Behavioral Ecology, 2024, Vol. 35, No. 4, arae042
https://doi.org/10.1093/beheco/arae042
Advance access publication 23 May 2024
Original Article

Scope and adaptive value of modulating aggression over breeding stages in a competitive female bird

Elizabeth M. George[1,2,*], Abigail M. Weber[1,3], Kimberly A. Rosvall[1]

[1]Department of Biology, Indiana University, 1001 E. Third St., Bloomington, IN, USA
[2]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Aronoff Lab Rm 300, 318 W. 12th Ave., Columbus, OH, USA
[3]Department of Zoology, Southern Illinois University Carbondale, Life Science II Rm 351, Carbondale, IL, USA

[*]Corresponding author: Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Aronoff Lab Rm 300, 318 W. 12th Ave., Columbus, OH, USA. Email: e.george914@gmail.com
Handling Editor: Kees van Oers

In seasonally breeding animals, the costs and benefits of territorial aggression should vary over time; however, little work thus far has directly examined the scope and adaptive value of individual-level plasticity in aggression across breeding stages. We explore these issues using the tree swallow (*Tachycineta bicolor*), a single-brooded bird species in which females compete for limited nesting sites. We measured aggressiveness in nearly 100 females within 3 different stages: (1) shortly after territory-establishment, (2) during incubation, and (3) while caring for young chicks. Based on the timing, direction, and magnitude of behavioral changes between stages, we used k-means clustering to categorize each female's behavior into a "plasticity type." We then tested whether plasticity type and stage-specific aggression varied with key performance metrics. About 40% of females decreased aggressiveness across consecutive breeding stages to some degree, consistent with population-level patterns. 33% of females exhibited comparatively little plasticity, with moderate to low levels of aggression in all stages. Finally, 27% of females displayed steep decreases and then increases in aggression between stages; females exhibiting this pattern had significantly lower body mass while parenting, they tended to hatch fewer eggs, and they had the lowest observed overwinter survival rates. Other patterns of among-stage changes in aggressiveness were not associated with performance. These results reveal substantial among-individual variation in behavioral plasticity, which may reflect diverse solutions to trade-offs between current reproduction and future survival.

Key words: age, behavioral plasticity, female–female aggression, individual variation, reaction norm, territorial aggression, tree swallow.

Introduction

When animals compete over limited resources, aggression may increase their likelihood of success (Archer 1988; Rosvall 2011b; Fuxjager et al. 2017). However, aggressive interactions also risk injury (Jaeger 1981), death (Gómez et al. 2016), or more subtle forms of harm, such as DNA damage via oxidative stress (Georgiev et al. 2015). Furthermore, the time, energy, and/or physiological mechanisms that facilitate aggression may induce trade-offs with other critical functions (Wingfield et al. 2001). Selection should, therefore, favor aggressive phenotypes in systems or individuals for which the benefits are great enough to offset these costs.

Critically, these costs and benefits may not be fixed over time, even within the same individual in a single breeding season. The benefits of aggression are driven in part by the value of resources over which individuals compete. However, resource value is influenced by its quality and abundance, which can change with various environmental conditions (Yang et al. 2008). Resource value can also affect the costs of aggression, since potentially higher benefits can lead to more intense contests with greater risk of injury or death (Enquist and Leimar 1987). Costs of aggression also may vary in relation to energy intake (e.g. food availability; Marler and Moore 1991) or trade-offs with parental care (e.g. McGlothlin et al. 2007). The dynamic nature of the myriad ecological factors influencing aggression highlights a key question: when selective pressures acting on a behavior vary temporally, what behavioral phenotypes occur in a population, and which of these are adaptive?

Selection could favor individuals that flexibly adjust their behavioral phenotypes across contexts, in other words, those that exhibit behavioral plasticity (Snell-Rood 2013). For example, individual zebrafish (*Danio rerio*) become more aggressive and bolder when placed in tanks with more vegetation and reduced water flow, respectively (Bhat et al. 2015). Likewise, individual tree swallows (*Tachycineta bicolor*) adjust their anti-predator behavior in response to apparent food availability (Betini and Norris 2012), and individual male field crickets (*Gryllus bimaculatus*) adjust their aggression in response to being housed with other males (Stevenson and Rillich 2013). Many seasonally-breeding animals also adjust aggression based on their reproductive stage, with higher aggression seen during initial territory establishment or courtship and lower aggression while parenting (Araya-Ajoy and Dingemanse 2017; Bowman et al. 2018). Even within a population, not all individuals change aggression in the same way (Dingemanse and Wolf 2013; Alonzo 2015; Stamps 2016; Biro et al. 2018; Laskowski et al. 2021). While there is growing recognition that individual variation in behavioral plasticity is widespread, its causes and consequences are poorly understood (Dingemanse and Wolf 2013; Alonzo 2015; Stamps 2016; Biro et al. 2018; Laskowski et al. 2021).

Received: 11 June 2023; **Revised:** 2 May 2024; **Editorial decision:** 7 May 2024; **Accepted:** 19 May 2024.

Female tree swallows are a good system to investigate how individuals adjust their aggressiveness across contexts. These migratory songbirds are single-brooded (Winkler et al. 2020) and highly synchronous breeders (Stutchbury and Robertson 1987b). They are obligate secondary cavity nesters, which means they require a natural or artificial cavity (i.e., a nest box) to reproduce but cannot excavate a nest site. Females do not defend a general-purpose or feeding territory; rather, they only defend the nesting site and the surrounding 20-30 m (Muldal et al. 1985). Nest boxes also afford some standardization in resource value because all boxes are identical and placed in high quality habitat (i.e. open wet meadows). When nest sites are limited, as is often the case (Holroyd 1975), a female's aggressiveness predicts her likelihood of acquiring a nesting cavity during initial territory establishment each spring (Rosvall 2008). Territorial aggression can be easily measured in the field using standardized resident-intruder assays, and an individual female's aggressiveness is highly consistent (repeatable) within a breeding stage and independent of her male partner's aggressiveness (Rosvall 2008). Maintaining aggressiveness beyond territorial establishment may be advantageous because there are often high numbers of non-territory-holding, or "floater," females that intrude upon claimed territories throughout the breeding season (Stutchbury and Robertson 1987a) and can evict or kill territory-holders (Leffelaar and Robertson 1985). On the other hand, costs of territorial aggression may increase as a breeding season progresses, as females face greater energetic demands of incubating eggs and provisioning nestlings. For example, female aggression and provisioning rates can be negatively correlated (Rosvall 2011a), and when female aggression is experimentally enhanced via testosterone implants, incubation behavior declines (Rosvall 2013). These patterns suggest that there may be opportunities for selection to act on the degree to which females adjust aggressiveness across different breeding stages. Consistent with this view, prior work in tree swallows shows that female aggression declines over a breeding season at the *population* level (Bentz et al. 2019a). However, we do not know the extent of *individual* variation in the degree of plasticity, and we do not know if such changes in aggression over the course of a breeding season are in fact adaptive.

To address these questions, we assayed the aggressiveness of nearly 100 female tree swallows within 3 breeding stages: (1) shortly after territory establishment, (2) during early incubation, and (3) during chick-rearing. We explored variation in plasticity by quantifying repeatability across stages and by statistically categorizing females' behavioral changes based on the timing, direction, and magnitude of aggression changes across these stages within a single breeding year. We call these categories "plasticity types," acknowledging that it is unknown how stable these types are across a lifetime. An important first step in understanding the biological relevance of behavioral plasticity is to explore its performance-related correlates. To this end, we examined whether a female's plasticity type or her aggression at particular stages related to her current age, condition, survival, and reproductive success in the current year. Based on the aforementioned costs and benefits of aggression specific to each breeding stage, we predicted that most, if not all, individuals would exhibit higher aggressiveness in earlier breeding stages than later breeding stages, and those who exhibited greater decreases in aggression would also exhibit higher reproductive quality or performance. Such a correlative result would be consistent with the fitness consequences of particular behavioral phenotypes. Alternatively, if an individual's stage-to-stage plasticity is unrelated to reproductive quality or performance, this would instead suggest there may be multiple routes to success. These analyses represent key steps toward understanding the scope and biological relevance of behavioral plasticity.

Methods

Monitoring the study population

This study used females breeding in nest boxes near Bloomington, Indiana, USA (39°9 N, 86°31 W) in springs 2018 (hereafter Y1) and 2020 (hereafter Y2); we did not collect data for this project in 2019 for logistical reasons. This nest box population has at least 100 tree swallows breeding each year.

Our analyses focus on 101 separate breeding attempts (Y1 n = 31, Y2 n = 70). All focal females were on their first or only breeding attempt of the year, selected at random from among the larger study population. Thus, our study does not include females that failed to obtain a nest box or failed to keep that nest box at least through the early chick period (see *Discussion*). Four females were included in both years' datasets. Excluding one of their breeding seasons does not fundamentally change any results, and so we retained these points.

We checked nest boxes every 1 to 4 d (dependent on nesting stage). We documented the progression of nest construction and determined the exact dates for egg-laying, the onset of incubation, and hatching (denoted as postnatal D1). All breeding females are banded annually (elaborated below), and we checked nests on postnatal D12 to count and measure their chicks and again on ~D21 to assess fledging success.

Measuring aggression at multiple breeding stages

Defining breeding stages

We assayed aggression during 3 distinct breeding stages, which were determined based on nest box contents and direct observation of the females themselves:

1. **Shortly after** Terr**itory** Est**ablishment** (hereafter "TerrEst"). Nest box ownership during this phase is readily apparent via direct observation, as females arrive at their nest sites in the early morning, defend it from other birds, and leave only to forage and roost (Winkler et al. 2020). In addition, as females claim a nesting site, they place a small amount of nesting material inside (e.g. 1 to 2 large, downy feathers or a few pieces of hay), and they may begin the earlier phases of nest building. We did not include putatively fertile females who were producing eggs because associated changes in body size may affect aerial abilities (Lee et al. 1996), and most aggressive behaviors occur in flight. Therefore, we excluded aggression data collected in the 5 d before a female initiated laying (determined post-hoc).
2. Inc**ubation** (hereafter "Inc"). Females begin incubation on the day they lay their penultimate egg, and hatch day typically occurs on the 14th day of incubation. We standardized the timing of this aggression assay to early incubation (average = 3.3 ± 0.4 d from last egg laid, range = 1 to 7 d).
3. C**hick-rearing** (hereafter "C"). Females and males care for chicks, who fledge ~3 wk after hatching (Winkler et al. 2020). We measured aggression early in this chick-rearing period (average = 2.8 ± 0.6 d post-hatch, range = 2 to 5 d).

Aggression assays

To measure each female's stage-specific aggressiveness, we assayed her territorial aggression in response to a standardized resident-intruder assay (sensu Wingfield 1985; van Duyse et al. 2004; George et al. 2022). Specifically, we placed a conspecific female decoy on a nest box for 5 min and recorded the focal female's behavioral response. Decoys were made from skins/feathers, affixed to a 3D-printed swallow body, and positioned directly below, but not occluding, the entrance to a nest box in defensive postures matching those exhibited by actual intruders (described in Bentz et al. 2019a). Skins were sourced from birds collected for our genomics work (e.g. Bentz et al. 2019b). All skins were from 2+ yr old females with iridescent blue/green plumage because some studies have found that younger (brown) females evoke less aggression (Coady and Dawson 2013; but see Stutchbury and Robertson 1987c). We rotated among 34 separate decoys to minimize pseudoreplication (Y1 $n = 18$; Y2 $n = 16$, no overlap between years). We also played an audio file containing aggressive tree swallow calls. Audio files were made by repeating a single vocalization clip several times within a file so that one playback file consisted of a [10 s call, 10 s silence] sequence repeated 15 times. We rotated among 6 audio files made from 6 separate clips, which were each recorded from unique individuals in our population in years prior to the study (2016 to 2017). Audio was played using a Sandisk Sportclip mp3 player and a portable speaker (Altec Lansing), which were attached to the nest box, within 0.5 m of the box entrance. Playbacks were normalized to naturalistic volume. We have previously found no effects of decoy or playback ID on behavioral or hormonal responses (George et al. 2022).

Prior to the start of an assay, we first visually identified the subject female. We differentiated each female from her mate using a combination of morphological features (e.g. plumage coloration), behavior, and leg bands. After spotting the female, we placed the decoy and speaker at her nest box, then retreated to a distance of at least 30 m. We recorded observations into a digital voice recorder and later transcribed using the event-recording software JWatcher (Blumstein et al. 2006). Observed aggressive behaviors included actions without direct physical contact: approaches (flying within 1m of the decoy), dives (steep, rapid flights within 1 m of the decoy, often accompanied by vocalizations), hovers (stationary flight within 1 m of the decoy), as well as direct physical contact: hits, pecks, and perches directly on decoy. Occasionally, females left the field of view during a trial and did not return within the 5 min period, and so we tried again the following day. On average, females were present and reliably identifiable for 99.2% of each 5-min assay. To quantify aggressiveness from each assay, we calculated an aggression score as the number of 5-s intervals in which the female performed at least one aggressive behavior (Rosvall 2008). Thus, females that displayed no aggression during the 5-min trial received a score of 0, and females that performed aggressive behaviors at least once during every 5-s interval received the maximum score of 60.

Determining and confirming individual identity:

Annually since 2015, we have tagged every breeding female in this population with a unique numbered USGS aluminum band on one leg. We also marked a subset of individuals with unique color combinations of nontoxic acrylic paint on the tips of their wings to use for identification during behavioral observations (Dunn et al. 1994). Since 2017, we have also outfitted each female with a color leg band with an embedded PIT tag (Passive Integrated Transponder; 2.3 mm EM4102, IB Technology, UK), modified to fit tree swallow tarsi.

Some females are banded during Terr Est, but most were banded during Inc or C; therefore, assigning female IDs to aggression assays conducted during the parental stages was straightforward. Except for the rare case in which we knew of a death or turnover in territory ownership, we assumed any female captured at a box during one parental stage was the same female present at that box during the other parental stage.

Assigning female IDs to assays during TerrEst required multiple methods. Past banding efforts allowed us to passively detect PIT-tag leg bands using RFID readers installed at some boxes (Radio Frequency Identification, Bridge and Bonter 2011). RFID readers work by detecting every instance of a PIT tag's close proximity to an antenna at the nest box entrance and saving that bird's unique hexademical code to a text file. To use RFID for identification, we collected data at a given nest box for at least 24 h. Then, we visually inspected the resulting text files. If territory owners were banded with PIT tags, their ownership was readily apparent based on the abundance and consistency of their tag recordings over 24 h, compared to occasional reads from other individuals. We matched PIT tag codes to past banding records to differentiate male and female IDs, focusing only on females. We assigned IDs to $n = 85$ TerrEst assays based on capture or RFID detection of that female within 4 d of each trial. For $n = 10$ additional TerrEst assays, we assigned IDs based on capture or RFID within 5 to 7 d of an assay, or via documentation of some notable feature seen during the assay (e.g. auxiliary paint marking on wingtips, missing tail feather, distinctive plumage marking) and confirmed during their capture at the same box during a parental stage. In the remaining cases ($n = 6$), we did not directly identify females during TerrEst; instead, we retroactively assigned the IDs based on those obtained from captures at those boxes during parental stages. This assumes that a female remained at a given nest box after she established it as her territory—an assumption that is supported by the aforementioned firmly-identified females that remained at their boxes in all 3 breeding stages. We confirmed that our main results are qualitatively similar if we exclude these 6 females, and therefore, we retained them in our analyses.

Inferring female age

We used plumage coloration to age females in this study. Females aged two years and older have iridescent blue/green upperparts (> 90% iridescent, resembling males), whereas 1-yr-old females have distinctly brown upperparts (< 50% iridescent; Hussell 1983). Females with intermediate plumage (50% to 90% iridescent) are likely to be 1 or 2 yr olds (Hussell 1983). We estimate % iridescence during the annual census of the population while the bird is in the hand, and we noted that most focal females had blue-green plumage characteristic of 2+ yr olds (Y1 $n = 23$, Y2 $n = 62$), and the remainder (Y1 $n = 8$, Y2 $n = 8$) had brown or intermediate plumage at the time of the study.

Tree swallows have high breeding site fidelity from year to year, indicating that most unbanded birds in an otherwise monitored population are young (Winkler et al. 2004). We, therefore, capitalized on our historic banding records to obtain a more specific age (in years). Fifty-eight females were captured in a previous year (Y1 $n = 16$, Y2 $n = 42$), and 43 were captured for the first time during the focal year (Y1 $n = 15$, Y2 $n = 28$). If the female was brown in plumage upon their initial capture or banded as a chick in our study population, we knew their exact hatch year. If the subject was already iridescent blue-green upon initial capture, we could not determine the exact hatch year and instead conservatively assigned the lowest possible value. Thus, females with intermediate plumage at the time of initial banding were assumed to be 1 yr old, and females with fully blue-green plumage at the initial banding were assumed to be 2 yr old. For example, a female with blue-green plumage when banded for the first time in 2016 was assigned an age of 4 yr old for her 2018 behavioral data, whereas a female with intermediate plumage when banded for the first time in 2016 was assigned an age of 3 yr old for her 2018 behavioral data. Using this method, we could determine that out of 101 females in this behavioral dataset, 16 were 1 yr, 49 were at least 2 yr, 18 were at least 3 yr, and collectively 18 were at least 4 to 6 yr.

Measuring female quality and performance metrics

To understand the causes or consequences of plasticity type and/or stage-specific variation in aggression, we explored 6 measures of quality or performance for each focal female:

Body mass

We measured each female's body mass using a Pesola scale at the time of banding. To avoid fluctuations in body mass associated with egg production, we focused here on measurements taken opportunistically during parental stages (Y1 $n = 14$, Y2 $n = 70$; total $n = 84$). Most females measured in Y1 were captured and weighed shortly after the hatching of their chicks, while all females in Y2 were captured and weighed in the final days of incubation, and so we statistically controlled for this in our analyses (see below).

Overwinter survival

We used banding and capture records from the years immediately following Y1 and Y2 of this study (2019 and 2021, respectively) to determine whether or not females successfully migrated, overwintered, and returned to our field sites. Because females show high site fidelity as adults (Winkler et al. 2004), we use this binary "return Y/N" variable as a proxy for female overwinter survival.

Other reproductive metrics

We focused on 4 additional metrics: lay date, hatching success, average chick mass, and chick recruitment. We recorded a female's lay date, which describes when she initiated egg-laying. Females that initiate laying earlier routinely have greater reproductive success (Verhulst et al. 1995), and therefore, lay date is thought to be a proxy of female quality. Next, we quantified hatching success based on whether eggs present at the end of incubation ultimately hatched or not. Parents routinely remove dead newly hatched chicks from nests but do not readily remove unhatched eggs (*personal obs.*), so we assumed any eggs that 'disappeared' around the time of hatching did successfully hatch but died shortly thereafter. Third, we recorded chick mass on postnatal D12 and calculated average chick mass for each nest; this metric is an established proxy for quality because heavier chicks are more likely to survive to adulthood (McCarty 2001). Finally, we used banding records to determine whether each breeding attempt in this study led to successful recruitment of offspring into the population in a subsequent year. Of the 24 cases in which a female had some offspring return as adults, only one female had multiple (2) chicks recruit from the same brood. Thus, we created a binary variable ("any chicks recruited, Y/N").

We did not analyze all aspects of female reproductive success because of additional co-occurring research projects in this population. Specifically, 81 of 101 nests had an egg or chick removed for terminal sampling (elaborated in Supplementary Methods). To account for potential downstream effects of clutch or brood reduction from these other experiments, we created a binary variable ("prior artificial brood reduction, Y/N"), relative in time to each dependent variable. If an egg was collected at a nest before hatching, then "artificial brood reduction" = "Y" for statistical models of female body mass, female survival, egg hatching success, D12 chick mass, and chick recruitment. If a chick was collected from a nest on D6 post-hatch, "artificial brood reduction" = "Y" for analyses of female survival, D12 chick mass, and chick recruitment, but = "N" for analyses of hatching success and female mass, since chick collection occurred after hatching and female mass measurements. This variable did not need to be included in analyses of lay date, since any brood reduction occurred after that variable was measured.

Statistical analyses

All analyses were carried out in R (R Core Team 2019).

Testing for effects of breeding stage and individual on aggression

Visual inspection of aggression scores (Fig. 1a) revealed a large degree of variation in aggression levels within and among breeding stages. To test whether breeding stage and individual ID explained this behavioral variation, we used the package "lme4" to create a linear mixed model (LMM) with aggression scores as the dependent variable, breeding stage as a categorical fixed effect, and individual as a random effect (Bates et al. 2007). Year was also included as a categorical fixed effect. We tested for the significance of the fixed effects and interaction terms using the package "lmertest" (Kuznetsova et al. 2015). We calculated repeatability (*R*) as the intra-class coefficient, which reports the variance attributed to the random effect of female ID, out of the total phenotypic variance (Nakagawa and Schielzeth 2010); *R* is therefore a measure of the proportion of phenotypical variance that can be attributed to between-subject variation. We used the "rptR" package to estimate the uncertainty in the estimate of R and perform significance testing (Stoffel et al. 2017).

Categorizing plasticity types

Next, we characterized how females differed from one another in their stage-to-stage behavioral plasticity. Visual inspection of individual-level aggression scores (Fig. 1b) showed nonlinear patterns of change, suggesting that a linear reaction norm approach would not be appropriate. Instead, we used unsupervised clustering to group together females based on how their aggression levels changed from TerrEst to Inc (Δ agg $_{\text{TerrEst, Inc}}$) and from Inc to C (Δ agg $_{\text{Inc, C}}$; Supplementary Fig. S2). This approach was modified from genomics studies that cluster genes based on how their expression levels vary across particular timepoints, or the shape of their gene expression reaction norms (e.g. Bukhari et al. 2017). We performed k-means clustering of Δ agg $_{\text{TerrEst, Inc}}$ and Δ agg $_{\text{Inc, C}}$ values with 25 iterations, grouping individuals based on their Euclidian distance from one another. We used the package 'NbClust' (Charrad et al. 2014) to determine the optimal number of clusters. Briefly, this package uses 25 indices that evaluate intra-cluster compactness and inter-cluster separation for a range of possible cluster numbers. These tests overwhelmingly pointed to 4 clusters as optimal (Supplementary Methods). For convenience, we labeled these 4 clusters as "plasticity types." Females within each cluster exhibited quantitatively similar behavioral changes to one another across breeding stages, though there is still some variation in both Δ agg $_{\text{TerrEst, Inc}}$ and Δ agg $_{\text{Inc, C}}$

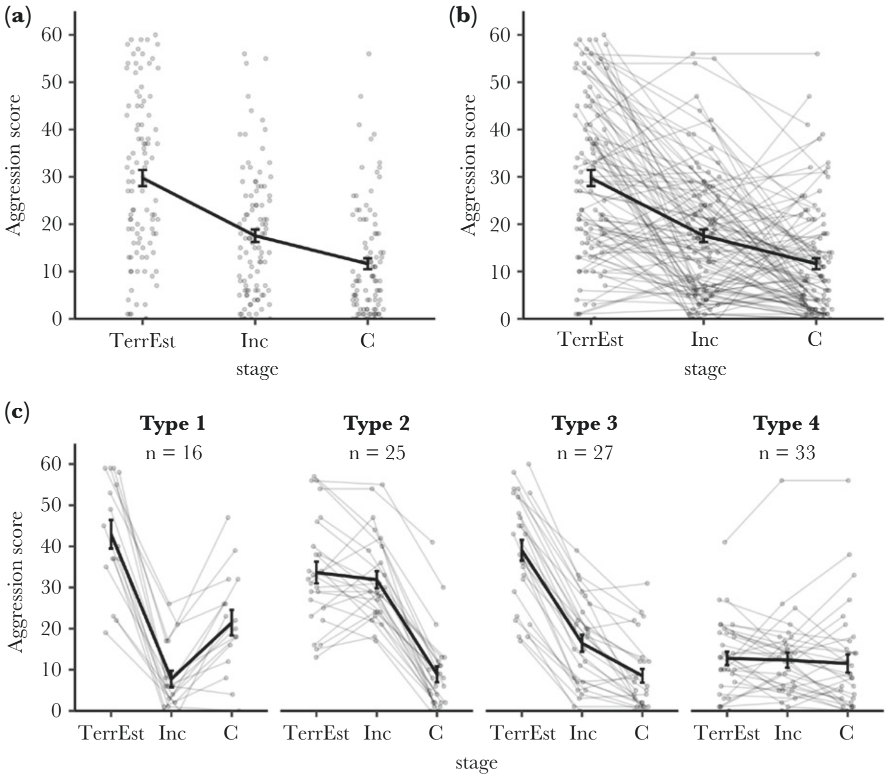

Fig. 1. (a) Female aggression differed by breeding stage at the population level, but (b) individuals varied in how their aggression changed between consecutive stages, resulting in (c) 4 main plasticity types. Each gray line is a female, and black lines summarize stage-specific averages for each panel.

within each cluster (Supplementary Fig. S2). We explored alternative analytical approaches, such as random regression (Nussey et al. 2007), which does not fit the non-linearity of these data, or models that use statistical interactions to test for joint effects of Δ agg $_{\text{TerrEst, Inc}}$ and Δ agg $_{\text{Inc, C}}$. Results of these latter models are qualitatively similar to our categorical clustering, yet more difficult to interpret and visualize, and so, we have included them in the Supplementary Materials (Supplementary Table S7).

Testing for effects of age on behavior

To test for age differences among plasticity types, we used a simple ANOVA. To test for age effects on aggression scores, we used an LMM with fixed effects of year, breeding stage, age, and stage × age interaction. Female ID was included as a random effect. We checked that appropriate model assumptions were met using the 'DHARMa'package (Hartig 2019). Post-hoc, we used the "emtrends" function in the "emmeans" package to test whether there was a significant effect of age on aggression within each breeding stage (Lenth et al. 2018), with False Discovery Rate (FDR) corrections for multiple tests (Benjamini and Hochberg 1995).

Testing for relationships between behavior and performance metrics

We separately tested relationships with TerrEst aggression, Inc aggression, C aggression, and plasticity type, thus resulting in 4 tests per variable of interest, for the 6 variables listed below. We globally adjusted aggression-associated *P*-values using FDR, as above. Based on variable type, statistical tests were either linear models (LMs) or generalized linear models (GLMs). We used the package "rsq" to obtain partial R^2 values for all fixed effects in LMs and GLMs (Zhang et al. 2018). We checked that appropriate model assumptions were met using the "gvlma" package for LMs (Pena and Slate 2014) and "DHARMa" package for GLMs (Hartig 2019). Because we were interested in the performance of specific plasticity types relative to the population, we further explored each significant or trending effect of plasticity type ($P < 0.1$) with a post-hoc test; in each case, we compared each type-specific response to the population average response, using the "contrasts" function in the "emmeans" package and with FDR corrections for multiple tests.

Each model included a fixed effect of age class (1 yr or 2+ yr old) because yearling females differ in many elements of morphology and reproductive performance (Stutchbury and Robertson 1988; Lozano and Handford 1995; Robertson and Rendell 2001). Each model also included a binary variable to account for any artificial brood reduction occurring prior to measurement of the outcome variable (described above and in Supplementary Methods). Finally, we accounted for any between-year differences by including Year as a fixed effect, except as noted below.

1. Body mass: LMs tested for relationships between female body mass and stage-specific aggression or plasticity type. We included an effect of the stage in which morphology was measured due to expected stage-related variation in mass. We did not include Year as a fixed effect due to collinearity with stage.
2. Overwinter survival: GLMs with binomial distributions tested whether aggression scores during each breeding stage or plasticity type predicted the female's overwinter survival (a binary response variable).
3. Lay date: LMs tested for relationships between lay date and stage-specific aggression or plasticity category.

4. Hatching success: GLMs with binomial distributions tested for effects of stage-specific aggression or plasticity category on the proportion of eggs successfully hatched within a clutch. Specifically, the response variable was specified as “cbind(# hatched, clutch size – # hatched),” i.e. the number of hatching “successes” and “failures” in each nest. Lay date was included as an additional fixed effect due to established relationships among lay date, weather, and hatching success (Shipley et al. 2020).
5. Chick mass: LMs tested for effects of aggression or plasticity type on average D12 mass for that female’s brood (sum D12 chick mass/ # chicks alive at D12). Hatch date and brood size on the day of sampling were also included as additional fixed effects because both affect chick growth (Murphy et al. 2000; Dawson 2008). Residuals of the models using untransformed chick body mass were not normally distributed, which we attributed to a negative skew in the distribution of the mass measurements. We therefore performed a square-root transformation on the response variable using the following formula: (square-root(max(D12 mass + 1)- D12 mass). Outputs were then multiplied by –1 to restore original directionality.
6. Chick recruitment: GLMs with binomial distributions tested whether any chicks returned the following year, a binary response variable, depended on stage-specific aggression scores or plasticity type.

Permutation analyses

To explore the robustness of our results, we performed a set of permutation tests to guard against Type 1 error. We created 10 000 permuted data sets, in which we assigned plasticity types randomly rather than according to their quantitative behavioral data. In each permuted data set, the number of females assigned to each type was held constant, corresponding to the sample sizes from our actual data ($n = 16$ Type 1, $n = 25$ Type 2, etc.). For each permuted data set, we then fit the same 6 types of statistical models as described above, with “permuted type” as a predictor variable alongside any other fixed effects previously specified. We saved the results of the significance tests for the permuted type variable and calculated the frequency at which we observed significant results from the permuted data for each variable of interest. This frequency thus represents the probability of obtaining a false positive effect of plasticity type for that response variable.

Ethical note

All applicable institutional and national guidelines for the care and use of animals were followed, including IACUC protocol at our university, federal banding permit, and state and local research and collecting permits. Aggression assays lasted only 5 min, and we observed that females resumed normal behavior within minutes of each assay’s completion.

Results

Effects of breeding stage and overall repeatability

Adjusted repeatability (R) obtained from the random effect of individual ID was significantly different than 0 ($R = 0.37 \pm 0.06$; $P < 0.0001$). Thus, individual identity explains ~37% of the variation in aggression. We found a significant effect of stage on aggression scores ($F_{2,200} = 68.73$, $P < 0.0001$), such that TerrEst > Inc > C (Fig. 1a). The effect of year was not statistically significant ($F_{1,\,99} = 2.67$, $P = 0.11$).

Plasticity types

We categorized the stage-to-stage behavioral changes of females into 4 plasticity types (Fig. 1c). Females that exhibited Type 1 changes were characterized by a steep decrease in aggression from TerrEst to Inc, followed by a steep increase in aggression from Inc to C. We consider these females to have exhibited higher plasticity than females of other types because their cumulative, absolute value of change in aggression was highest (i.e. $|\Delta\ \text{agg}_{\text{TerrEst, Inc}}| + |\Delta\ \text{agg}_{\text{Inc, C}}|$; Supplementary Fig. S3). Comparatively, females with Type 2 and Type 3 changes exhibited moderate plasticity, and both of these types became less aggressive over the course of the breeding season, generally consistent with population-level patterns. However, Type 2 females exhibited a greater decrease in aggression from Inc to C, while Type 3 females exhibited a greater decrease from TerrEst to Inc. Finally, Type 4 females exhibited very little plasticity overall, maintaining moderate to low aggression across all 3 stages.

Relationships between behavior and age

We found a significant main effect of age on female aggression ($\beta = -3.14 \pm 1.08$, $F_{1,\,198} = 3.99$, $P = 0.049$ and a trending age by breeding stage interaction ($F_{2,\,198} = 2.74$, $P = 0.067$), in addition to the highly significant effect of stage ($F_{2,98} = 27.62$, $P < 0.0001$) and a trending effect of year ($F_{1,98} = 3.25$, $P = 0.074$). Post-hoc analysis of the interaction term revealed a significant effect of age within TerrEst ($\beta = -3.14 \pm 1.08$, $t(231) = -2.90$, $P = 0.012$) but not within Inc ($\beta = -1.49 \pm 1.08$, $t(231) = -1.38$, $P = 0.25$), or C ($\beta = -0.31 \pm 1.08$, $t(231) = -0.29$, $P = 0.77$). Visually, differences in TerrEst and Inc aggression appear greatest between generally young (1–3 yr) and generally old (4+ yr) birds (Fig. 2a–c). Despite relationships between age and aggression scores, females of different plasticity types did not differ in age ($F_{3,\,100} = 0.42$, $P = 0.74$, permuted $P = 0.054$; Fig. 2d).

Relationships between behavior and other variables

Table 1 summarizes relationships between behavior and other variables of interest. Results related to additional fixed effects (year, age class, artificial brood reduction, capture stage) are detailed in the Supplementary Results (Supplementary Tables S1–S6).

Body mass

Body mass was not related to aggression at any one breeding stage (Supplementary Table S1, $F < 2.4$ and $P > 0.12$ for all tests). Plasticity type was related to body mass before FDR ($P = 0.043$; Fig. 3a), though this effect was no longer significant after adjustment for FDR ($P = 0.34$). Post-hoc tests indicated that Type 1 females had lower-than-average body mass (adjusted $P = 0.03$, Fig. 3a).

Overwinter survival

Stage-specific aggression did not predict likelihood to recapture the female the following year ($P > 0.40$ in all models, Supplementary Table S2). Plasticity type, however, exhibited a nonsignificant trend with survival, though the global effect was not significant after FDR correction ($c^2(3) = 6.98$, $P = 0.07$, adjusted $P = 0.29$; Fig. 3b). Post-hoc tests indicated that Type 1 females tended to be less likely than average to return the following year (adjusted $P = 0.10$) with only 19% of Type 1 females returning, compared to 44% to 56% for other types.

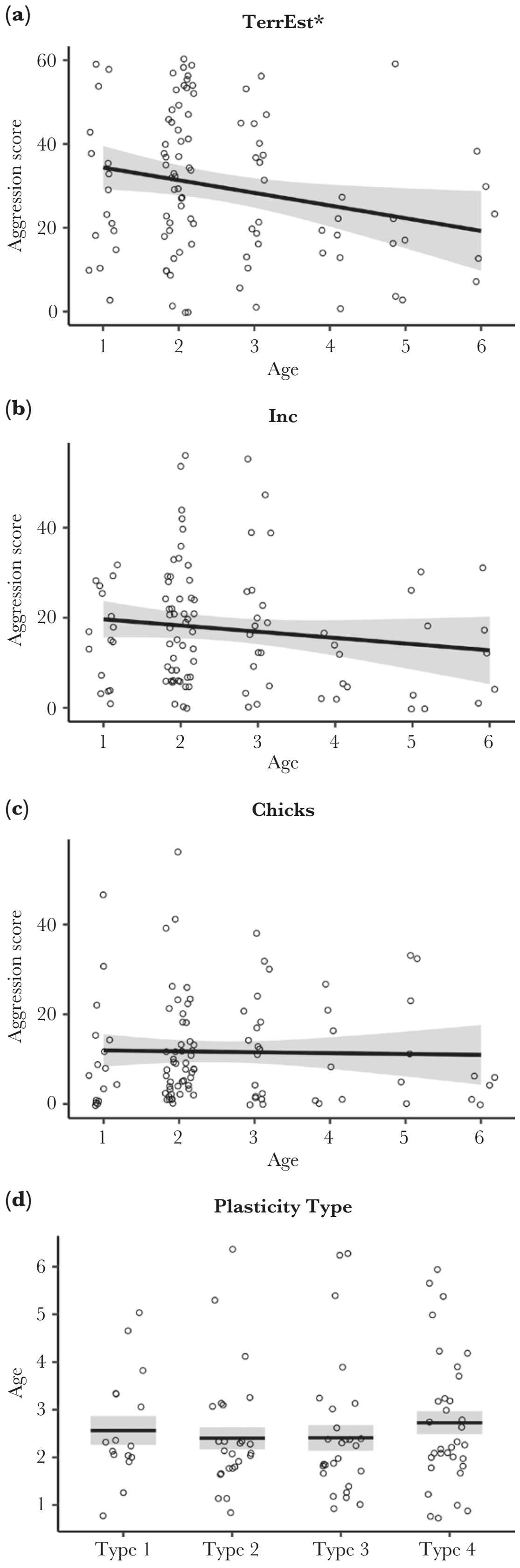

Fig. 2. Age (in years) was significantly (*) related to aggression score during TerrEst (a), but not during Inc (b) or Chick (c) stages. Age did not significantly differ among plasticity types (d). Solid lines in a–c designate trend lines, with 95% CI shaded; solid lines in d designate mean age, with standard error shaded.

Lay date

Stage-specific aggression scores were not significantly related to lay dates ($F < 0.50$ and $P > 0.48$ in all models, Supplementary Table S3). Plasticity type did not predict the start of laying either ($F_{3,95} = 0.81$, adjusted $P = 0.80$, Fig. 3c).

Hatching success

We found some trends when associating behavior with hatching success, though these patterns were not significant after FDR correction. Specifically, aggression during TerrEst was related to hatching success, with fewer eggs hatched for more aggressive compared to less aggressive females ($P = 0.017$, adjusted $P = 0.07$, Supplementary Table S4 and Table 1); aggression during other stages was unrelated to hatching success (Supplementary Table S4). Similarly, plasticity type showed some relationship to hatching success (Fig. 3d), though the main effect was not significant after FDR correction (χ^2 (3) = 7.08, $P = 0.069$, adjusted $P = 0.13$). Post-hoc tests revealed that Type 1 females tended to have lower than average hatching success (adjusted $P = 0.081$), and Type 4 females tended to have higher than average hatching success (adjusted $P = 0.081$).

Chick mass on postnatal D12

Stage-specific aggression scores were not significantly related to chick mass, except for a positive trend with Inc aggression (unadjusted $P = 0.088$; Supplementary Table S5); however, this relationship was no longer significant after FDR adjustment (Table 1). Plasticity type also did not predict average chick mass (Supplementary Table S5, Fig. 3e).

Chick recruitment

Aggression scores during TerrEst and Inc were not related to subsequent chick recruitment into the breeding population ($P > 0.24$, Supplementary Table S6). Aggression scores during chick-rearing showed some relationship with chick recruitment, such that more aggressive females were less likely to have any of their chicks return the following year (unadjusted $P = 0.08$, Supplementary Table S6); however, this result was not significant after correcting for multiple comparisons (Table 1). Finally, females of different plasticity types did not differ in the likelihood of their chicks surviving to return the following year ($c^2(3) = 1.29$, $P = 0.73$, Fig. 3f).

Discussion

We measured individual-level plasticity in territorial aggression across 3 breeding stages in which we expected the costs and benefits of aggression to differ: shortly after territory-establishment (TerrEst), during incubation (Inc), and during chick-rearing (C). At the population level, we found that aggression decreased with subsequent breeding stages, consistent with the idea that females modulate their aggression levels alongside changes in the competitive environment and other energetic demands (i.e. parental duties). However, repeatability within individuals across the entire breeding season was 0.37. While consistent with Bell et al. (2009) meta-analysis of the repeatability of behavioral traits (overall average $R = 0.37$), this value is substantially lower than the repeatability of territorial aggression measured within one breeding stage ($R = 0.75$; Rosvall 2008), suggesting that individuals vary in how their aggressiveness shifts as they progress from one breeding stage to the next. Indeed, clustering analyses identified 4 general patterns of how behavior changed (or did not change); we termed these patterns “plasticity types.” Notably, only two of

Table 1. Summary of significant or trending relationships with stage-specific aggression or plasticity type.

	TerrEst Aggression	Inc Aggression	C Aggression	Plasticity type
Body mass	P = 0.63 (P = 0.63)	P = 0.63 (P = 0.50)	P = 0.25 (P = 0.12)	Overall P = 0.17 (P = 0.04); permuted P = 0.19) * **Type 1: P = 0.03 (P = 0.0075)** Type 2: P = 0.40 (P = 0.30) † **Type 3: P = 0.10 (P = 0.05)** Type 4: P = 0.83 (P = 0.83)
Overwinter survival	P = 0.48 (P = 0.48)	P = 0.48 (P = 0.48)	P = 0.48 (P = 0.39)	Overall P = 0.29 (P = 0.072); permuted P = 0.052 † **Type 1: P = 0.10 (P = 0.026)** Type 2: P = 0.76 (P = 0.76) Type 3: P = 0.19 (P = 0.15) Type 4: P = 0.19 (P = 0.14)
Lay date	P = 0.78 (P = 0.41)	P = 0.68 (P = 0.58)	P = 0.80 (P = 0.80)	Overall P = 0.80 permuted P = 0.057 (P = 0.49)
Hatching success	† P = 0.07 (P = 0.017)	P = 0.55 (P = 0.55)	P = 0.13 (P = 0.10)	Overall P = 0.13 (P = 0.07) permuted P = 0.052 † **Type 1: P = 0.081 (P = 0.020)** Type 2: P = 0.99 (P = 0.91) Type 3: P = 0.99 (P = 0.99) † **Type 4: P = 0.081 (P = 0.040)**
Chick D12 mass	P = 0.36 (P = 0.18)	P = 0.35 (P = 0.088)	P = 0.63 (P = 0.63)	Overall P = 0.63 permuted P = 0.047 (P = 0.47)
Chick recruitment	P = 0.63 (P = 0.23)	P = 0.73 (P = 0.73)	P = 0.32 (P = 0.08)	Overall P = 0.73 permuted P = 0.069 (P = 0.73)

* and bolding designate P-values < 0.05 after adjustments, † and bolding designates P < 0.10 after adjustments, and p-values in parentheses are unadjusted (taken from Supplementary tables). Full model outputs, with all additional fixed effects are reported in Supplementary Tables S1–S6.

these types consisted of successive decreases in aggression from breeding stage to stage, mirroring the population-level averages. Another type was characterized by minimal changes in aggression across breeding stages. A final type of plasticity was instead characterized by a sharp drop in aggression from TerrEst to Inc, followed by an increase during chick rearing. Females displaying this form of behavioral plasticity also exhibited poorer performance, but females exhibiting other forms of plasticity, including low stage-to-stage plasticity, did not differ from one another in performance. Thus, individuals exhibiting different types of behavioral plasticity (or lack thereof) within a breeding season can experience similar fitness outcomes. Though correlative in nature, these results shed light on the diverse behavioral strategies within a population and set the stage for future research on their ecological drivers and evolutionary consequences.

Using 4 plasticity types to describe patterns of behavioral changes

Three of the 4 identified behavioral types consisted of marked changes in aggression across a breeding season. Type 1 included the greatest cumulative changes in aggression over the entire season: females exhibiting this pattern started with some of the highest levels of aggression seen in the population, then decreased steeply to have some of the lowest responses during incubation, only to increase in aggressiveness during chick-rearing. In this way, their aggression can be seen as hyper-plastic, relative to the other females. Types 2 and 3 both followed the population-level pattern of decreasing aggression over the course of the entire breeding season (sensu Bentz et al. 2019a), differing only in which stage-to-stage transition had the greater decrease, TerrEst to Inc (Type 3), or Inc to C (Type 2).

The behavioral profiles captured by Type 4 differed from that of the population average and from our predictions. This type involved relatively constant levels of low to moderate aggression across all 3 breeding stages investigated. In this way, females of Type 4 exhibited low plasticity compared to other females. Comparable patterns, in which plasticity and trait expression co-vary, have been reported for other behaviors, such as anti-predator defense and provisioning rates (Betini and Norris 2012; Westneat et al. 2015), as well as morphological features, such as body size in tadpoles and shell thickness in freshwater snails (Auld et al. 2010). In each of these cases, the slopes of reaction norms (i.e. plasticity) and intercepts of reaction norms (i.e. average trait values) are positively correlated, such that individuals with the most extreme or exaggerated traits in at least one context are also the most plastic (but see He et al. 2017; Jolles et al. 2019; Marchetti and Drent 2000 for examples of negative correlations between traits and plasticity). In the case of female tree swallows, the complete absence of a “high aggression, low plasticity” phenotype is notable and suggests that this strategy has been selected against, has not been selected for, or does not exist due to other constraints. These constraints also could be physiological or ecological (Ketterson et al. 2005; Beever et al. 2017), which could result in the same females expressing different patterns of among-stage plasticity across different years. We believe the further study of these potential drivers of plasticity within and among individuals is an important future direction.

Comparing female characteristics and performance among types

To further explore the potential costs or benefits of different behavioral types, we compared female performance and found that

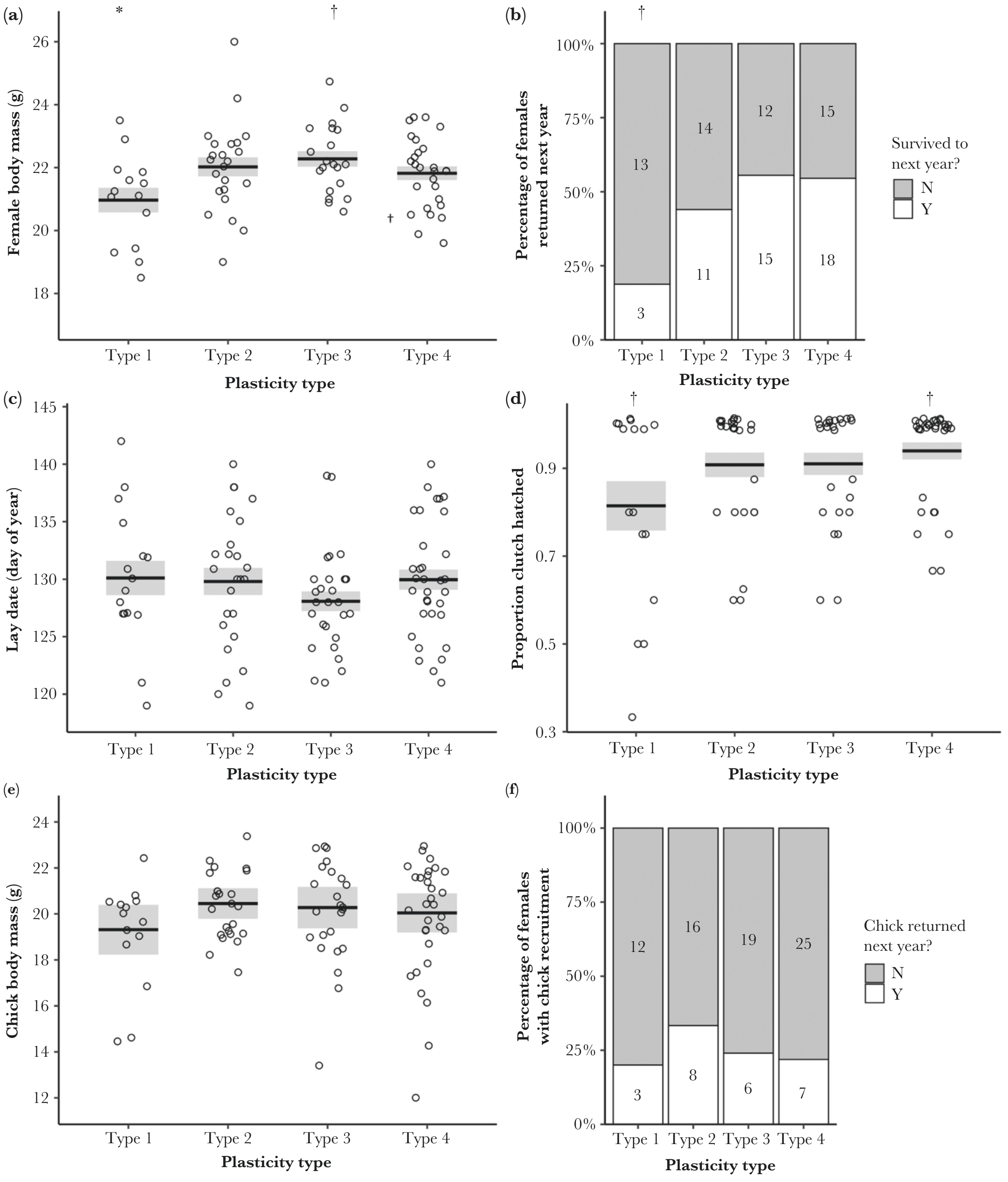

Fig. 3. Relationship between plasticity type and key performance metrics: (a) female body mass, (b) overwinter survival, (c) lay date, (d) hatching success, (e) chick body mass, and (f) chick recruitment. Continuous response variables (a, c, e) are plotted as partial-residual adjusted values, accounting for other fixed effects in the models (Supplementary Results). Data in (e) were back-transformed. Horizontal lines and shading in plots a, c, d, e represent means and SE. * designates a plasticity type that significantly differed from the average response in post-hoc analyses (FDR-adjusted $P < 0.05$). † designates a plasticity type with a trending effect (FDR-adjusted $P < 0.10$).

females exhibiting Type 1 plasticity tend to differ from other females in the population in key ways. Specifically, we saw that fewer Type 1 females tended to return to breed the following year, suggesting that they may experience greater mortality in the non-breeding season. We also found that females exhibiting Type 1 plasticity tended to have lower-than-average hatching success that year, while females with Type 4 plasticity tended to have higher-than average hatching success. Females with Type 1 plasticity also weighed less than average, while Types 3 females tended to weigh more than average. Our permutation analysis flagged this mass result as potentially spurious, though the direction of the effect is consistent with the previously described patterns involving overwinter survival and hatching success. Notably, we did not observe any direct relationships between stage-specific aggression and performance. Therefore, any negative performance metrics associated with Type 1 plasticity coincide with a particular *combination* of aggression levels across breeding stages (i.e. plasticity types), rather than aggression levels in any given breeding stage.

There are at least two possible explanations for why females that exhibited Type 1 plasticity also experienced poorer fitness outcomes. First, especially high aggression during territorial establishment could have caused females to be in poor condition as they entered incubation. Afterall, there is substantial evidence that high levels of aggression have energetic, metabolic, and other costs (Marler and Moore 1991; Rosvall 2011c; Georgiev et al. 2015). If such costs accrue quickly or leave a lingering mark (Bentz et al. 2021), these females may not be able to maintain high or moderate aggression through incubation, at least until they recuperate and recover their high aggressiveness during the chick period. Second, the process or machinery of changing behavioral phenotypes may itself be costly (DeWitt et al. 1998; Murren et al. 2015), such that oscillating aggression levels to such an extent (for whatever reasons) caused females to be in worse condition (lower body mass), and thus less likely to survive the following months. Seasonal changes in aggressive behavior are known to be mediated by seasonal changes in neural tissues (Goodson et al. 2005). However, the energetics of seasonal neuroplasticity are still poorly understood, especially in females (Bentz et al. 2019b; Rose et al. 2022). Because of the correlative nature of this study, we cannot determine whether this behavioral pattern leads to poor condition, or whether a female's condition affects her behavioral changes; nevertheless, it is clear that females exhibiting this behavioral phenotype appear worse off than other breeding females.

On the other hand, Type 1 females still have a fitness advantage compared to another group of females: floaters who do not have a nesting territory and were therefore excluded from the study here. Females of this 'zero-class' are known to be abundant in this system (Stutchbury and Robertson 1987a), and they skew younger in age (Stutchbury and Robertson 1985). In failing to obtain or retain a nesting site for a reproductive cycle, floaters have *zero* reproductive success in that breeding season. Past experimental work has shown that females with higher aggression during territorial establishment are more likely to win contests over nesting sites (Rosvall 2008), indicating a very real reproductive advantage to high aggressiveness, like that seen during TerrEst in Type 1 females. Our analyses do not indicate that Type 1 females are younger than any other type of female, though we note that the highest aggression scores were seen in younger birds (elaborated below). Perhaps Type 1 females exhibit especially high aggression during TerrEst in order to increase their probability of having at least some reproductive success in a season, even though this is correlated with some costs relative to other breeding females. The benefit of obtaining a cavity may explain why this plasticity type was still present in our population, particularly if females vary in how they resolve life history trade-offs related to current reproduction versus future survival.

Of course, this argument makes the prevalence of females exhibiting little behavioral change across all stages (i.e. Type 4 plasticity) all the more puzzling, as the vast majority of these low-plasticity females exhibited moderate to low aggression levels during all 3 breeding stages. We can think of a few potential explanations as to how these low-aggression, low-plasticity individuals retained a nesting site and went on to have comparable reproductive success to the more aggressive or moderately more plastic females. For one, escalated or direct aggressive interactions may not be necessary to prevail during competitive interactions if animals can effectively communicate their aggressive intent or resource-holding potential (Searcy and Nowicki 2010). There is evidence that female tree swallows with brighter white breast plumage exhibit a more "passive" response to both simulated (Beck and Hopkins 2019) and actual territorial intrusions (Taff et al. 2019). Some females may also be able to claim a nest site without as much competition, based on the timing of their arrival at breeding sites or their prior experience (Thys et al. 2021). Additional longitudinal work is necessary to determine whether a correlation between aggression and age is driven by changes in aggression across females' lifetimes, differential survival among females, or both. However, our finding that the oldest, most experienced females that had been breeding at our study site for several years were significantly less aggressive than younger females during TerrEst is consistent with the idea that older females may compete in different ways than younger females.

Another explanation for the persistence of these low-aggression, low-plasticity females relates to the nature of our aggression assays. Specifically, a female who scores relatively low via our aggression assay could still be aggressive enough to deter most real intruders. These relatively low aggression females rarely exhibited *no* aggression response to the short assay. Indeed, low-plasticity females were aggressive for 20% of their 5-min trial (average score = 12.8 ± 1.6 five-second bins, out of 60). While our aggression assays used stationary decoys whose "intrusion" lasted a full 5 min, most real intruders retreat quickly upon being chased (Stutchbury and Robertson 1987a). Finally, we must again consider the 'zero-class' floater females not captured in our current dataset, whose aggressive phenotype we could not measure but whose reproductive success was zero. Given that more aggressive females are more likely to obtain a nest box during competition (Rosvall 2008), we might predict that floaters are less aggressive than *all* of the subjects of this study, even those with "low" aggression relative to the other breeding females.

Conclusions

In summary, we found that—at the population level—aggression decreased across successive breeding stages. However, individuals do not necessarily match the population level pattern. High levels of plasticity in aggression were associated with worse qualities and performance, compared to other patterns of behavioral plasticity. Critically, we did not find significant relationships between these same performance metrics and aggression measured in just one breeding stage. This suggests that the interactive effect of behavioral variation across stages may be more salient than the aggressiveness at any one stage. We also found many females displayed very little plasticity in aggression across the breeding season, despite stage-to-stage differences in the demands of competition and care. Thus, in at least some cases, different behavioral strategies are associated with equivalent fitness outcomes, at least within a single reproductive cycle. The existence of these diverse plasticity patterns also challenges assumptions that there is one "best" route to success; after all, many individuals do not adjust their behavior in the way we expected. These behavioral strategies may vary over an individual's lifetime and in relation to year-to-year environmental factors as well (sensu Araya-Ajoy and Dingemanse 2017), further contributing to the maintenance of behavioral variation within a population.

Animal behavior has a strong foundation in both consistent individual differences (Dall et al. 2004; Sih et al. 2004) and in behavioral flexibility (Stamps 2016). The integration of these topics has grown in recent years (Mathot and Dingemanse 2014; Mitchell and Houslay 2020; O'Dea et al. 2022). As animal behavior continues to investigate both proximate mechanisms underlying individual differences in behavioral plasticity (Hau and Goymann 2015; Bengston et al. 2018; Biro et al. 2018) as well as their evolutionary consequences (Mathot et al. 2012; Dingemanse and Wolf 2013; Snell-Rood 2013), we are well posed to integrate these ideas across multiple scales, including within-individual variation within a breeding cycle and among breeding years, to among-individual variation in these individual reaction norms. Past work in other areas of biology suggests that the mechanisms operating at one of these timescales may not apply to other timescales (George and Rosvall 2022). All of these points emphasize the need for repeated measures and longitudinal research studies, which will collectively test whether and why some individuals behave like the "average" population and others do not.

Supplementary material

Supplementary material is available at *Behavioral Ecology* online.

Acknowledgments

Many thanks to Indiana University's Research and Teaching Preserve, Indiana Department of Natural Resources, and Bloomington Parks and Recreation for access to field sites. We are also grateful to many people who made this research possible: especially SE Wolf, AB Bentz, MJ Woodruff, SE Lipshutz, SD Myers, KR Content, EK Dossey Curole, KR Stansberry, TA Empson, and the entire 2018 and 2020 TRES field crews. AMW also thanks GT Smith and GE Demas, for valuable feedback at earlier stages of this project. Thanks also to the editor and an anonymous referee for helpful comments on an earlier version of this manuscript.

Funding

This work was supported by National Science Foundation grants to KAR (IOS-1656109, IOS- 1942192) and EMG (NSF graduate research fellowship). EMG and KAR were all also supported by the National Institute of Health (T32HD049336). AMW was supported by Indiana University's Hutton Honors College, Cox Research Scholarship Program, and the Department of Biology Best Undergraduate Thesis Award.

Conflicts of Interest

The authors declare no conflict of interest.

Data availability

Analyses reported in this article can be reproduced using the data provided by George et al. (2024).

References

Alonzo SH. 2015. Integrating the how and why of within-individual and among-individual variation and plasticity in behavior. Curr Opin Behav Sci. 6:69–75. https://doi.org/10.1016/j.cobeha.2015.09.008

Araya-Ajoy YG, Dingemanse NJ. 2017. Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. J Anim Ecol. 86(2):227–238. https://doi.org/10.1111/1365-2656.12621

Archer J. 1988. The behavioural biology of aggression. Cambridge, UK: CUP Archive.

Auld JR, Agrawal AA, Relyea RA. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proc Biol Sci. 277(1681):503–511. https://doi.org/10.1098/rspb.2009.1355

Bates D, Sarkar D, Bates MD, Matrix L. 2007. The lme4 package. R package version, 2(1), 74.

Beck ML, Hopkins WA. 2019. The relationship between plumage coloration and aggression in female tree swallows. J Avian Biol. 50(11):1–11. https://doi.org/10.1111/jav.01910

Beever EA, Hall LE, Varner J, Loosen AE, Dunham JB, Gahl MK, Smith FA, Lawler JJ. 2017. Behavioral flexibility as a mechanism for coping with climate change. Front Ecol Environ. 15(6):299–308. https://doi.org/10.1002/fee.1502

Bell AM, Hankison SJ, LaskowskiKL. The repeatability of behaviour: a meta-analysis. Animal Behaviour 2009:77(4):771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Bengston SE, Dahan RA, Donaldson Z, Phelps SM, van Oers K, Sih A, Bell AM. 2018. Genomic tools for behavioural ecologists to understand repeatable individual differences in behaviour. Nat Ecol Evol. 2(6):944–955. https://doi.org/10.1038/s41559-017-0411-4

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bentz AB, George EM, Wolf SE, Rusch DB, Podicheti R, Buechlein A, Nephew KP, Rosvall KA. 2021. Experimental competition induces immediate and lasting effects on the neurogenome in free-living female birds. Proc Natl Acad Sci USA. 118(13):e2016154118. https://doi.org/10.1073/pnas.2016154118

Bentz AB, Philippi KJ, Rosvall KA. 2019a. Evaluating seasonal patterns of female aggression: Case study in a cavity-nesting bird with intense female–female competition. Ethology 125(8):555–564. https://doi.org/10.1111/eth.12881

Bentz AB, Rusch DB, Buechlein A, Rosvall KA. 2019b. The neurogenomic transition from territory establishment to parenting in a territorial female songbird. BMC Genomics. 20(1):819. https://doi.org/10.1186/s12864-019-6202-3

Betini GS, Norris DR. 2012. The relationship between personality and plasticity in tree swallow aggression and the consequences for reproductive success. Anim Behav. 83(1):137–143. https://doi.org/10.1016/j.anbehav.2011.10.018

Bhat A, Greulich MM, Martins EP. 2015. Behavioral plasticity in response to environmental manipulation among Zebrafish (Danio rerio) populations. PLoS One. 10(4):e0125097. https://doi.org/10.1371/journal.pone.0125097

Biro PA, Garland T, Beckmann C, Ujvari B, Thomas F, Post JR. 2018. Metabolic scope as a proximate constraint on individual behavioral variation: Effects on personality, plasticity, and predictability. Am Nat. 192(2):142–154. https://doi.org/10.1086/697963

Blumstein D, Evans C, Daniel J. 2006. JWatcher 1.0.

Bowman RM, Schmidt S, Weeks C, Clark H, Brown C, Latta LC, Edgehouse M. 2018. Phenotypic plasticity in a population of odonates. Sci Rep. 8(1):8442. https://doi.org/10.1038/s41598-018-26301-y

Bridge ES, Bonter DN. 2011. A low-cost radio frequency identification device for ornithological research. J Field Ornithol. 82(1):52–59. https://doi.org/10.1111/j.1557-9263.2010.00307.x

Bukhari SA, Saul MC, Seward CH, Zhang H, Bensky M, James N, Zhao SD, Chandrasekaran S, Stubbs L, Bell AM. 2017. Temporal dynamics of neurogenomic plasticity in response to social interactions in male threespined sticklebacks. PLoS Genet. 13(7):e1006840. https://doi.org/10.1371/journal.pgen.1006840

Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Soft. 61(6):1–36. https://doi.org/10.18637/jss.v061.i06.

Coady CD, Dawson RD. 2013. Subadult plumage color of female tree swallows (Tachycineta bicolor) reduces conspecific aggression during the breeding season. Wilson J Ornithol. 125(2):348–357. https://doi.org/10.1676/12-155.1

Dall SR, Houston AI, McNamara JM. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol Lett. 7(8):734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x.

Dawson RD. 2008. Timing of breeding and environmental factors as determinants of reproductive performance of tree swallows. Can J Zool. 86(8):843–850. https://doi.org/10.1139/z08-065

DeWitt TJ, Sih A, Wilson DS. 1998. Costs and limits of phenotypic plasticity. Trends Ecol Evol. 13(2):77–81. https://doi.org/10.1016/s0169-5347(97)01274-3

Dingemanse NJ, Wolf M. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. Anim Behav. 85(5):1031–1039. https://doi.org/10.1016/j.anbehav.2012.12.032

Dunn PO, Whittingham LA, Lifjeld JT, Robertson RJ, Boag PT. 1994. Effects of breeding density, synchrony, and experience on extrapair paternity in tree swallows. Behav Ecol. 5(2):123–129. https://doi.org/10.1093/beheco/5.2.123

Enquist M, Leimar O. 1987. Evolution of fighting behaviour: The effect of variation in resource value. J Theor Biol. 127(2):187–205. https://doi.org/10.1016/s0022-5193(87)80130-3

Fuxjager MJ, Zhao X, Rieger NS, Marler CA. 2017. Why animals fight: Uncovering the function and mechanisms of territorial aggression, APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior, Vol. 1. Washington, DC, USA: American Psychological Association. p. 853–875.

George EM, Rosvall KA. 2022. Bidirectional relationships between testosterone and aggression: A critical analysis of four predictions. Integr Comp Biol. 62(3):474–486. https://doi.org/10.1093/icb/icac100

George EM, Weber AM, Rosvall KA. 2024. Data from: The scope and adaptive value of modulating aggression across breeding stages: Case study in a competitive female songbird. Dryad. https://doi.org/10.5061/dryad.3xsj3txkn

George EM, Wolf SE, Bentz AB, Rosvall KA. 2022. Testing hormonal responses to real and simulated social challenges in a competitive female bird. Behav Ecol. 33(1):233–244. https://doi.org/10.1093/beheco/arab129

Georgiev AV, Thompson ME, Mandalaywala TM, Maestripieri D. 2015. Oxidative stress as an indicator of the costs of reproduction among free-ranging rhesus macaques. J Exp Biol. 218(Pt 13):1981–1985. https://doi.org/10.1242/jeb.121947

Gómez JM, Verdú M, González-Megías A, Méndez M. 2016. The phylogenetic roots of human lethal violence. Nature. 538(7624):233–237. https://doi.org/10.1038/nature19758

Goodson JL, Saldanha CJ, Hahn TP, Soma KK. 2005. Recent advances in behavioral neuroendocrinology: Insights from studies on birds. Horm Behav. 48(4):461–473. https://doi.org/10.1016/j.yhbeh.2005.04.005

Hartig F. 2019. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.1. 0. See http://florianhartig.github.io/DHARMa

Hau M, Goymann W. 2015. Endocrine mechanisms, behavioral phenotypes and plasticity: known relationships and open questions. Front Zool. 12(Suppl 1):S7. https://doi.org/10.1186/1742-9994-12-S1-S7

He R, Pagani-Núñez E, Chevallier C, Barnett CRA. 2017. To be so bold: boldness is repeatable and related to within individual behavioural variability in North Island robins. Behav Process. 140:144–149. https://doi.org/10.1016/j.beproc.2017.04.014

Holroyd G. 1975. Nest site availability as a factor limiting population size of swallows.

Hussell DJT. 1983. Age and plumage color in female tree swallows. J Field Ornithol. 54(3):312–318. https://www.jstor.org/stable/4512839.

Jaeger RG. 1981. Dear enemy recognition and the costs of aggression between salamanders. Am Nat. 117(6):962–974. https://doi.org/10.1086/283780

Jolles JW, Briggs HD, Araya-Ajoy YG, Boogert NJ. 2019. Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. Anim Behav. 154:193–202. https://doi.org/10.1016/j.anbehav.2019.06.022

Ketterson ED, Nolan Jr V, Sandell M. 2005. Testosterone in females: Mediator of adaptive traits, constraint on sexual dimorphism, or both? Am Nat. 166(S4):S85–S98. https://doi.org/10.1086/444602

Kuznetsova A, Brockhoff PB, Christensen RHB. 2015. Package “lmertest.” R package version, 2(0).

Laskowski KL, Moiron M, Niemelä PT. 2021. Integrating behavior in life-history theory: Allocation versus acquisition? Trends Ecol Evol. 36(2):132–138. https://doi.org/10.1016/j.tree.2020.10.017

Lee SJ, Witter MS, Cuthill IC, Goldsmith AR. 1996. Reduction in escape performance as a cost of reproduction in gravid starlings, *Sturnus vulgaris*. Proc Biol Sci. 263(1370):619–623. https://doi.org/10.1098/rspb.1996.0093

Leffelaar D, Robertson RJ. 1985. Nest usurpation and female competition for breeding opportunities by tree swallows. Wilson Bull. 97(2):221–224. https://www.jstor.org/stable/4162075.

Lenth R, Singmann H, Love J, Buerkner P, Herve M. 2018. Emmeans: Estimated marginal means, aka least-squares means. R package version, 1(1), 3.

Lozano GA, Handford PT. 1995. A test of an assumption of delayed plumage maturation hypotheses using female tree swallows. The Wilson Bulletin 107(1):153–164. https://www.jstor.org/stable/4163520.

Marchetti C, Drent PJ. 2000. Individual differences in the use of social information in foraging by captive great tits. Anim Behav. 60(1):131–140. https://doi.org/10.1006/anbe.2000.1443

Marler CA, Moore MC. 1991. Supplementary feeding compensates for testosterone-induced costs of aggression in male mountain spiny lizards, *Sceloporus jarrovi*. Anim Behav. 42(2):209–219. https://doi.org/10.1016/s0003-3472(05)80552-4

Mathot KJ, Dingemanse NJ. 2014. Plasticity and personality, integrative organismal biology. Hoboken, NJ, USA: John Wiley & Sons, 55–69.

Mathot KJ, Wright J, Kempenaers B, Dingemanse NJ. 2012. Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. Oikos 121(7):1009–1020. https://doi.org/10.1111/j.1600-0706.2012.20339.x

McCarty JP. 2001. Variation in growth of nestling tree swallows across multiple temporal and spatial scales. Auk. 118(1):176–190. https://doi.org/10.1642/0004-8038(2001)118[0176:vigont]2.0.co;2

McGlothlin JW, Jawor JM, Ketterson ED. 2007. Natural variation in a testosterone-mediated trade-off between mating effort and parental effort. Am Nat. 170(6):864–875. https://doi.org/10.1086/522838

Mitchell DJ, Houslay TM. 2020. Context-dependent trait covariances: how plasticity shapes behavioral syndromes. Behav Ecol. 32(1):25–29. https://doi.org/10.1093/beheco/araa115

Muldal A, Gibbs HL, Robertson RJ. 1985. Preferred nest spacing of an obligate cavity-nesting bird, the tree swallow. Condor. 87(3):356–363. https://doi.org/10.2307/1367216

Murphy MT, Armbrecht B, Vlamis E, Pierce A. 2000. Is reproduction by tree swallows cost free? Auk. 117(4):902–912. https://doi.org/10.1093/auk/117.4.902

Murren CJ, Auld JR, Callahan H, Ghalambor CK, Handelsman CA, Heskel MA, Kingsolver JG, Maclean HJ, Masel J, Maughan H, *et al*. 2015. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. Heredity. 115(4):293–301. https://doi.org/10.1038/hdy.2015.8

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol Rev Camb Philos Soc. 85(4):935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Nussey DH, Wilson AJ, Brommer JE. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. J Evol Biol. 20(3):831–844. https://doi.org/10.1111/j.1420-9101.2007.01300.x

O'Dea RE, Noble DWA, Nakagawa S. 2022. Unifying individual differences in personality, predictability and plasticity: A practical guide. Methods Ecol Evol. 13(2):278–293. https://doi.org/10.1111/2041-210x.13755

Pena EA, Slate EH. 2014. gvlma: Global validation of linear models assumptions. R package version, 1(0.2).

R Core Team. 2019. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Robertson RJ, Rendell WB. 2001. A long-term study of reproductive performance in tree swallows: the influence of age and senescence on output. J Anim Ecol. 70(6):1014–1031. https://doi.org/10.1046/j.0021-8790.2001.00555.x

Rose EM, Haakenson CM, Ball GF. 2022. Sex differences in seasonal brain plasticity and the neuroendocrine regulation of vocal behavior in songbirds. Horm Behav. 142:105160. https://doi.org/10.1016/j.yhbeh.2022.105160

Rosvall KA. 2008. Sexual selection on aggressiveness in females: evidence from an experimental test with tree swallows. Anim Behav. 75(5):1603–1610. https://doi.org/10.1016/j.anbehav.2007.09.038

Rosvall KA. 2011a. Cost of female intrasexual aggression in terms of offspring quality: A cross-fostering study. Ethology 117(4):332–344. https://doi.org/10.1111/j.1439-0310.2011.01881.x

Rosvall KA. 2011b. Intrasexual competition in females: evidence for sexual selection? Behav Ecol. 22(6):1131–1140. https://doi.org/10.1093/beheco/arr106

Rosvall KA. 2011c. Maintenance of variation in sexually selected traits in females: a case study using intrasexual aggression in tree swallows Tachycineta bicolor. J Avian Biol. 42(5):454–462. https://doi.org/10.1111/j.1600-048x.2011.05322.x

Rosvall KA. 2013. Life history trade-offs and behavioral sensitivity to testosterone: an experimental test when female aggression and maternal care co-occur. PLoS One. 8(1):e54120. https://doi.org/10.1371/journal.pone.0054120

Searcy WA, Nowicki S. 2010. The evolution of animal communication. Princeton, NJ, USA: Princeton University Press.

Shipley JR, Twining CW, Taff CC, Vitousek MN, Flack A, Winkler DW. 2020. Birds advancing lay dates with warming springs face greater risk of chick mortality. Proc Natl Acad Sci USA. 117(41):25590–25594. https://doi.org/10.1073/pnas.2009864117

Sih A, Bell A, Johnson JC. 2004. Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol. 19(7):372–378. https://doi.org/10.1016/j.tree.2004.04.009

Snell-Rood EC. 2013. An overview of the evolutionary causes and consequences of behavioural plasticity. Anim Behav. 85(5):1004–1011. https://doi.org/10.1016/j.anbehav.2012.12.031

Stamps JA. 2016. Individual differences in behavioural plasticities. Biol Rev Camb Philos Soc. 91(2):534–567. https://doi.org/10.1111/brv.12186

Stevenson PA, Rillich J. 2013. Isolation associated aggression – a consequence of recovery from defeat in a territorial animal. PLoS One. 8(9):e74965. https://doi.org/10.1371/journal.pone.0074965

Stoffel MA, Nakagawa S, Schielzeth H. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol. 8(11):1639–1644. https://doi.org/10.1111/2041-210x.12797

Stutchbury BJ, Robertson RJ. 1985. Floating populations of female tree swallows. Auk. 102(3):651–654. https://doi.org/10.1093/auk/102.3.651

Stutchbury BJ, Robertson RJ. 1987a. Behavioral tactics of subadult female floaters in the tree swallow. Behav Ecol Sociobiol. 20(6):413–419. https://doi.org/10.1007/bf00302984

Stutchbury BJ, Robertson RJ. 1987b. Do nest building and first egg dates reflect settlement patterns of females? Condor. 89(3):587–593. https://doi.org/10.2307/1368647

Stutchbury BJ, Robertson RJ. 1987c. Signaling subordinate and female status: Two hypotheses for the adaptive significance of subadult plumage in female tree swallows. Auk. 104(4):717–723. https://doi.org/10.1093/auk/104.4.717

Stutchbury BJ, Robertson RJ. 1988. Within-season and age-related patterns of reproductive performance in female tree swallows (*Tachycineta bicolor*). Can J Zool. 66(4):827–834. https://doi.org/10.1139/z88-122

Taff CC, Zimmer C, Vitousek MN. 2019. Achromatic plumage brightness predicts stress resilience and social interactions in tree swallows (*Tachycineta bicolor*). Behav Ecol. 30(3):733–745. https://doi.org/10.1093/beheco/arz010

Thys B, Eens M, Pinxten R, Iserbyt A. 2021. Pathways linking female personality with reproductive success are trait- and year-specific. Behav Ecol. 32(1):114–123. https://doi.org/10.1093/beheco/araa110

van Duyse E, Pinxten R, Darras V, Arckens L, Eens M. 2004. Opposite changes in plasma testosterone and corticosterone levels following a simulated territorial challenge in male great tits. Behaviour 141(4):451–467. https://doi.org/10.1163/156853904323066739

Verhulst S, van Balen JH, Tinbergen JM. 1995. Seasonal decline in reproductive success of the great tit: Variation in time or quality? Ecology. 76(8):2392–2403. https://doi.org/10.2307/2265815

Westneat DF, Wright J, Dingemanse NJ. 2015. The biology hidden inside residual within-individual phenotypic variation. Biol Rev Camb Philos Soc. 90(3):729–743. https://doi.org/10.1111/brv.12131

Wingfield JC. 1985. Short-term changes in plasma levels of hormones during establishment and defense of a breeding territory in male song sparrows, *Melospiza melodia*. Horm Behav. 19(2):174–187. https://doi.org/10.1016/0018-506x(85)90017-0

Wingfield JC, Lynn S, Soma KK. 2001. Avoiding the "Costs" of testosterone: Ecological bases of hormone-behavior interactions. Brain Behav Evol. 57(5):239–251. https://doi.org/10.1159/000047243

Winkler DW, Hallinger KK, Ardia DR, Robertson RJ, Stutchbury BJ, Cohen RR. 2020. Tree swallow (*Tachycineta bicolor*), version 1.0, In: Poole AF, editor. Birds of the World. Ithaca, NY: Cornell Lab of Ornithology. https://doi-org.proxyiub.uits.iu.edu/10.2173/bow.treswa.2101

Winkler DW, Wrege PH, Allen PE, Kast TL, Senesac P, Wasson MF, Llambias PE, Ferretti V, Sullivan PJ. 2004. Breeding dispersal and philopatry in the tree swallow. Condor. 106(4):768–776. https://doi.org/10.1650/7634

Yang LH, Bastow JL, Spence KO, Wright AN. 2008. What can we learn from resource pulses. Ecology. 89(3):621–634. https://doi.org/10.1890/07-0175.1

Zhang D, Zhang MD, generalized R-squared, C. 2018. Package “rsq.”