No evidence for behavioural syndrome and genetic basis for three personality traits in a wild bird population

Jennifer Morinay, Grégory Daniel, Lars Gustafsson, Blandine Doligez

No evidence for behavioural syndrome and genetic basis for three personality traits in a wild bird population

Jennifer Morinay[1, 2,*], **Gregory Daniel**[1], **Lars Gustafsson**[2], **Blandine Doligez**[1]

[1] Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, Villeurbanne, France

[2] Department of Ecology and Evolution/Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Received 3 September 2018

Initial acceptance 31 October 2018

Final acceptance 25 March 2019

MS number 18-00633R

* *Correspondence:* J. Morinay, LBBE, CNRS UMR 5558, Univ Lyon–Université Claude Bernard Lyon 1, Bâtiment Gregor Mendel, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France.

E-mail address: jennifer.morinay@gmail.com

Keywords: aggressiveness, between- and within-individual correlations, boldness, collared flycatcher, *Ficedula albicollis*, heritability, neophobia, parental effects, pedigree, quantitative genetics, repeatability

ABSTRACT

Personality traits and their correlations have been shown to be linked with life history strategies and fitness in various species. Between-individual correlations (i.e. behavioural syndromes) between personality traits can affect the evolutionary responses of these traits to environmental variation. Understanding the genetic and ecological determinants of personality traits and their interactions as behavioural syndromes in the wild is thus needed to shed light on the mechanisms shaping their evolution. Partitioning the observed (co)variance in these traits, however, requires large numbers of repeated behavioural measures on many individuals of known relatedness level. In the absence of such data, it is thus often assumed that phenotypic (co)variances inform about (1) underlying between-individual (co)variances (i.e. ignoring within-individual (co)variances) and (2) underlying genetic (co)variances. We tested these assumptions using three personality traits collected during 3 years on a long-term monitored breeding population of collared flycatchers, *Ficedula albicollis*. We partitioned the observed phenotypic (co)variance of aggressiveness, boldness and neophobia into genetic, permanent environment and parental components, and we estimated the repeatability and heritability of these traits and their between-individual correlations. All three traits were repeatable between years (at least on the latent scale) but none were heritable. Permanent environment effects explained 15% of the phenotypic variance in aggressiveness, and parental effects explained 25% of the phenotypic variance in neophobia, in line with previous studies in wild populations. The three traits showed phenotypic correlations but no between-individual correlations and no additive genetic covariance. Thus, our results did not support the assumptions that phenotypic covariance reflects behavioural syndromes and genetic covariance. We discuss the reasons for the absence of heritability and between-individual and

genetic covariance between these three personality traits in light of the possible selective pressures acting on this population.

INTRODUCTION

Over the past two decades, personality traits, that is, repeatable between-individual behavioural differences across time and contexts (Réale, Reader, Sol, McDougall, & Dingemanse, 2007), have received increasing attention in animal behavioural and evolutionary ecology studies (Bell, 2007; Bell, Hankison, & Laskowski, 2009; Carere & Maestripieri, 2013; Dingemanse & Réale, 2005; Réale et al., 2007; Sih, Bell, & Johnson, 2004; Sih, Bell, Johnson, & Ziemba, 2004). Five ecologically important personality axes have been identified to characterize the behavioural responses of individuals when interacting with their environment (activity, exploration, boldness) and with others (aggressiveness, sociability; Réale et al., 2007). Personality traits have been shown to depend on ecological parameters (e.g. Réale et al., 2007; Sih, Cote, Evans, Fogarty, & Pruitt, 2012), to be heritable (e.g. van Oers, de Jong, van Noordwijk, & Drent, 2005; van Oers & Sinn, 2013), to be linked to life history traits or fitness (Dingemanse, Both, Drent, & Tinbergen, 2004; Dingemanse & Réale, 2013; Duckworth & Kruuk, 2009; Reale et al., 2010; Schuett, Tregenza, & Dall, 2010; Smith & Blumstein, 2008) and often to correlate with each other at the individual level, forming so-called behavioural syndromes (Garamszegi, Markó, & Herczeg, 2012; Andrew Sih, Bell, & Johnson, 2004; van Oers & Sinn, 2013). Such correlations may result from a functional integration of personality traits favoured by selection when interactions between these traits increase individuals' fitness in given environmental conditions (e.g. Dingemanse et al., 2007). Altogether, these various results reveal the crucial role that personality traits may play in shaping evolutionary processes in wild populations (Dingemanse et al., 2004; Duckworth & Badyaev, 2007; Karlsson Green, Eroukhmanoff, Harris, Pettersson, & Svensson, 2016; Niemelä, Lattenkamp, & Dingemanse, 2015).

Understanding the evolution of personality traits and their associations in behavioural syndromes requires understanding the mechanisms underlying these between-individual differences in behaviour and their interactions, including their genetic basis. Phenotypic correlations between personality traits result from the addition of between-individual correlations (defining behavioural syndromes per se, whether genetic or nongenetic) and within-individual (or residual) correlations (Brommer, 2013; Dingemanse & Réale, 2013; Dingemanse, Kazem, Réale, & Wright, 2010). Assessing the relative contribution of between- and within-individual correlation components in observed phenotypic correlations can be crucial because between-individual correlations may constrain the independent evolution of the traits involved and thus may have major evolutionary consequences in the wild (Sih, Bell, & Johnson, 2004).

To reliably quantify this relative contribution of between- and within-individual correlations, multiple measurements of the personality traits considered must be collected on a large number of individuals. When only single measurements are available, it is often assumed that a phenotypic correlation observed between personality traits reflects an underlying between-individual correlation, i.e. the within-individual correlation is negligible (the so-called 'individual gambit', Brommer, 2013). Furthermore, to quantify genetic variance in personality traits and genetic covariance between them, the level of genetic relatedness between individuals measured has to be incorporated (e.g. via pedigree information) in the models. When relatedness information is unavailable, it is often assumed that the observed phenotypic (co)variance reflects the underlying genetic (co)variance (the so-called 'phenotypic gambit'; Grafen, 1984; Hadfield, Nutall, Osorio, & Owens, 2007; van Oers & Sinn, 2011).

These two crucial assumptions have been tested empirically in various species, and recent meta-analyses including over 30 studies, among which 25 are from wild populations, have confirmed their overall validity (Brommer & Class, 2017; Dochtermann, 2011; Dochtermann, Schwab, & Sih, 2015). Across these studies, the sign (and to a certain extent the magnitude) of the phenotypic correlations between personality traits reliably informed on the sign (and the magnitude) of the between-individual correlations (Brommer & Class, 2017) and of the genetic correlations (Dochtermann, 2011). Furthermore, 52% of the between-individual variation in personality traits taken separately was explained by additive genetic variance (Dochtermann et al., 2015). A recent empirical study on a wild population of yellow-bellied marmots, *Marmota flaviventris*, estimated the proportion of phenotypic (co)variance explained by genetic, permanent environment and maternal (co)variances between four different personality traits: docility, exploration, activity and sociability (Petelle, Martin, & Blumstein, 2015). Results showed additive genetic variations, as well as maternal and permanent environment variations, in all four traits and a positive genetic correlation between activity and sociability (Petelle et al., 2015). More of such integrative studies partitioning the observed phenotypic (co)variance in multiple personality traits simultaneously are needed in different biological models with contrasting life histories to better understand the mechanisms underlying and possibly constraining the evolution of correlated personality traits.

In this study, we assessed the genetic basis of three personality traits, together with the relative contribution of between- and within-individual variations in, and correlations between, these traits, chosen for their potentially important effects on crucial ecological processes (here, nest site acquisition and defence against competitors and predators) in a natural population of a small territorial, short-lived, migrant passerine bird, the collared flycatcher, *Ficedula albicollis*. During 3 consecutive years, we measured for several hundreds of breeding pairs in the field (1) aggressiveness towards competitors (as the agonistic reaction to simulated territorial intrusions by intra- and interspecific competitors), (2) boldness towards predators (as the latency to resume nestling feeding after human disturbance) and (3) neophobia, possibly reflecting exploration (as the latency to resume nestling feeding in the presence of a novel object on the nest site; following Réale et al.'s 2007 definitions). To identify the mechanisms underlying the phenotypic (co)variation in these behavioural traits, we assessed to what extent (1) additive genetic, parental or permanent environment effects contributed to the observed phenotypic (co)variance, while accounting for fixed individual (sex, age) covariates, and (2) between-individual correlations explained phenotypic correlations between these three traits. Based on many previous studies on personality traits in populations of passerines of similar ecology (e.g. Dingemanse, Both, Drent, van Oers, & van Noordwijk, 2002; Drent, Oers, & Noordwijk, 2003; Duckworth & Badyaev, 2007; Garamszegi et al., 2015; Garamszegi, Rosivall, et al., 2012; van Oers, Drent, de Goede, & van Noordwijk, 2004), we expected heritable differences to partly explain variation in aggressiveness, boldness and neophobia in our study population. Furthermore, high aggressiveness, high boldness and low neophobia may allow individuals to efficiently secure and defend a breeding site when they are unfamiliar with the environment (e.g. for dispersers: Cote, Clobert, Brodin, Fogarty, & Sih, 2010; Duckworth & Kruuk, 2009; Korsten, van Overveld, Adriaensen, & Matthysen, 2013). Thus we predicted a functional integration and (possibly genetically based) between-individual correlations between these traits. In another population of collared flycatchers, male aggressiveness and boldness, but not neophobia, were found to be phenotypically correlated in some years (Garamszegi et al., 2015; Garamszegi, Eens, & Török, 2009). This population and ours, however, differ greatly in both demographic functioning (e.g. male age structure: Hegyi, Rosivall, & Török, 2006) and selective pressures (e.g. nest predation: Doligez & Clobert, 2003; sexual selection: Qvarnström, 1997; see also Rosivall, Török, Hasselquist, & Bensch, 2004), which may affect the fitness consequences of interactions between personality traits, and thus their potential functional integration.

METHODS

Study species and population monitoring

Collared flycatchers are migratory cavity nesters and readily breed in artificial nestboxes, providing easy access to parents' identity and breeding data. Between 2011 and 2013, we conducted the behavioural tests (see below) on 1131 pairs breeding in nestboxes spread over 14 to 22 forest patches in our study population located on the island of Gotland (Sweden, Baltic Sea). Each year since 1980, nests in boxes have been monitored at least weekly from late April until early July, allowing us to record major breeding variables (laying and hatching dates; clutch size; nestling number, condition and fledging success). Breeding pairs were captured, identified and ringed if previously unringed; females were caught during incubation and males while feeding nestlings. Nestlings were ringed between day 8 and day 13 after hatching; fledging typically occurs 16 days after hatching. Adult and nestling identification every year combined with a high return rate of both adults (approximately 40%) and juveniles (approximately 10%) for such a short-lived passerine bird (Gustafsson, 1989) allowed us to establish a high-quality social pedigree of the population which has previously been used in several quantitative genetic studies (e.g. Evans & Gustafsson, 2017; Merilä & Sheldon, 2000; Sheldon, Kruuk, & Merilä, 2003; Appendix Table A1). In this population, approximately 15% of all nestlings are extrapair (Sheldon & Ellegren, 1999), a percentage considered low enough for quantitative genetic models to provide valid (i.e. only slightly underestimated) heritability estimates from the social pedigree (Charmantier & Réale, 2005; Firth, Hadfield, Santure, Slate, & Sheldon, 2015), even though no information is available yet on how extrapair paternity may affect genetic covariances. The clear sexual dimorphism in plumage coloration in this species allowed an easy discrimination of adult males (black and white plumage with a white forehead patch) from females (brown plumage; Svensson, 1992), even from several metres away during behavioural tests. Adults could also be aged by plumage criteria (yearlings versus older adults; Svensson, 1992).

Aggressiveness score

We measured the level of aggressiveness of breeding flycatchers soon after settlement, during nest building or early laying, that is, when the risk of losing a nestbox to a competitor is highest in this single-clutch species. During the breeding season, collared flycatchers compete for nest sites with conspecifics but also with great tits, *Parus major*, the second most abundant species breeding in nestboxes in the study area (Gustafsson, 1987). Aggressiveness towards conspecific intruders was shown to decrease after the start of incubation (Král & Bícík, 1989) even though aggressiveness towards great tit intruders remained high throughout the breeding cycle (Král & Bícík, 1992). To elicit an aggressive response from a focal flycatcher pair, we simulated the intrusion of competitors at the nest of the pair by attaching to its nestbox clay decoys mimicking either a flycatcher pair or a single (male) great tit. We used a pair (one male and one female) for flycatcher decoys to elicit and measure an aggressive response by both pair members, that is, to avoid a sex-specific response towards this intraspecific stimulus, while the response to the interspecific stimulus (great tit decoy) was not expected to differ depending on the sex of the decoy. In addition, we simultaneously broadcast male songs corresponding to the decoy(s) species with a loudspeaker placed just under the nestbox. To avoid pseudoreplication, we randomly used one of eight different sets of decoys and one of five different song tracks per species for each test. After attaching the decoys to the nestbox and the loudspeaker under the box, the observer sat under a camouflage net approximately 8–10 m away from the nestbox and recorded the following behaviours for each pair member: (1) movements between perches and perching position (within 2 m, between 2 and 5 m or between 5 and 10 m away from the nestbox), (2) agonistic behaviours towards a decoy (attacks and stationary flights in front of the decoy) and (3) chases towards living birds attracted by the stimulus. A behavioural test started with an observation period of 15 min but we lengthened the test by up to 5 min when an individual arrived between 10 and 15 min after the start of the test, and up to 5 additional min if its partner arrived during this extra time, so that we could observe the behavioural response of each pair member for at least 5 min. Each test thus lasted between 15 and 25 min. If an individual was observed during less than 5 min before the end of the test, it was discarded from the analyses and these observations were therefore not used later on.

Aggressiveness level was measured as the sum of the number of movements within 2 m of the nestbox, attacks, stationary flights and chases. We included this latter behaviour because chasing a live intruder may have prevented the focal flycatcher from interacting with the decoy, while reflecting an aggressive territory defence response. The number of each type of behavioural response (movements, agonistic behaviours and chases) was standardized by the time interval between the first observation of the individual and the end of the test, rescaled to 15 min. Using alternative scores did not qualitatively change the results (see Appendix and Table A2). We conducted aggressiveness tests two to four times per focal pair over a 5-day interval, with at most one test per day and tests on 2 days in a row. The stimulus used (intra- / interspecific decoys) was alternated between tests after a random assignment for the first test. An aggressiveness score was computed for each individual for each test. We obtained aggressiveness responses (for more than 5 min at least once per year) for 1974 individuals (including unidentified ones; 961 females and 1014 males in 1046 nests). Among those, 825 females and 667 males were later captured and identified, and thus used for heritability estimation. In 601 breeding pairs both partners were identified and responded to the tests and in 273 only one partner responded. We obtained repeated estimates for 502 and 445 identified females and males, respectively.

Boldness and neophobia scores

During nestling rearing, we estimated (1) boldness level by measuring the individual's reaction towards the presence of a human observer near the nestbox and (2) neophobia level by measuring the reaction towards the presence of a novel object on the nestbox (i.e. in a familiar environment), following the definitions from Réale et al. (2007). As advised in Greenberg and Mettke-Hofmann (2001), we measured our behavioural responses as the latency to perform a highly motivating action (here, feeding their nestlings) after disturbance. We conducted one combined boldness/neophobia test per breeding pair when the nestlings were 5 days old, that is, at the beginning of the period of highest provisioning by parents (and before we caught the parents to avoid any behavioural interference). A test consisted of two consecutive periods of approximately 1 h each: the behaviour of the parents was recorded first without any change in the surroundings of the nestbox, that is, without the novel object, and second with a novel object (here a coloured figurine approximately 7 cm high) attached near the entrance hole of the nestbox. Both periods were video-recorded from a distance (6–8 m). At the beginning of each period, the observer checked the camouflaged video recorder, walked to the nestbox, opened it to check nestling satiety, closed it, and then left the area. The test was abandoned if the nestlings were very hungry to avoid them starving if the parents were too disturbed by the test.

We estimated boldness score using the latency to enter the nestbox after the departure of the observer in the first period (i.e. without the novel object). Reaction to disturbance by humans has previously been used in boldness tests in this species (e.g. Garamszegi et al., 2009). To ease interpretation (i.e. increasing values of boldness score corresponding to increasing level of boldness), we transformed the latency to enter the nestbox such that the boldness score of an individual was the maximum latency observed in the entire data set minus the latency for this individual. We estimated neophobia score based on the latency to enter the box after the departure of the observer in the second period (i.e. in the presence of the novel object). For both boldness and neophobia scores, individuals that did not enter the nestbox during the first period of the test were not used in the analysis (187 of 1251 observations, i.e. 15%). Individuals that entered the nestbox during the first but not the second part of the test (411 of 1064 observations, i.e. 39%) were considered as the most neophobic ones but could not be assigned a latency. To include them in the analyses, we discretized the latency to enter the nestbox in the second period into four categories based on its quartiles, assigning values from 1 to 4 for increasing latencies, and adding a fifth category including individuals that did not enter in the second part of the test. Using alternative scores for boldness and neophobia did not qualitatively change the results (see Appendix and Tables A2 and A3). We obtained boldness and neophobia estimates for 849 identified individuals (472 females and 378 males). Over the 3 years, we assessed 403 unique breeding pairs where both identified partners responded to the tests and 185 pairs where only one partner responded. We obtained repeated boldness and neophobia scores (i.e. several years in a row) for 66 females and 65 males.

Repeatability and heritability of aggressiveness, boldness and neophobia

We estimated the repeatability of aggressiveness, boldness and neophobia scores as well as their heritability by fitting three separate univariate animal models. The models included the following random effects: additive genetic effect (associated with the pedigree), individual identity for the repeated measures per individual (permanent environment effect), maternal and paternal identities, forest patch, observer identity (the person observing and reporting the behaviours onsite for the aggressiveness tests and the person extracting latencies from the video recording for the boldness and neophobia tests). The models of the aggressiveness score also included the broadcast song track and decoy set identifiers. In addition, the models included the following fixed effects, to control for potential confounding factors: sex, age (two levels: yearling versus older) and their interaction, as well as year (three levels: 2011, 2012, 2013). The aggressiveness model also included stimulus type (two levels: flycatcher versus great tit decoys), the order of the test within a year (continuous variable: first to fourth), the presence of the partner during the test (binary variable: yes/no) and the presence of other live flycatchers or great tits (binary variable: yes/no). The boldness and neophobia models included the number of ringed nestlings as a proxy of the motivation to enter the nestbox to feed nestlings. The neophobia models included the boldness score to control for the effect of the human disturbance at the beginning of the period with the novel object. All continuous fixed terms were centred and standardized prior to analysis to allow comparisons between effects.

Repeatabilities (R) were estimated as the ratio of the sum of the additive genetic (V_A), permanent environment (V_{PE}), maternal and paternal identities variances (V_M and V_F, respectively) over the total phenotypic variance (V_P, sum of all variances; Falconer & Mackay, 1996; with possibly an additional term accounting for the distribution variance, Nakagawa & Schielzeth, 2010). Narrow-sense heritabilities (h^2) were estimated as the ratio of the additive variance V_A over the phenotypic variance V_P. The presence of fixed effects in models did not result in over- or underestimating repeatability and heritability estimates (as warned against in Nakagawa & Schielzeth, 2010), because estimates were similar when obtained from models with only the intercept (presented in the main text) and from models with the previously described fixed effects (see Appendix Tables A1, A2, A3; except for a slight overestimation of the boldness and neophobia repeatabilities on the latent scale). For aggressiveness, we also estimated within-year repeatability by replacing the additive genetic and permanent environment effects by a unique identifier per individual per breeding season.

Correlations between aggressiveness, boldness and neophobia scores

We estimated the between- and within-individual correlations between aggressiveness, boldness and neophobia scores by fitting a trivariate mixed-effects model (Dingemanse & Dochtermann, 2013). For this model, aggressiveness was averaged over all aggressiveness scores of an individual in a given year (i.e. over up to four estimates). We chose this averaging approach because aggressiveness score (1) differed depending on decoy species, chosen at random for the first test, and (2) decreased with the order of the test (i.e. due to habituation). Averaging all aggressiveness scores of an individual in a given year allowed us to control for these differences and decrease the effect of varying environmental (meteorological) conditions between tests. In this trivariate model, we included sex and year as fixed effects and ring number as a random effect. The phenotypic correlation between two traits *A* and *B*, $r_{\mathrm{P_A,P_B}}$, and the between-individual correlation between the traits *A* and *B*, $r_{\mathrm{ind_A,ind_B}}$, were estimated as follows (Snijders & Bosker 1999):

$$r_{\mathrm{P_A,P_B}} = \frac{\mathrm{Cov}_{\mathrm{ind_A,ind_B}} + \mathrm{Cov}_{\varepsilon_\mathrm{A},\varepsilon_\mathrm{B}}}{\sqrt{V_{\mathrm{P_A}} \times V_{\mathrm{P_B}}}} \qquad (1)$$

$$r_{\mathrm{ind_A,ind_B}} = \frac{\mathrm{Cov}_{\mathrm{ind_A,ind_B}}}{\sqrt{V_{\mathrm{ind_A}} \times V_{\mathrm{ind_B}}}} \qquad \text{(eq. 2)}$$

where $\mathrm{Cov}_{\mathrm{ind_A},\mathrm{ind_B}}$ and $\mathrm{Cov}_{\varepsilon_A,\varepsilon_B}$ are the between- and the within-individual covariances between traits A and B, and $V_{\mathrm{P,\ A\ or\ B}}$ the total phenotypic variance (sum of the between- and within-individual variances) associated with trait A or B. Combining aggressiveness, boldness and neophobia score data for a given individual in a given year, we obtained 841 observations for which estimates for all three traits were available, and 152 observations for which only boldness and neophobia estimates were available. Among these 841 observations, 49 females and 52 males were repeatedly assessed over several years for all three traits, providing 100 and 111 repeated observations, respectively.

To estimate the additive genetic correlations between traits, we fitted a model with the same fixed effects but with the additive genetic effect instead of the individual ring as a random factor. Fitting both additive genetic and individual (permanent environment) effects together in a single model indeed led to convergence failures. We thus fitted only one effect at a time.

Implementation of Bayesian models

All statistical analyses were performed within the Bayesian framework in R v.3.3.2 (R Core Team, 2016). Both univariate and trivariate models were fitted using the function MCMCglmm ('MCMCglmm' R package, Hadfield, 2010). The pedigree was prepared using the function fixPedigree ('pedantics' R package, Morrissey & Wilson, 2010) and pruned using the function prunePed ('nadiv' R package, Wolak, 2012; see Appendix Table A1 for a description of the pedigree). Aggressiveness and boldness scores were fitted with a Poisson family (logit link), and neophobia scores with a Threshold family with the residual variance fixed to 10 (instead of the usual value of 1, to improve the mixing of the chains for low variances, which were expected from preliminary analyses; Hadfield, 2016). We used wide normally distributed priors for fixed effects (large variance V=10^8; Hadfield, 2016) and parameter expanded χ^2 distributed priors with 1 degree of freedom for random effects. For the univariate models, we adjusted the number of iterations, burn-in and thinning interval for each model so as to obtain an effective sample size over 1500 (see Appendix) and autocorrelations of posterior samples below 0.1 in all cases. For the trivariate models, we used 4 x 10^6 iterations, a burn-in of 10^5 and a thinning interval of 2000 to reach the same criteria. We visually assessed the convergence of each MCMC chain and compared three chains per model using the Gelman & Rubin approach (gelman.diag and gelman.plot functions, 'coda' R package, Plummer et al., 2016). Following Nakagawa and Schielzeth (2010), we retained in our univariate models data from individuals tested only once. For all three traits, estimates are presented as posterior modes with the associated 95% credible intervals (CI). Variance, repeatability and heritability estimates are presented on the latent scale (R_{latent}, following Nakagawa & Schielzeth, 2010; h^2_{latent} following de Villemereuil, Schielzeth, Nakagawa, & Morrissey, 2016). We also provide repeatability and heritability estimates on the observed scale (R_{obs} using the QGicc function, h^2_{obs} using the QGparams function, from the 'QGglmm' R package; de Villemereuil, 2018; de Villemereuil, Schielzeth, Nakagawa, & Morrissey, 2015). It is not possible here to discuss all results on the observed scale, even though it is the scale of the realized behaviour, that is, where natural selection can act (see de Villemereuil et al., 2016). Indeed, the estimation of R_{obs} for ordinal traits (here the neophobia score) is complex and is not currently implemented in QGicc. To allow us to compare estimates between traits on the observed scale, we computed repeatabilities for neophobia using two alternative scores with nonordinal distributions, namely (1) the latency to enter the nestbox in the presence of the novel object for individuals that entered the nestbox (N=653, fitted with a Poisson distribution), and (2) whether the individual entered the nestbox in the presence of the novel object (N=1064, binary variable fitted with the threshold family). Moreover, the estimation of h^2_{obs} for ordinal traits provides one heritability estimate per level, which in the case of an artificial categorization as here is not biologically relevant (de Villemereuil, 2018). As the heritability estimates on the observed scale were fairly similar between neophobia levels, we compared the range of heritability estimates found for neophobia with the heritability estimates found on the observed scale for aggressiveness and boldness. Correlations are provided on the latent scale.

Ethical note

Permission for catching and ringing adults (here 838 yearlings, 1074 older birds) and nestlings (here 9750) with individually numbered aluminium rings was granted every year by the Ringing Centre from the Museum of Natural History in Stockholm (licence nb. 471: M015 to B.D.). Adults were caught in the nest, either directly (females during incubation) or using swinging-door traps (both parents during nestling rearing). Traps were set for at most 30 - 60 min depending on nestling age (30 min when nestlings were 5 days old or younger), to avoid nestling starvation if parents did not resume feeding during the catching period; traps were checked every 5–10 min, and removed as soon as the adults had been caught. Catching sessions started after 0600 hours to let birds feed and provision nestlings undisturbed for at least 2 h after the night period (sunrise occurs at approximately 0400 hours during spring). Adults were handled for 5 - 10 min and released straight after manipulation or (when catching both parents during nestling feeding) kept until the partner was captured (up to 40 min maximum). For nestling ringing, whole broods were taken directly from the nest and ringed just beside the nestbox (for approximately 10 min); nestlings that were not handled were kept warm using small heating packs. During the aggressiveness tests, we minimized disturbance by approaching the nestbox as quietly as possible and hiding below a camouflage net. Conversely, for the combined boldness/neophobia test, which aimed at measuring (or controlling for) the reaction towards human presence, we on purpose approached the nestbox conspicuously. During the boldness/neophobia test, nestling satiety was checked at the beginning and in between the two parts of the test, and the test was aborted if nestlings were begging too strongly, to avoid any harmful effect of temporarily decreased provisioning by parents. All the manipulations were performed in accordance with the Swedish legislation applying at the time.

RESULTS

Univariate models

The level of repeatability for aggressiveness was 0.18 on the latent scale and 0.03 on the observed scale (Table 1). In addition, aggressiveness score was repeatable within years (R_{latent}=0.22, 95% CI = [0.18; 0.26]; R_{obs}=0.04, 95% CI = [0.03; 0.06]) and between years (when averaging the aggressiveness score of 1 year; R_{latent}=0.26, 95% CI = [0.11; 0.38]; R_{obs}=0.11, 95% CI = [0.04; 0.20]). The level of repeatability for boldness was 0.11 on the latent scale and 0.10 on the observed scale (Table 1). Neophobia was slightly more repeatable that the other traits on the latent scale (Table 1) as well as on the observed scale when estimated from alternative nonordinal measures (R_{latent}=0.25, 95% CI = [0.06; 0.35] and R_{obs}=0.14, 95%CI = [0.04; 0.23] for the latency to enter the nestbox in the presence of the novel object; R_{latent}=0.44, 95% CI = [0.22; 0.67] and R_{obs}=0.29, 95%CI = [0.12; 0.44] for whether the individual entered the nestbox during the test or not). None of the three behavioural scores, however, were heritable (all 95% CI for V_A and h^2 values included zero; Table 1). Permanent environment effects explained 15% of the phenotypic variance in aggressiveness score (95% CI of V_{PE} = [0.28; 0.80]; Appendix Table A2) and parental identities (i.e. maternal and paternal identities summed) explained 25% of the phenotypic variance in neophobia score (95% CI of $V_M + V_F$ = [0.38; 8.46], even though the lower limit of the 95% CI for each parent identity separately was 0: [0.00; 6.49] for V_M and [0.00; 5.75] for V_F). When excluding the maternal (paternal) identity from the model, the paternal (maternal) identity explained 19% (18%) of the phenotypic variance. Fitting the neophobia model without the maternal and paternal identities did not change the heritability estimate, revealing that these effects were not confounded with the additive genetic effect (results not detailed). Observer identity explained 7% of the phenotypic variance for aggressiveness (95% CI of $V_{Observer}$ = [0.14; 0.55]); paternal identity explained 11% of the phenotypic variance for aggressiveness but only when measuring aggressiveness as the first axis of a principal component analysis (see Appendix). All other variances were low (less than 4 % of the phenotypic variance) or not different from zero (Appendix Tables A1, A2 and A3).

Males were more aggressive than females, especially among yearlings (interaction sex*age, with yearling males as reference: 95% CI = [0.44, 0.81]; Appendix Fig. A1a). In addition, males were slightly shyer (longer latency to enter in the absence of a novel object) and less neophobic (shorter latency in the presence of a novel object) than females (with female as a reference 95% CI = [-0.12; -0.02] and [-2.78; -1.07] respectively; Appendix Fig. A1b), and this did not depend on age (see Appendix Tables A2 and A3 for sex*age interactions). In addition, individuals with larger broods were bolder (95% CI = [0.002; 0.06]; Table A3) and less neophobic (95% CI = [-1.03; -0.37]; Table A4). Regarding environmental effects, aggressiveness, boldness and neophobia scores depended on the year: individuals were less aggressive and less neophobic in 2011 than 2012 (Appendix Tables A2 and A4), and shyer in 2011 than 2013 (Appendix Table A3). Finally, individuals were more aggressive in the presence of their partner or neighbouring tits attracted by the stimulus, and during the first tests of the sequence (Appendix Table A2).

Trivariate model

Aggressiveness and neophobia scores were phenotypically correlated: more aggressive individuals were less neophobic (Table 2, Appendix Fig. A2a). Boldness and neophobia scores were also phenotypically correlated: bolder individuals were less neophobic (Table 2, Appendix Fig. A2b). The corresponding within-individual correlations were negative (Table 2). However, there was no phenotypic correlation between boldness and aggressiveness (Table 2; see Appendix Table A5 for the full model output) and none of the between-individual correlations differed from zero (Table 2). When accounting for an additive genetic effect, none of the additive genetic covariances differed from zero (Table 2; see Table A6 for the full model output).

DISCUSSION

In this study, we tested whether three personality traits (aggressiveness, boldness and neophobia) had a genetic basis in a wild population of collared flycatchers and formed (genetically based) behavioural syndromes during breeding, to shed light on constraints in their possible evolution. None of the three traits were heritable and their repeatability estimates were low (0.11–0.39 on the latent scale for all traits; 0.03–0.10 on the observed scale for aggressiveness and boldness; 0.14 and 0.29 on the observed scale for nonordinal measures of neophobia) compared to average estimates previously found for behavioural traits in two meta-analyses (0.37 in Bell et al., 2009; 0.41 in Holtmann et al., 2017), suggesting strong phenotypic plasticity in these traits. The repeatability originated mainly from permanent environment effects for aggressiveness and from parental effects for neophobia. In addition, we found that the three traits showed phenotypic covariance but no between-individual covariance and no additive genetic covariance. The absence of behavioural syndromes among these personality traits may be due either to a lack of statistical power to detect between-individual covariances or to an absence of functional integration of these traits at the individual level and no genetic correlation at the population level.

Factors at the origin of behavioural trait repeatability

Our levels of repeatability, estimated both within and between years for aggressiveness score and between years for boldness and neophobia scores, were lower than usually reported for such behaviours: around 0.50 for aggressiveness and exploration, and around 0.40 for antipredator behaviours (Bell et al., 2009). Interestingly, the repeatability level estimated here for aggressiveness score was similar within and between years, contrary to the usual decrease observed when the time interval between recordings increases (Bell et al., 2009; Chervet, Zöttl, Schürch, Taborsky, & Heg, 2011; Dingemanse et al., 2012; Garamszegi et al., 2015; Holtmann et al., 2017; Wuerz & Krüger, 2015; but see David, Auclair, & Cézilly, 2012 for differences between traits). Overall, our lower levels of repeatability, especially on the observed scale, suggest higher plasticity, both within and between years compared to studies on other species.

The observed repeatability in aggressiveness score resulted mostly from permanent environment effects, which explained 15% of the phenotypic variance and 72% of the repeatability in aggressiveness score. Here, because we controlled for the identity of the parents, permanent environment effects could be linked to differences in individual condition or experience. Some measures of personality traits have indeed been found to depend on individual condition or experience (reviewed in Stamps & Groothuis, 2010). In our population, condition and experience also affect breeding habitat choice depending on social cues (e.g. Doligez, Danchin, Clobert, & Gustafsson, 1999; Doligez, Pärt, Danchin, Clobert, &

Gustafsson, 2004; Kivelä et al., 2014), which could shape individuals' response to the risk of competition for nest sites. Permanent environment effects may also include a dominance effect (Kruuk & Hadfield, 2007; Wilson et al., 2010), which could not be directly modelled here because full- and half-sib links were too rare in our pruned pedigree for running such complex models (Wilson et al., 2010).

In turn, the observed repeatability in neophobia resulted mostly from parental identities, which accounted for 25% of the phenotypic variance and 55% of the between-individual variance. Both pre- and postnatal parental effects have been found to affect exploration and neophobia behavioural responses later in life (e.g. nestling provisioning and exploration in birds, Carere, Drent, Koolhaas, & Groothuis, 2005; maternal hormones early in life and neophobia, Spencer & Verhulst, 2007; see the review in Groothuis & Maestripieri, 2013). However, parental identities did not explain between-individual differences in aggressiveness and boldness here, contrary to previous findings (e.g. Eising, Muller, & Groothuis, 2006; reviewed in Groothuis & Maestripieri, 2013). These behavioural responses may be more dependent on individual or local environmental conditions, in particular individual competitive ability and neighbour/predator presence or density, at the time of the test(s). To better understand how parental effects shape behavioural responses later in life in our study population, further experiments (e.g. nestling cross-fostering) would be necessary.

Personality traits with no genetic basis

We found no genetic basis for our three personality traits. A meta-analysis on personality traits in wild animal populations estimated an average heritability level of 0.28 for aggressiveness, 0.31 for boldness and 0.58 for exploration-avoidance (including estimates from novel environment and novel object tests; van Oers & Sinn, 2013). The absence of heritability for our personality traits here was not due to a lack of statistical power to detect significant additive genetic variance based on our sample and social pedigree, because based on the same sample with the same pedigree, we obtained positive heritability estimates for tarsus and wing length (h^2=0.59, 95% CI = [0.44; 0.69] for tarsus and h^2=0.30, 95% CI = [0.11; 0.50] for wing length, while accounting for maternal and permanent environment effects) which are consistent with previous estimates in this population (h^2 = 0.53 and 0.51 for tarsus and wing length, respectively, in Merilä & Gustafsson, 1993). The absence of heritability in our personality traits was therefore likely to be the result of very low additive genetic variance combined with large environmental variance as illustrated, for instance, by between-year differences in behavioural scores, which reflected large variations in environmental conditions between the 3 years of our study (see Morinay, Forsman, Kivelä, Gustafsson, & Doligez, 2018 for differences between 2012 and 2013). Large environmental variance could originate from individuals being tested in different environments (including the social context) in different years, because between-year fidelity to the nestbox and/or partner is very low in this population (approximately 6.7% of 240 individuals bred in the same nestbox several years and 1.0% of 214 identified pairs were faithful over several years). This, however, limited the risk of pseudoreplication (Niemelä & Dingemanse, 2017).

Because we measured personality traits at the nest during breeding, the reaction of the partner may have affected the reaction of the focal bird during a behavioural test, as found here with a higher aggressiveness score when the partner was present. To account for this effect, we could have included the partner's identity and genetic background (i.e. pedigree) as random effects in our models (see Morinay et al., 2018 for an example in the same population; and Wolf, Brodie III, Cheverud, Moore, & Wade, 1998 for so-called indirect genetic effects). However, this could have led to pseudoreplication, because the behavioural score of the partner itself was most of the time also analysed in this dataset (e.g. around 74% of the females and 90% of the males had their partner tested). To keep exploring a response at the individual level (rather than combining behavioural scores at the pair level), a solution could be to fit a bivariate model of the two partners' responses and include both their pedigree, permanent environment and parental effects in the model. This would, however, require a larger data set than used here to reach sufficient statistical power to detect such effects with such complex models. Furthermore, because the focal bird chooses at least partly its partner (like its nest site), partner's effects can also be expected to be at least partly included in the individual's genetic and permanent environment effects. Disentangling such complex effects may require a more balanced sample of faithful and divorced pairs breeding in the same and different sites over several years than observed in our population.

No personality syndrome?

Phenotypic correlations were observed between our personality traits, even though they did not constitute behavioural syndromes (i.e. there was no between-individual correlations): less neophobic individuals were more aggressive and bolder. This was in line with previous studies reporting bolder individuals to be more explorative in a novel environment (or less neophobic in a novel object test; e.g. Garamszegi et al., 2009; van Oers, De Jong, Drent, & van Noordwijk, 2004). Conversely, the absence of correlation between aggressiveness and boldness partly contrasts with previous results reporting more aggressive individuals to be bolder, as part of the proactive–reactive axis, in different species (Koolhaas et al., 1999; Sih, Bell, Johnson, et al., 2004) including the collared flycatcher (Garamszegi et al., 2015).

The observed phenotypic correlations resulted solely from correlated changes in behaviours between measurements for the same individuals, that is, within-individual correlations. Within-individual correlations could be due to micro-environmental effects (e.g. nestbox environment), to individual effects (e.g. long-term between-year plasticity but short-term within-year behavioural constraints, for instance due to experience) or to correlated measurement errors (Dingemanse & Dochtermann, 2013). Error correlation, however, is more likely to occur between boldness and neophobia scores, which were extracted from the same test and might both be correlated with feeding rate (with birds investing more in nestling provisioning returning more rapidly to their nest in both situations), than between aggressiveness and neophobia scores, which were measured several weeks apart by different persons in different tests. To tease these sources of within-individual correlations apart, aggressiveness, boldness and neophobia scores need to be estimated several times during the same breeding season and possibly the same phase(s) of the reproductive cycle. The limited number of individuals measured several times here (211 observations of 101 individuals), however, is likely to be why we did not detect between-individual covariance (Dingemanse & Dochtermann, 2013 recommended sample sizes of at least 200 individuals tested twice; see also Garamszegi & Herczeg, 2012). Indeed, based on the same sample, we were not able to obtain positive genetic or between-individual covariances between tarsus and wing length, two morphological traits previously reported as genetically correlated in the same population (Merilä & Gustafsson, 1993).

Even though our limited statistical power does not allow us to conclude the absence of behavioural syndromes, this absence, if true, would suggest that selective pressures did not yield or maintain a functional integration between the personality traits investigated here. A true absence of behavioural syndromes among the traits we studied could be explained by specific breeding conditions in our population, possibly altering the selective regimes compared to other populations or species. In our population, the high availability of high-quality nest sites (i.e. nestboxes, provided in excess since the early 1980s) may have released joint selective pressures on exploration to find suitable nest sites and aggressiveness to acquire and defend this resource against dominant competitors (in particular tit species) in a natural context. In turn, providing nestboxes probably increased local breeding densities and thereby competition for food resources during the nestling period, especially in a highly synchronous species such as the collared flycatcher. Furthermore, our population is subjected to very low nest predation rates, due to the absence of mustelid species on Gotland (Doligez & Clobert, 2003), which may have released selective pressures on boldness through the decrease in the need to defend the brood. Overall, these specific breeding conditions may have strongly modified the selective regime for personality traits and for a functional integration between them if they are costly.

In conclusion, we showed that aggressiveness, boldness and neophobia are repeatable but not heritable traits and do not seem to form behavioural syndromes in our population of collared flycatchers since only phenotypic correlations were observed between neophobia and the other two traits. Our study thus brings insights on the evolutionary potential of these personality traits alone and in interaction with each other during breeding in a wild population experiencing particular breeding conditions (low competition for nest sites, low nest predation rate). To understand the absence of individual covariance between, and heritable variations in, personality traits in our population, a first step would be to investigate the fitness benefits (i.e. reproductive success and survival) associated with each trait and their interactions. Plasticity in the associations between personality traits should be selected for if the fitness costs and benefits of expressing each trait relative to the others depend on the environmental (including social) context (e.g. competition level or predation risk), which remains to be explored for example by experimentally manipulating these environmental conditions.

Conflict of interest

We declare no conflict of interest.

Acknowledgments

We are grateful to the landowners from Gotland for allowing us to collect data on their properties, to all fieldworkers who collected the data between 2011 and 2013, and to all the students who contributed to the extraction of data from the audio and video recordings and to building the behavioural database. We also thank Ingrid Ahnesjö and Emmanuel Desouhant, and three anonymous referees for their critical comments on the manuscript. J.M. was funded by the Ministère de l'Enseignement Supérieur et de la Recherche (PhD grant), by Uppsala University (PhD grant and Stiftelsen research grant) and by the University of Lyon (IDEX mobility grant). G.D. was funded by the Région Auvergne Rhône-Alpes (PhD and Explora'Doc grants). L.G. was funded by the Swedish Research Council and B.D. by the Centre National pour la Recherche Scientifique.

References

Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 755–761. http://doi.org/10.1098/rspb.2006.0199

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*(4), 771–783. http://doi.org/10.1016/j.anbehav.2008.12.022

Brommer, J. E. (2013). On between-individual and residual (co)variances in the study of animal personality: Are you willing to take the 'individual gambit'? *Behavioral Ecology and Sociobiology*, *67*(6), 1027–1032. http://doi.org/10.1007/s00265-013-1527-4

Brommer, J. E., & Class, B. (2017). Phenotypic correlations capture between-individual correlations underlying behavioral syndromes. *Behavioral Ecology and Sociobiology*, *71*(3). http://doi.org/10.1007/s00265-017-2278-4

Carere, C., Drent, P. J., Koolhaas, J. M., & Groothuis, T. G. G. (2005). Epigenetic effects on personality traits: Early food provisioning and sibling competition. *Behaviour*, *142*(9), 1329–1355. http://doi.org/10.1163/156853905774539328

Carere, C., & Maestripieri, D. (2013). *Animal personalities*. Chicago, IL: The University of Chicago Press.

Charmantier, A., & Réale, D. (2005). How do misassigned paternities affect the estimation of heritability in the wild? *Molecular Ecology*, *14*(9), 2839–2850. http://doi.org/10.1111/j.1365-294X.2005.02619.x

Chervet, N., Zöttl, M., Schürch, R., Taborsky, M., & Heg, D. (2011). Repeatability and Heritability of Behavioural Types in a Social Cichlid. *International Journal of Evolutionary Biology*, *2011*, 1–15. http://doi.org/10.4061/2011/321729

Cote, J., Clobert, J., Brodin, T., Fogarty, S., & Sih, A. (2010). Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1560), 4065–4076. http://doi.org/10.1098/rstb.2010.0176

David, M., Auclair, Y., & Cézilly, F. (2012). Assessing short- and long-term repeatability and stability of personality in captive zebra finches using longitudinal data. *Ethology*, *118*(10), 932–942. http://doi.org/10.1111/j.1439-0310.2012.02085.x

de Villemereuil, P. (2018). Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences*, *1422*(1), 29–47. http://doi.org/10.1111/nyas.13571

de Villemereuil, P., Schielzeth, H., Nakagawa, S., & Morrissey, M. (2016). General methods for evolutionary quantitative genetic inference from generalised mixed models. *Genetics*, *204*(3), 1–30. http://doi.org/10.1534/genetics.115.186536

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B:*

Biological Sciences, *271*(1541), 847–852. http://doi.org/10.1098/rspb.2004.2680

Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, *64*(6), 929–938. http://doi.org/10.1006/anbe.2002.2006

Dingemanse, N. J., Bouwman, K. M., van de Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E., & Quinn, J. L. (2012). Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology*, *81*(1), 116–126. http://doi.org/10.1111/j.1365-2656.2011.01877.x

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. http://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. http://doi.org/10.1016/j.tree.2009.07.013

Dingemanse, N. J., & Réale, D. (2005). Natural selection and animal personality. *Behaviour*, *142*(9/10), 1165–1190. http://doi.org/10.1163/156853905774539445

Dingemanse, N. J., & Réale, D. (2013). What is the evidence that natural selection maintains variation in animal personalities? In C. Carere & D. Maestripieri (Eds.), *Animal personalities* (pp. 201–220). Chicago, IL: The University of Chicago Press.

Dingemanse, N. J., Wright, J., Kazem, A. J. N., Thomas, D. K., Hickling, R., & Dawnay, N. (2007). Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. *Journal of Animal Ecology*, *76*(6), 1128–1138. http://doi.org/10.1111/j.1365-2656.2007.01284.x

Dochtermann, N. A. (2011). Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution*, *65*(6), 1814–1820. http://doi.org/10.1111/j.1558-

5646.2011.01264.x

Dochtermann, N. A., Schwab, T., & Sih, A. (2015). The contribution of additive genetic variation to personality variation: heritability of personality. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798), 20142201. http://doi.org/10.1098/rspb.2014.2201

Doligez, B., & Clobert, J. (2003). Clutch size reduction as a response to increased nest predation rate in the collared flycatcher. *Ecology*, *84*(10), 2582–2588. http://doi.org/10.1890/02-3116

Doligez, B., Danchin, E., Clobert, J., & Gustafsson, L. (1999). The use of conspecific reproductive success for breeding habitat selection in a non-colonial, hole-nesting species, the collared flycatcher. *Journal of Animal Ecology*, *68*(6), 1193–1206. http://doi.org/10.1046/j.1365-2656.1999.00362.x

Doligez, B., Pärt, T., Danchin, E., Clobert, J., & Gustafsson, L. (2004). Availability and use of public information and conspecific density for settlement decisions in the collared flycatcher. *Journal of Animal Ecology*, *41*, 75–87.

Drent, P. J., Oers, K. v., & Noordwijk, A. J. v. (2003). Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society B: Biological Sciences*, *270*(1510), 45–51. http://doi.org/10.1098/rspb.2002.2168

Duckworth, R. A., & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(38), 15017–22. http://doi.org/10.1073/pnas.0706174104

Duckworth, R. A., & Kruuk, L. E. B. (2009). Evolution of genetic integration between dispersal and colonization ability in a bird. *Evolution*, *63*(4), 968–977. http://doi.org/10.1111/j.1558-5646.2009.00625.x

Eising, C. M., Muller, W., & Groothuis, T. G. . (2006). Avian mothers create different phenotypes by hormone deposition in their eggs. *Biology Letters*, *2*(1), 20–22. http://doi.org/10.1098/rsbl.2005.0391

Evans, S. R., & Gustafsson, L. (2017). Climate change upends selection on ornamentation in a wild bird. *Nature Ecology and Evolution*, *1*(2), 1–5. http://doi.org/10.1038/s41559-016-0039

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th editio). London, U.K.: Prentice Hall.

Firth, J. A., Hadfield, J. D., Santure, A. W., Slate, J., & Sheldon, B. C. (2015). The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution*, *69*(5), 1336–1344. http://doi.org/10.1111/evo.12649

Garamszegi, L. Z., Eens, M., & Török, J. (2009). Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. *Animal Behaviour*, *77*(4), 803–812. http://doi.org/10.1016/j.anbehav.2008.12.012

Garamszegi, L. Z., & Herczeg, G. (2012). Behavioural syndromes, syndrome deviation and the within- and between-individual components of phenotypic correlations: When reality does not meet statistics. *Behavioral Ecology and Sociobiology*, *66*(12), 1651–1658. http://doi.org/10.1007/s00265-012-1439-8

Garamszegi, L. Z., Markó, G., & Herczeg, G. (2012). A meta-analysis of correlated behaviours with implications for behavioural syndromes: Mean effect size, publication bias, phylogenetic effects and the role of mediator variables. *Evolutionary Ecology*, *26*(5), 1213–1235. http://doi.org/10.1007/s10682-012-9589-8

Garamszegi, L. Z., Markó, G., Szász, E., Zsebők, S., Azcárate, M., Herczeg, G., & Török, J. (2015). Among-year variation in the repeatability, within- and between-individual, and phenotypic correlations of behaviors in a natural population. *Behavioral Ecology and*

Sociobiology, *69*(12). http://doi.org/10.1007/s00265-015-2012-z

Garamszegi, L. Z., Rosivall, B., Rettenbacher, S., Markó, G., Zsebok, S., Szöllosi, E., … Török, J. (2012). Corticosterone, avoidance of novelty, risk-taking and aggression in a wild bird: no evidence for pleiotropic effects. *Ethology*, *118*(7), 621–635. http://doi.org/10.1111/j.1439-0310.2012.02049.x

Grafen, A. (1984). Natural selection, kin selection and group selection. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: an Evolutionary Approach* (pp. 62–84). Oxford, U.K.: Blackwell Science.

Greenberg, R., & Mettke-Hofmann, C. (2001). Ecological aspects of neophobia and neophilia in birds. *Current Ornithology, 16I,* 119–178. http://doi.org/10.1007/978-1-4615-1211-0

Groothuis, T. G. G., & Maestripieri, D. (2013). Parental influence on offspring personality traits in oviparous and placental vertebrates. In C. Carere & D. Maestripieri (Eds.), *Animal personalities: behaviour, physiology, and evolution* (pp. 317–352). Chicago, IL: The University of Chicago Press.

Gustafsson, L. (1987). Interspecific competition lowers fitness in collared flycatchers *Ficedula albicollis*: an experimental demonstration. *Ecology*, *68*(2), 291–296.

Gustafsson, L. (1989). Collared flycatcher. In: I. Newton (ed.): *Lifetime reproduction in birds* (pp. 75-88). London, U.K.: Academic Press.

Hadfield, J. D. (2010). MCMC methods for multi-response Generalized Linear Mixed Models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22. http://doi.org/10.1002/ana.22635

Hadfield, J. D. (2016). MCMCglmm course notes. https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf

Hadfield, J. D., Nutall, A., Osorio, D., & Owens, I. P. F. (2007). Testing the phenotypic gambit: Phenotypic, genetic and environmental correlations of colour. *Journal of*

Evolutionary Biology, *20*(2), 549–557. http://doi.org/10.1111/j.1420-9101.2006.01262.x

Hegyi, G., Rosivall, B., & Török, J. (2006). Paternal age and offspring growth: Separating the intrinsic quality of young from rearing effects. *Behavioral Ecology and Sociobiology*, *60*(5), 672–682. http://doi.org/10.1007/s00265-006-0211-3

Holtmann, B., Lagisz, M., & Nakagawa, S. (2017). Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: a meta-analysis. *Functional Ecology*, *31*(685–696). http://doi.org/10.1111/1365-2435.12779

Karlsson Green, K., Eroukhmanoff, F., Harris, S., Pettersson, L. B., & Svensson, E. I. (2016). Rapid changes in genetic architecture of behavioural syndromes following colonisation of a novel environment. *Journal of Evolutionary Biology*, *29*, 144–152. http://doi.org/10.1111/jeb.12769

Kivelä, S. M., Seppänen, J.-T., Ovaskainen, O., Doligez, B., Gustafsson, L., Mönkkönen, M., & Forsman, J. T. (2014). The past and the present in decision-making: the use of conspecific and heterospecific cues in nest site selection. *Ecology*, *95*, 3428–3439. http://doi.org/10.1890/13-2103.1

Koolhaas, J. ., Korte, S. ., De Boer, S. ., Van Der Vegt, B. ., Van Reenen, C. ., Hopster, H., … Blokhuis, H. . (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*(7), 925–935. http://doi.org/10.1016/S0149-7634(99)00026-3

Korsten, P., van Overveld, T., Adriaensen, F., & Matthysen, E. (2013). Genetic integration of local dispersal and exploratory behaviour in a wild bird. *Nature Communications*, *4*(2362). http://doi.org/10.1038/ncomms3362

Král, M., & Bícík, V. (1989). Intraspecific aggressive responses of male collared flycatchers (*Ficedula alcicollis* temm.). *Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Biologica XXIX*, *96*, 107–122.

Král, M., & Bícík, V. (1992). Nest defence by the collared flycatcher (*Ficedula albicollis*) against the great tit (*Parus major*). *Folia Zooogica*, *41*(3), 263–269.

Kruuk, L. E. B., & Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, *20*(5), 1890–1903. http://doi.org/10.1111/j.1420-9101.2007.01377.x

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18.

Merilä, J., & Gustafsson, L. (1993). Inheritance of size and shape in a natural population of collared flycatchers, *Ficedula albicollis*. *Journal of Evolutionary Biology*, *6*(3), 375–395. http://doi.org/10.1046/j.1420-9101.1993.6030375.x

Merilä, J., & Sheldon, B. C. (2000). Lifetime Reproductive Success and Heritability in Nature. *The American Naturalist*, *155*(3), 301–310. http://doi.org/10.1086/303330

Morinay, J., Forsman, J. T., Kivelä, S. M., Gustafsson, L., & Doligez, B. (2018). Heterospecific nest site copying behavior in a wild bird: Assessing the influence of genetics and past experience on a joint breeding phenotype. *Frontiers in Ecology and Evolution*, *5*(January), 1–14. http://doi.org/10.3389/fevo.2017.00167

Morrissey, M. B., & Wilson, A. J. (2010). Pedantics: an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology Resources*, *10*(4), 711–719. http://doi.org/10.1111/j.1755-0998.2009.02817.x

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*, 935–956. http://doi.org/10.1111/j.1469-185X.2010.00141.x

Niemelä, P. T., & Dingemanse, N. J. (2017). Individual versus pseudo-repeatability in behaviour: Lessons from translocation experiments in a wild insect. *Journal of Animal Ecology*, *86*(5), 1033–1043. http://doi.org/10.1111/1365-2656.12688

Niemelä, P. T., Lattenkamp, E. Z., & Dingemanse, N. J. (2015). Personality-related survival and sampling bias in wild cricket nymphs. *Behavioral Ecology*, *26*(3), 936–946. http://doi.org/10.1093/beheco/arv036

Petelle, M. B., Martin, J. G. A., & Blumstein, D. T. (2015). Heritability and genetic correlations of personality traits in a wild population of yellow-bellied marmots (*Marmota flaviventris*). *Journal of Evolutionary Biology*, *28*(10), 1840–1848. http://doi.org/10.1111/jeb.12700

Plummer, M., Best, N., Cowles, K., & Vines, K (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News, 6,* 7-11.

Qvarnström, A. (1997). Experimentally increased badge size increases male competition and reduces male parental care in the collared flycatcher. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1385), 1225–1231. http://doi.org/10.1098/rspb.1997.0169

R Core Team. (2016). R*: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. http://doi.org/10.1098/rstb.2010.0208

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. http://doi.org/10.1111/j.1469-185X.2007.00010.x

Rosivall, B., Török, J., Hasselquist, D., & Bensch, S. (2004). Brood sex ratio adjustment in collared flycatchers (*Ficedula albicollis*): results differ between populations. *Behavioral Ecology and Sociobiology*, *56*(4), 346–351. http://doi.org/10.1007/s00265-004-0796-3

Schuett, W., Tregenza, T., & Dall, S. R. X. (2010). Sexual selection and animal personality.

Biological Reviews, *85*, 217–246. http://doi.org/10.1111/j.1469-185X.2009.00101.x
Sheldon, B. C., & Ellegren, H. (1999). Sexual selection resulting from extrapair paternity in collared flycatchers. *Animal Behaviour*, *57*(2), 285–298. http://doi.org/10.1006/anbe.1998.0968
Sheldon, B. C., Kruuk, L. E. B., & Merilä, J. (2003). Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution*, *57*(2), 406–420. http://doi.org/10.1111/j.0014-3820.2003.tb00274.x
Sih, A., Bell, A. M., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. http://doi.org/10.1016/j.tree.2004.04.009
Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: an integrative overview. *The Quarterly Review of Biology*, *79*(3), 241–277. http://doi.org/10.1086/422893
Sih, A., Cote, J., Evans, M., Fogarty, S., & Pruitt, J. (2012). Ecological implications of behavioural syndromes. *Ecology Letters*, *15*(3), 278–289. http://doi.org/10.1111/j.1461-0248.2011.01731.x
Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: A meta-analysis. *Behavioral Ecology*, *19*(2), 448–455. http://doi.org/10.1093/beheco/arm144
Spencer, K. A., & Verhulst, S. (2007). Delayed behavioral effects of postnatal exposure to corticosterone in the zebra finch (*Taeniopygia guttata*). *Hormones and Behavior*, *51*(2), 273–280. http://doi.org/10.1016/j.yhbeh.2006.11.001
Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biological Reviews*, *85*(2), 301–325. http://doi.org/10.1111/j.1469-185X.2009.00103.x
Svensson, L. (1992). *Identification guide to european passerines* (Märstatryc). Stockholm,

Sweden: Svensson.

van Oers, K., De Jong, G., Drent, P. J., & van Noordwijk, A. J. (2004). A genetic analysis of avian personality traits: Correlated response to artificial selection. *Behavior Genetics*, *34*(6), 611–619. http://doi.org/10.1007/s10519-004-5588-z

van Oers, K., de Jong, G., van Noordwijk, A. J., & Drent, P. J. (2005). Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour*, *142*(9), 1185–1206. http://doi.org/10.1163/156853905774539364

van Oers, K., Drent, P. J., de Goede, P., & van Noordwijk, A. J. (2004). Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society B: Biological Sciences*, *271*, 65–73. http://doi.org/10.1098/rspb.2002.2168

van Oers, K., & Sinn, D. L. (2011). Towards a basis for the phenotypic gambit: advances in the evolutionary genetics of animal personality. In M. Inoue-Murayama, S. Kawamura, & A. Weiss (Eds.), *From genes to behavior* (pp. 165–183). Tokyo, Japan: Springer.

van Oers, K., & Sinn, D. L. (2013). Quantitative and molecular genetics of animal personality. In C. Carere & D. Maestripieri (Eds.), *Animal personalities: behaviour, physiology, and evolution* (pp. 149–200). Chicago, IL: The University of Chicago Press.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., … Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, *79*(1), 13–26. http://doi.org/10.1111/j.1365-2656.2009.01639.x

Wolak, M. E. (2012). nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*, *3*(5), 792–796. http://doi.org/10.1111/j.2041-210X.2012.00213.x

Wolf, J. B., Brodie III, E. D. B., Cheverud, J. M., Moore, A. J., & Wade, M. J. (1998). Indirect genetic effects. *Trends in Ecology & Evolution*, *13*(2), 64–69.

http://doi.org/10.1016/S0169-5347(97)01233-0

Wuerz, Y., & Krüger, O. (2015). Personality over ontogeny in zebra finches: Long-term repeatable traits but unstable behavioural syndromes. *Frontiers in Zoology*, *12*(1). http://doi.org/10.1186/1742-9994-12-S1-S9

Appendix

Aggressiveness score

We estimated aggressiveness score using the following alternative measures and modelled them using the parameters given in parentheses. (1) Number of aggressive behaviours (movements < 2 m from the nestbox, stationary flights and attacks towards the decoys, chases of live birds) standardized per 15 min (Poisson family; number of iterations = 10^6; burn-in = 10^4; thinning interval = 400). (2) First axis of the PCA presented below (Table A7; Gaussian family; number of iterations = 10^6; burn-in = 10^4; thinning interval = 400). (3) Discrete score (threshold family, residual variance V_R = 10; number of iterations = 13 x 10^5; burn-in = 8 x 10^4; thinning interval = 700). This score was based on the distinction between activity (number of movements/min, including stationary flights and chases) performed far from (> 2 m) and close to (< 2 m) the nestbox, and on attacks, subdivided into six categories (Fig. A3a): **0:** individuals that performed no movements (either far from or close to the nestbox); 1: individuals that performed no attack or movements close to the nestbox and performed less than 0.440 movements/min far from (> 5 m) the nestbox; 2: individuals that performed no attack or movements close to the nestbox and performed more than 0.440 movements/min far from (> 5 m) the nestbox; 3: individuals that performed no attacks and less than 0.282 movements/min close to (< 2 m) the nestbox; 4: individuals that performed no attacks and between 0.282 and 0.784 movements/min close to (< 2 m) the nestbox; 5: individuals that performed no attacks and above 0.784 movements/min close to (< 2 m) the nestbox; 6:

individuals that performed attacks towards decoys(s). The thresholds were chosen so as to distribute individuals equally among categories for scores 1 and 2 on the one hand and scores 3, 4 and 5 on the other.

Boldness score

We estimated boldness score using the following alternative measures and modelled them using the parameters given in parentheses. (1) Maximum latency to enter the nestbox after human departure (from all individuals) minus the same latency for the focal individual on the focal test (Poisson family; number of iterations = 1 x 10^6; burn-in = 10^4; thinning interval = 500). (2) Inverse logarithmic ratio of the latency to enter the nestbox after human departure divided by the feeding rate for the remaining time after the first entrance in the nestbox. We divided by the feeding rate for this alternative measurement because the average feeding interval might have affected the latency to return to the nestbox, for example if individuals feeding more frequently entered the nestbox faster. The feeding rate after the first entrance was estimated as the average time interval between two feeding events by the focal birds, after the first entrance; it was thus only computable for individuals that fed at least twice in the period; for technical reasons, further individuals could not be used for this variable, leading to a final used data set of 641 observations (Gaussian family; number of iterations = 25 x 10^4; burn-in = 10^4; thinning interval = 100). (3) Discrete score (threshold family, V_R = 10; number of iterations = 3 x 10^6; burn-in = 2 x 10^5; thinning interval = 1000). The score was based on entrance in the nestbox during the first part of the boldness-neophobia test (no novel object) and latency to enter after human disturbance; individuals that did not enter were given a score of 0, and individuals that entered the nestbox were given a score of 1 - 5 based on five quantiles of the inverse latency to enter (Fig. A3b).

Neophobia score

We estimated neophobia score for individuals that entered the nestbox during the first period of the test, using the following alternative measures and modelled them using the parameters given in parentheses. (1) Discrete score based on the latency to enter the nestbox in the presence of a novel object (second period of the test), discretized in four quantiles, the fifth category including individuals that did not enter the nestbox at all in the presence of the novel object (threshold family; V_R = 10; number of iterations = 15 x 10^5; burn-in = 2 x 10^5; thinning interval = 500; Fig. A3c); (2) Binary variable separating individuals that did and did not enter during the second period of the test (threshold family; V_R = 10; number of iterations = 10^6; burn-in = 10^5; thinning interval = 500). (3) Latency to enter the nestbox in the presence of the novel object, excluding the individuals that did not enter the nestbox during the second period of the test (Poisson family; number of iterations= 15 x 10^5; burn-in = 15 x 10^4; thinning interval = 500).

Results obtained for these alternative scores for the three behavioural traits are given in Tables A2 to A4, and the main text presents the first score in each case.

TABLES

Table 1. Between-year repeatability and heritability estimates for aggressiveness, boldness and neophobia scores

	R_{latent}	R_{obs}	h^2_{latent}	h^2_{obs}
Aggressiveness	0.18 *	0.03 *	0.00	0.00
	[0.15; 0.23]	[0.02; 0.04]	[0.00; 0.08]	[0.00; 0.008]
Boldness	0.11 *	0.10 *	0.00	0.00
	[0.01; 0.21]	[0.01; 0.19]	[0.00; 0.10]	[0.00; 0.09]
Neophobia	0.39 *		0.00	0.00 for all scores
	[0.25; 0.54]		[0.00; 0.15]	From [0.00; 0.00] to [0.00; 0.09]

Repeatabilities and heritabilities (posterior modes and 95% credible intervals) are given on the latent scale (R_{latent}, h^2_{latent}) and on the observed scale (R_{obs} obtained with the QGicc function and h^2_{obs} with the QGparams function from 'QGglmm' R package; de Villemereuil et al. 2016; de Villemereuil 2018). Asterisks indicate estimates whose 95% CI do not encompass zero. For aggressiveness, estimates are given using all scores. For neophobia, we provide a range of heritability values on the observed scale, since one value is provided per neophobia score level (i.e. five values in total); however, we could not derive repeatability estimates on the observed scale for ordinal variables.

Table 2. Phenotypic, between- and within-individual, and additive genetic correlations between aggressiveness, boldness and neophobia scores

Correlation level	Aggressiveness - Neophobia	Boldness - Neophobia	Aggressiveness - Boldness
Phenotypic	-0.20*	-0.30*	0.02
	[-0.25; -0.12]	[-0.38; -0.23]	[-0.04; 0.10]
Between-individual	0.00	0.00	-0.02
	[-0.06; 0.06]	[-0.07; 0.05]	[-0.06; 0.06]
Within-individual	-0.28*	-0.43*	0.03
	[-0.38; -0.18]	[-0.53; -0.32]	[-0.04; 0.12]
Additive genetic	0.00	-0.01	0.01
	[-0.06; 0.006]	[-0.06; 0.06]	[-0.07; 0.06]

Posterior modes and 95% credible intervals (CI) on the latent scale are shown. Asterisks indicate estimates whose 95% CI do not encompass zero.

Table A1. Detailed description of the collared flycatcher pedigree from the Gotland Island population

Pedigree statistics	
Records	2 218
Founders	1423
Maternities	728
Paternities	781
Mothers with ≥ 2 offspring	120
Fathers with ≥ 2 offspring	133
Full sibs	130
Maternal sibs	222
Maternal half sibs	92
Paternal sibs	230
Paternal half sibs	100
Maternal grandmothers	249
Maternal grandfathers	268
Paternal grandmothers	220
Paternal grandfathers	236
Maximum pedigree depth	15
Mean relatedness	5.29×10^{-4}

The pedigree statistics were obtained from all identified individuals involved in either aggressiveness, boldness or neophobia assays, and were extracted using the pedigreeStats and pedStatSummary functions from ‘pedantics’ R package (Morrissey & Wilson, 2010).

Table A2. Output of the univariate models fitting aggressiveness score

	Number of aggressive behaviours/15 min						PC1			D
	Without fixed effect			With fixed effects						
	Posterior mode	95% CI		Posterior mode	95% CI		Posterior mode	95% CI		P n
Fixed effects										
Intercept	1.22	[0.91 ; 1.50]	*	-0.29	[-0.72 ; 0.26]		-0.98	[-1.30 ; -0.65]	*	5
Sex (male)				0.63	[0.44 ; 0.81]	*	0.50	[0.35 ; 0.62]	*	1
Age (young)				0.23	[-0.01 ; 0.40]		0.06	[-0.05 ; 0.25]		0
Presence of tits (present)				0.34	[0.19 ; 0.50]	*	0.05	[-0.07 ; 0.15]		0
Presence of flycatchers (present)				0.27	[-0.02 ; 0.55]		0.29	[0.05 ; 0.49]	*	0
Presence of the partner (present)				0.62	[0.44 ; 0.80]	*	0.34	[0.20 ; 0.44]	*	1
Test number				-0.20	[-0.27 ; -0.14]	*	-0.16	[- ; -]	*	-0

								0.20 ; 0.11		
Dummy type (flycatcher)				0.08	[-0.23 ; 0.34]		0.16	[-0.01 ; 0.38]		0
Year (2012)				0.48	[0.08 ; 0.87]	*	0.19	[-0.05 ; 0.50]		0
Year (2013)				0.29	[-0.09 ; 0.67]		0.10	[-0.09 ; 0.46]		0
Sex*age (male*young)				0.28	[0.03 ; 0.61]	*	0.24	[0.09 ; 0.50]	*	0
Random effects										
V_A	0.00	[0.00 ; 0.32]		0.00	[0.00 ; 0.41]		0.00	[0.00 ; 0.16]		0
V_{PE}	0.60	[0.28 ; 0.80]	*	0.52	[0.11 ; 0.71]	*	0.19	[0.03 ; 0.28]	*	1
V_M	0.00	[0.00 ; 0.20]		0.00	[0.00 ; 0.17]		0.00	[0.00 ; 0.12]		0
V_F	0.00	[0.00 ; 0.20]		0.00	[0.00 ; 0.24]		0.23	[0.06 ; 0.37]	*	0
V_{plot}	0.07	[0.02 ; 0.16]	*	0.05	[0.01 ; 0.15]	*	0.01	[0.00 ; 0.04]		0
V_{decoy}	0.00	[0.00 ; 0.03]		0.00	[0.00 ; 0.04]		0.00	[0.00 ; 0.03]		0
V_{song}	0.02	[0.00 ; 0.09]		0.01	[0.00 ; 0.09]		0.00	[0.00 ; 0.03]		0

V_{obs}	0.25	[0.14 ; 0.55]	*	0.30	[0.15 ; 0.63]	*	0.06	[0.02 ; 0.16]	*	0
$V\varepsilon$	2.56	[0.91 ; 1.50]	*	2.11	[1.97 ; 2.34]	*	1.49	[1.40 ; 1.57]	*	1
Derived estimates										
R_{latent}	0.18	[0.15 ; 0.23]	*	0.18	[0.13 ; 0.22]	*	0.23	[0.17 ; 0.27]	*	0
R_{obs}	0.03	[0.02 ; 0.04]	*	0.03	[0.02 ; 0.04]	*	0.21	[0.16 ; 0.25]	*	
h^2_{latent}	0.00	[0.00 ; 0.08]		0.00	[0.00 ; 0.12]		0.00	[0.00 ; 0.08]		0
h^2_{obs}	0.00	[0.00 ; 0.01]		0.00	[0.00 ; 0.01]		0.00	[0.00 ; 0.07]		P
N		4680			3271			3271		
Effective sample size		> 2357			> 2234			> 2209		

Models for the general aggressiveness score, based on the number of aggressive behaviours standardized per 15 min, are shown without and with fixed effects. We also present the posterior modes and 95% credible intervals (CI) of models for alternative aggressiveness scores: the first axis of the principal component analysis (PC1) and the discrete score (see Appendix). Asterisks indicate estimates whose 95% CI do not encompass zero. For categorical fixed terms, estimates refer to the category indicated in parentheses. V_A, V_{PE}, V_M, V_F and V_ε refer to the additive genetic, permanent environment, maternal, paternal and residual variances, respectively. V_{plot}, V_{decoy}, V_{song}, V_{obs} refer to the variances associated with the plot, the decoy set used, the song track played and observer identity, respectively. N is the sample size of the data set used in the model.

Repeatability and heritability estimates are given both on the latent scale (R_{latent}, h^2_{latent}) and on the observed scale (R_{obs}, h^2_{obs}) whenever these could be estimated.

Table A3 Output of the models fitting boldness scores.

	Maximum latency–individual latency						Latency/feeding rate			Discrete sc
	Without fixed effects			Without fixed effects						
	Posterior mode	95% CI		Posterior mode	95% CI		Posterior mode	95% CI		Posterior mode
Fixed effects										
Intercept	8.00	[7.97 ; 8.04]	*	8.05	[7.96 ; 8.10]	*	-3.78	[-4.07 ; -3.46]	*	4.80
Sex (male)				-0.04	[-0.12 ; 0.02]		-0.10	[-0.29 ; 0.13]		-1.33
Age (young)				-0.02	[-0.10 ; 0.05]		-0.12	[-0.39 ; 0.13]		-0.65
Year (2012)				-0.03	[-0.10 ; 0.04]		0.05	[-0.19 ; 0.43]		-0.38
Year (2013)				0.08	[0.00 ; 0.16]	*	0.47	[0.10 ; 0.73]	*	1.46
No. of chicks				0.03	[0.00 ; 0.06]	*	0.26	[0.15 ; 0.33]	*	0.36
Sex*age (male*young)				0.03	[-0.08 ; 0.13]		0.30	[-0.18 ; 0.55]		0.60
Random effects										
V_A	0.00	[0.00 ; 0.02]		0.00	[0.00 ; 0.03]		0.00	[0.00 ; 0.20]		0.01

V_{PE}	0.00	[0.00 ; 0.02]		0.00	[0.00 ; 0.03]		0.00	[0.00 ; 0.25]		0.04
V_{M}	0.00	[0.00 ; 0.02]		0.00	[0.00 ; 0.02]		0.00	[0.00 ; 0.13]		0.01
V_{F}	0.00	[0.00 ; 0.01]		0.00	[0.00 ; 0.01]		0.00	[0.00 ; 0.15]		0.01
V_{plot}	0.00	[0.00 ; 0.00]		0.00	[0.00 ; 0.01]		0.00	[0.00 ; 0.07]		0.00
V_{obs}	0.00	[0.00 ; 0.00]		0.00	[0.00 ; 0.00]		0.00	[0.00 ; 0.03]		0.00
$V\varepsilon$	0.16	[0.14 ; 0.17]	*	0.13	[0.11 ; 0.16]	*	1.20	[0.93 ; 1.37]	*	10.00
Derived estimates										
R_{latent}	0.11	[0.01 ; 0.21]	*	0.19	[0.07 ; 0.31]	*	0.12	[0.01 ; 0.29]	*	0.28
R_{obs}	0.10	[0.01 ; 0.19]	*	0.12	[0.02 ; 0.23]	*	0.04	[0.01 ; 0.14]	*	
h^2_{latent}	0.00	[0.00 ; 0.10]		0.00	[0.00 ; 0.18]		0.00	[0.00 ; 0.14]		0.00
h^2_{obs}	0.00	[0.00 ; 0.09]		0.00	[0.00 ; 0.09]		0.00	[0.00 ; 0.06]		Posterior m
N	1064			914			641			
Effective sample size	> 1718			> 1979			> 2053			

Models for the general boldness estimate (maximum latency to enter the nestbox after human disturbance observed in the entire data set minus the individual latency) are shown without and with fixed effects. We also present the posterior modes and 95% credible intervals (CI) of models for alternative boldness estimates: the log-transformed and inverse ratio of the latency to enter the nestbox after human disturbance over the feeding rate during the time remaining and the discrete score based on the latency (see Appendix). Asterisks indicate estimates whose 95% CI do not encompass zero. For categorical fixed terms, estimates refer to the category indicated in parentheses. V_{A}, V_{PE}, V_{M}, V_{F} and V_{ε} refer to the additive genetic, permanent environment, maternal, paternal and residual variances, respectively. V_{plot} and V_{obs} refer to the variances associated with the plot and the observer identity, respectively. *N* is the sample size of the data set used in the model. Repeatability and heritability estimates are given both on the latent scale (R_{latent}, h^2_{latent}) and on the observed scale (R_{obs}, h^2_{obs}) whenever these could be estimated. Repeatability and heritability estimates for the latency/feeding rate model (fitted with a Gaussian distribution) differ between the latent and observed scales because we accounted for the inverse log transformation in the QGicc and QGparams functions ('QGglmm' R pachage; de Villemereuil et al. 2016).

Table A4. Output of the models fitting neophobia scores

	Discrete score				Binary		La
	Without fixed effect		With fixed effects				
	Posterior mode	95% CI	Posterior mode	95% CI	Posterior mode	95% CI	Po m
Fixed effects							
Intercept	4.22	[3.40 ; 4.94] *	5.08	[3.63 ; 6.38] *	-1.67	[-3.39 ; -0.30] *	6.
Sex (male)			-1.89	[-2.78 ; -1.07] *	-1.45	[-2.44 ; -0.41] *	-0
Age (young)			-0.18	[-1.01 ; 0.76]	0.07	[-1.26 ; 0.92]	-0
Year (2012)			1.62	[0.72 ; 2.93] *	2.19	[0.59 ; 3.85] *	0.
Year (2013)			1.21	[-0.05 ; 2.37]	1.84	[-0.06 ; 3.34]	0.
Boldness			-1.58	[-1.93 ; -1.12] *	-1.34	[-1.95 ; -0.93] *	-0

No. of chicks				-0.71	[-1.03 ; -0.37]	*	-0.71	[-1.15 ; -0.30]	*	-0
Sex*age (male*young)				-1.02	[-2.44 ; 0.11]		-0.79	[-2.75 ; 0.69]		-0
Random effects										
V_{A}	0.02	[0.00 ; 2.90]		0.03	[0.00 ; 5.76]		0.08	[0.00 ; 8.41]		0.
V_{PE}	0.04	[0.00 ; 5.77]		0.06	[0.00 ; 9.03]		0.05	[0.00 ; 10.70]		0.
V_{M}	0.04	[0.00 ; 6.49]		0.06	[0.00 ; 8.81]		0.12	[0.00 ; 22.53]		0.
V_{F}	0.02	[0.00 ; 5.75]		0.03	[0.00 ; 7.85]		0.04	[0.00 ; 11.33]		0.
V_{plot}	0.98	[0.26 ; 2.50]	*	1.01	[0.27 ; 3.42]	*	0.80	[0.00 ; 3.08]		0.
V_{obs}	0.28	[0.02 ; 1.01]	*	0.38	[0.00 ; 1.41]		0.80	[0.14 ; 3.14]	*	0.
$V\varepsilon$	10.00	[10.00 ; 10.00]		10.00	[10.00 ; 10.00]		10.00	[10.00 ; 10.00]		0.
Derived estimates										
R_{latent}	0.39	[0.25 ; 0.54]	*	0.50	[0.33 ; 0.63]	*	0.56	[0.35 ; 0.75]	*	0.
R_{obs}							0.29	[0.17 ; 0.48]	*	0.
h^2_{latent}	0.00	[0.00 ; 0.15]		0.00	[0.00 ; 0.22]		0.00	[0.00 ; 0.26]		0.

h^2_{obs}	Posterior mode range: 0.00-0.00	Posterior mode range: 0.00-0.00	0.00	[0.00 ; 0.14]	0.
N	1 064	914	914		
Effective sample size	> 2302	> 2317	> 1583		

Models for the general neophobia discrete score, based on the latency to enter the nestbox in presence of the novel object, are shown without and with fixed effects We also present the posterior modes and 95% credible intervals (CI) of models for alternative neophobia estimates: a binary variable (individual entered versus did not enter in the presence of the novel object) and a continuous latency to enter the nestbox, for individuals that entered in the presence of the novel object (see Appendix). Asterisks indicate estimates whose 95% CI do not encompass zero. For categorical fixed terms, estimates refer to the category indicated in parentheses. V_A, V_{PE}, V_M and V_F refer to the additive genetic, permanent environment, maternal and paternal variances, respectively. The residual variance V_ε was set to 10. V_{plot} and V_{obs} refer to the variances associated with the plot and the observer identity, respectively. N is the sample size of the data set used in the model. Repeatability and heritability estimates are given both on the latent scale (R_{latent}, h^2_{latent}) and the observed scale (R_{obs}, h^2_{obs}) whenever these could be estimated.

Table A5. Output of the trivariate model fitting aggressiveness, boldness and neophobia scores with an individual random effect

	Posterior mode	95% CI	
Fixed effects			
Aggressiveness	1.73	[1.65 ; 1.82]	*
Boldness	7.97	[7.93 ; 8.02]	*
Neophobia	4.44	[3.79 ; 5.00]	*
Sex (male)	-0.02	[-0.06 ; 0.04]	
Year (2012)	0.04	[-0.02 ; 0.09]	
Year (2013)	0.09	[0.03 ; 0.15]	*
Random effects			
$V_{\text{ind, aggressiveness}}$	0.42	[0.16 ; 0.66]	*
$V_{\text{ind, boldness}}$	0.02	[0.00 ; 0.04]	
$V_{\text{ind, neophobia}}$	6.14	[2.96 ; 10.27]	*
$V_{\varepsilon\text{, aggressiveness}}$	1.36	[1.09 ; 1.62]	*

$V_{\varepsilon,\ \mathrm{boldness}}$	0.13	[0.12 ; 0.16]	*
$V_{\varepsilon,\ \mathrm{neophobia}}$	10.00	[10.00 ; 10.00]	
$\mathrm{Cov}_{\mathrm{ind,\ aggressiveness\text{-}neophobia}}$	0.00	[-0.09 ; 0.11]	
$\mathrm{Cov}_{\mathrm{ind,\ aggressiveness\text{-}boldness}}$	0.00	[-0.01 ; 0.01]	
$\mathrm{Cov}_{\mathrm{ind,\ boldness\text{-}neophobia}}$	0.00	[-0.03 ; 0.02]	
$\mathrm{Cov}_{\varepsilon,\ \mathrm{aggressiveness\text{-}neophobia}}$	-1.03	[-1.37 ; -0.63]	*
$\mathrm{Cov}_{\varepsilon,\ \mathrm{aggressiveness\text{-}boldness}}$	0.02	[-0.02 ; 0.05]	
$\mathrm{Cov}_{\varepsilon,\ \mathrm{boldness\text{-}neophobia}}$	-0.49	[-0.62 ; -0.36]	*
N		1689	
Effective sample size:		>1740	

V stands for variance terms and Cov for covariance terms (posterior mode and 95% credible interval (CI)). The individual effect was the only random term included; ‘ind’ and ‘ε’ stand for between-individual and residual terms, respectively. The residual variance for neophobia score was fixed to 10 (see text for the distributions used for the three scores). Asterisks indicate estimates whose 95% CI do not encompass zero.

Table A6. Output of the trivariate model fitting aggressiveness, boldness and neophobia scores with an additive genetic effect

	Posterior mode	95% CI	
Fixed effects			
Aggressiveness	1.71	[1.62 ; 1.79]	*
Boldness	7.97	[7.93 ; 8.02]	*
Neophobia	4.12	[3.55 ; 4.62]	*
Sex (male)	-0.01	[-0.05 ; 0.04]	
Year (2012)	0.04	[-0.02 ; 0.09]	
Year(2013)	0.09	[0.03 ; 0.15]	*
Random effects			
$V_{\text{A, aggressiveness}}$	0.15	[0.00 ; 0.38]	
$V_{\text{A, boldness}}$	0.02	[0.00 ; 0.04]	
$V_{\text{A, neophobia}}$	4.81	[2.07 ; 7.52]	*
$V_{\varepsilon\text{, aggressiveness}}$	1.46	[1.29 ; 1.78]	*

$V_{\varepsilon,\ \text{boldness}}$	0.14	[0.12 ; 0.16]	*
$V_{\varepsilon,\ \text{neophobia}}$	10.00	[10.00 ; 10.00]	
$\text{Cov}_{\text{A, aggressiveness-neophobia}}$	0.00	[-0.06 ; 0.07]	
$\text{Cov}_{\text{A, aggressiveness-boldness}}$	0.00	[0.00 ; 0.00]	
$\text{Cov}_{\text{A, boldness-neophobia}}$	0.00	[-0.02 ; 0.02]	
$\text{Cov}_{\varepsilon,\ \text{aggressiveness-neophobia}}$	-1.01	[-1.35 ; -0.64]	*
$\text{Cov}_{\varepsilon,\ \text{aggressiveness-boldness}}$	0.01	[-0.02 ; 0.05]	
$\text{Cov}_{\varepsilon,\ \text{boldness-neophobia}}$	-0.48	[-0.61 ; -0.37]	*
N		1686	
Effective sample size		> 1816	

V stands for variance terms and Cov for covariance terms (posterior mode and 95% credible interval (CI)). The additive genetic effect was the only random term included; ‘A’ and ‘ε’ stand for additive genetic and residual terms, respectively. The residual variance for neophobia score was fixed to 10 (see text for the distributions used for the three scores). Asterisks indicate estimates whose 95% CI do not encompass zero. Positive additive genetic variances are found here because permanent environment and parental effects are not taken into account.

Table A7. Output of a principal component analysis of behaviours recorded during the aggressiveness assays

	Coordinates		Contribution	
	PC1	PC2	PC1	PC2
Moves < 2 m from the nestbox	0.83	-0.12	36.12	0.95
Stationary flights towards the decoy	0.73	-0.21	27.74	2.99
Attacks towards the decoy	0.67	-0.24	23.24	4.06
Moves between 2 and 5 m from the nestbox	0.42	0.65	9.40	29.96
Moves between 5 and 10 m from the nestbox	-0.03	0.74	0.03	38.49
Chases of live birds	0.26	0.58	3.47	23.55

We used the function PCA from the 'FactoMineR' R package (Lê, Josse, & Husson, 2008).

FIGURE CAPTIONS

Figure A1. Sex differences in aggressiveness, boldness and neophobia scores (means ± 95% confidence interval). (a) Between-sexes differences in average aggressiveness score for a given individual in a given year depending on age (yearling versus older). (b) Between-sexes differences in the latency to return after human disturbance for the period without a novel object, as a proxy of (inverse) boldness, and for the period with the novel object, as a proxy of neophobia. Number of observations is indicated near each estimate.

Figure A2. (a) Aggressiveness and (b) boldness scores depending on the neophobia score (means ± 95% confidence interval). See text for the definitions of the scores. Aggressiveness is here the averaged value of all scores for a given individual in a given year. Number of observations is indicated near each estimate.

Figure A3. Distribution of the discrete scores for (a) aggressiveness, (b) boldness and (c) neophobia. Number of observations is indicated above each bar.

Figure A1

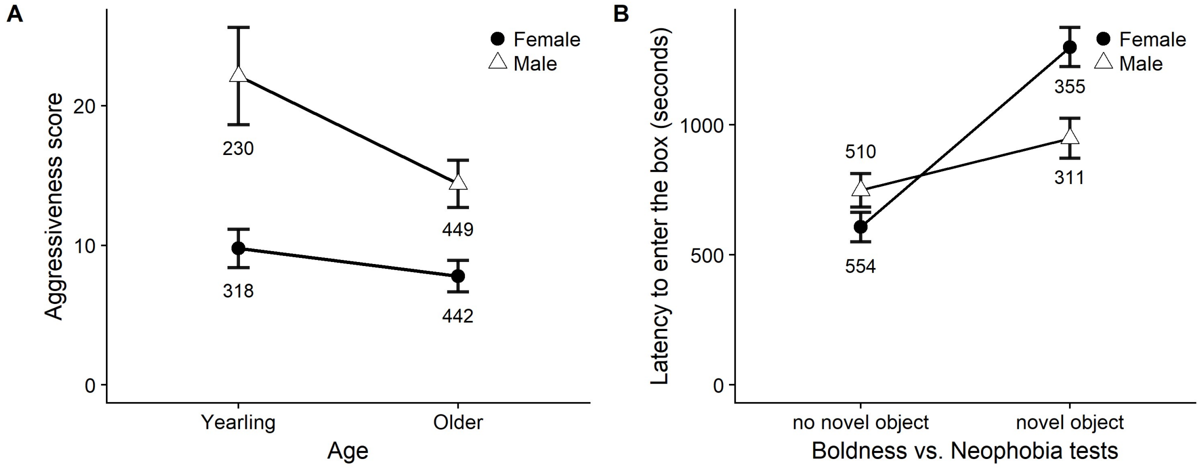
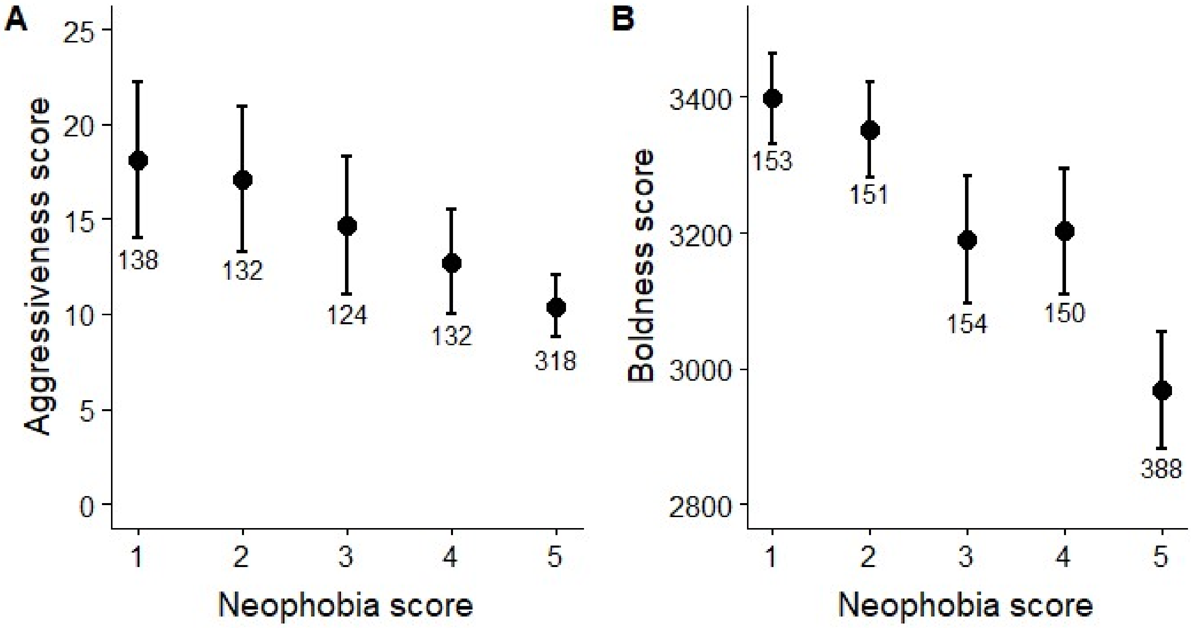
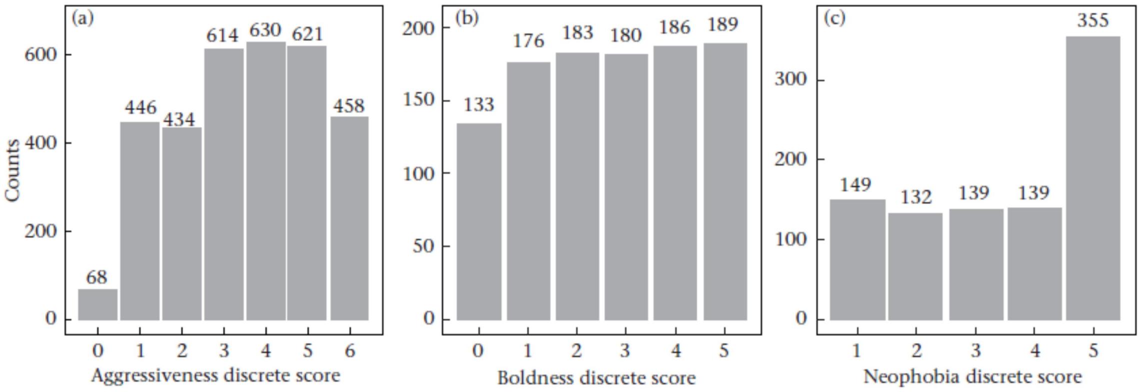

A
Aggressiveness score
0
10
20
230
449
318
442
Yearling
Older
Age
Female
Male
B
Latency to enter the box (seconds)
0
500
1000
510
554
355
311
no novel object
novel object
Boldness vs. Neophobia tests
Female
Male

Figure A2



A
B
25
20
15
10
5
0
Aggressiveness score
138
132
124
132
318
1
2
3
4
5
Neophobia score
3400
3200
3000
2800
Boldness score
153
151
154
150
388
1
2
3
4
5
Neophobia score

Figure A3


(a)
(b)
(c)
Counts
68
446
434
614
630
621
458
133
176
183
180
186
189
149
132
139
139
355
Aggressiveness discrete score
Boldness discrete score
Neophobia discrete score