe, L., Ludwig, D. & Schluter, D. (1994) Time, condition, and the seasonal decline in avian clutch size. *The American Naturalist*, **143**, 698–722.
- SAS Institute (2008) *SAS User's Guide: Statistics*. SAS Institute, Cary, North Carolina, USA.
- Scheiner, S.M. (1993) Genetics and evolution of phenotypic plasticity. *Annual Reviews in Ecology and Systematics*, **24**, 35–68.
- Schrey, A., Grispo, M., Awad, M., Cook, M., McCoy, E., Mushinsky, H. *et al.* (2011) Broad-scale latitudinal patterns of genetic diversity among native European and introduced house sparrow (*Passer domesticus*) populations. *Molecular Ecology*, **20**, 1133–1143.
- Schwagmeyer, P.L. & Mock, D.W. (2003) How consistently are good parents good parents? Repeatability of parental care in the house sparrow, *Passer domesticus*. *Ethology*, **109**, 303–313.
- Schwagmeyer, P.L., Mock, D.W. & Parker, G.A. (2008) Parental provisioning and offspring fitness: size matters. *Animal Behaviour*, **75**, 291–298.
- Siikamäki, P., Hovi, M. & Rätti, O. (1994) A trade-off between current reproduction and moult in the pied flycatcher—an experiment. *Functional Ecology*, **8**, 587–593.
- Stearns, S.C. (1989) The evolutionary significance of phenotypic plasticity. *BioScience*, **39**, 436–445.
- 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.
- Westneat, D.F., Stewart, I.R.K. & Hatch, M.I. (2009) Complex interactions among temporal variables affect the plasticity of clutch size in a multi-brooded bird. *Ecology*, **90**, 1162–1174.
- Williams, T.D. (2005) Mechanisms underlying the costs of egg production. *BioScience*, **55**, 39–48.
- Woltereck, R. (1909) Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden. *Verhandlungen des Deutsch-Tschechischen Zoologischen Gesellschaft*, **19**, 110–172.

Received 2 September 2013; accepted 20 November 2013

Handling Editor: Daniel Ardia

Supporting Information

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

Fig. S1. Analysis of residual clutch size of female house sparrows from a fitted model (Table 3) showing a generally good fit to a normal distribution but with a slight excess of smaller values.

Table S1. Individual population estimates for the within-female parameters in the clutch size reaction norm.