Received: 22 December 2022 | Accepted: 23 May 2023

DOI: 10.1111/1365-2656.13966

Journal of Animal Ecology BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Are animal personality, body condition, physiology and structural size integrated? A comparison of species, populations and sexes, and the value of study replication

Irene Gaona-Gordillo[1] | **Benedikt Holtmann**[1,2] | **Alexia Mouchet**[1,3] | **Alexander Hutfluss**[1,4] | **Alfredo Sánchez-Tójar**[5] | **Niels J. Dingemanse**[1]

[1]Behavioural Ecology, Department of Biology, Ludwig Maximilians University of Munich, Planegg-Martinsried, Germany

[2]Federal Maritime and Hydrographic Agency, Department Management of the Sea, Division Assessment and Monitoring, Section Environmental Assessment, Rostock, Germany

[3]IDEEV UMR EGCE, IRD, CNRS, Paris-Saclay University, Gif-sur-Yvette, France

[4]Bayerisches Landesamt für Umwelt, Ref. 51 Fachgrundlagen Naturschutz, Augsburg, Germany

[5]Department of Evolutionary Biology, Bielefeld University, Bielefeld, Germany

Correspondence

Irene Gaona-Gordillo

Email: gaona-gordillo@biologie.uni-muenchen.de

Niels J. Dingemanse

Email: n.dingemanse@lmu.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: DI 1694/1-1, DI 1694/1-2 and HO 6288/1-1; Deutscher Akademischer Austauschdienst, Grant/Award Number: 91731707

Handling Editor: Allert Bijleveld

Abstract

1. A major question in behavioural ecology is why behaviour, physiology and morphology are often integrated into syndromes. In great tits, *Parus major*, for example, explorative males are larger (vs. smaller) and leaner (vs. heavier) compared to less explorative individuals. Unfortunately, considerable debate exists on whether patterns found in specific studies are replicable. This debate calls for study replication among species, populations and sexes.
2. We measured behavioural (exploration), physiological (breathing rate) and morphological traits (body mass, tarsus length, wing length, bill length) in two species (great vs. blue tits *Cyanistes caeruleus*), two populations (Forstenrieder Park vs. Starnberg) and two sexes (males vs. females). We then tested whether the same pattern of integration characterized all unique combinations of these three biological categories (hereafter called datasets).
3. We used a multi-year repeated measures set-up to estimate among-individual trait correlation matrices for each dataset. We then used structural equation modelling to test for size-dependent behaviour and physiology, size-corrected (i.e. size-independent) behaviour-physiology correlations and size-corrected body mass-dependent behaviour and physiology. Finally, we used meta-analyses to test which structural paths were generally (vs. conditionally) supported (vs. unsupported).
4. We found general and consistent support for size-dependent physiology and size-corrected body mass-dependent physiology across datasets: faster breathers were smaller but heavier for their size. Unexpectedly, condition-dependent behaviour was not supported: explorative birds were neither leaner, nor was this relationship heterogeneous across datasets. All other hypothesized patterns were dataset-specific: the covariance between size and behaviour, and between behaviour and physiology differed in sign between datasets, and both were, on average, not supported. This heterogeneity was not explained by any of our moderators: species, population or sex.

---

5. The specific pattern of size- and condition-dependent physiology reported for a unique combination of species, population, and sex, thus generally predicted those in others. Patterns of size- or condition-dependent behaviour (i.e. 'personality'), or behaviour-physiology syndromes reported in specific datasets, by contrast, did not. These findings call for studies revealing the ecological background of this variation and highlight the value of study replication to help understand whether patterns of phenotypic integration reported in one study can be generalized.

KEYWORDS
animal personality, phenotypic integration, physiology, meta-analysis, state-dependent behaviour, study replication

1 | INTRODUCTION

Natural selection often favours particular combinations of phenotypic traits (Brodie et al., 1995; Kingsolver et al., 2001; Lande & Arnold, 1983). Such correlational forms of selection can ultimately result in the evolution of genetic correlations (Roff & Fairbairn, 2012; Sinervo & Svensson, 2002). In garter snakes, *Thamnophis ordinoides*, for example, colour pattern and anti-predator behaviour are genetically correlated because individuals with different colour patterns can best avoid predation by also exhibiting a different behaviour when under attack (Brodie, 1992, 1993). Correlational selection thereby may produce optimal trait combinations that organisms require for executing particular functions (Klingenberg, 2008, 2014; Pigliucci & Preston, 2004). The level of predation risk, intensity of competition for resources or other ecological conditions specific to a species, population or sex will thus ultimately shape the patterns of selection observed in nature (Kingsolver et al., 2012; Siepielski et al., 2009). Species-, population- or sex-specific differences in ecology are therefore expected to result in species-, population- or sex-specific patterns of phenotypic trait integration (Armbruster et al., 2014; Armbruster & Schwaegerle, 1996; Roff & Fairbairn, 2012).

Comparing patterns of phenotypic integration among species, populations and sexes is of general importance if we are to understand whether patterns observed in specific studies are replicable versus study-specific (Wilson, 1998). In behavioural ecology, for example, there is growing awareness that conclusions regarding the replicability of findings can only be indirectly assessed through the study of publication bias (e.g. Jennions & Møller, 2002; Yang et al., 2023) due to lack of study replication, specially 'exact replication' (Filazzola & Cahill, 2021; Kelly, 2006, 2019; Nakagawa & Parker, 2015). In this paper, we take up this challenge. Our objective is to understand whether results are study- (or dataset-) specific versus replicable across studies (or datasets). We focus on complex patterns of behavioural, physiological and morphological integration that have previously been predicted by adaptive animal personality theory (Dingemanse & Wolf, 2010; Sih et al., 2015; Wolf & Weissing, 2010) and subsequently demonstrated empirically for a specific sex (males) of a specific bird species (the great tit *Parus major*) of a specific population (Starnberg, Bavaria, Germany) (Moiron et al., 2019). We investigate here whether the same structure of phenotypic integration characterizes both sexes (females and males) within the same population of great tits, and whether any patterns of (sex-specific) phenotypic integration also characterize the same and another species (the blue tit *Cyanistes caeruleus*) studied in another population (the Forstenrieder Park, Munich, Bavaria, Germany). We note that we did not aim to test for effects of a priori hypothesized ecological differences between species, population, and/or sexes; this would require many replicates of populations/species differing in ecology (Kelly, 2006). We instead aimed to use any observed differences in phenotypic integration as an opportunity to generate ecological hypotheses to be tested by future research.

A major question is why individuals differ in behaviour and why behavioural traits are often (genetically) correlated with other phenotypic traits (Dall et al., 2004; Wolf & Weissing, 2010). This question is of importance as (behavioural) genetic correlations can impose major constraints on micro-evolutionary processes (Dochtermann & Dingemanse, 2013). Our replication study, in part, addresses the role of behaviour in phenotypic integration. Particularly, we focused on a risk-taking behaviour (Dall et al., 2004; Réale et al., 2007; Sih et al., 2004; Wolf & Weissing, 2010). Risk-taking behaviours are considered as those that facilitate resource acquisition at the expense of increased risk of mortality, predation, or parasitism (Barber & Dingemanse, 2010; Stamps, 2007; Wolf et al., 2007). Examples are aggressiveness, anti-predator boldness, neophilia and speed of exploration of novel environments and objects. Meta-analyses have revealed that populations generally exhibit repeatable individual variation in such behaviours (Bell et al., 2009; Holtmann et al., 2017), that different types of risk-taking behaviours are positively correlated among individuals (Garamszegi et al., 2012), and that they are underpinned by tight genetic correlations (Dochtermann, 2011).

Behavioural ecologists have thus developed a suite of models seeking to explain when repeatable among-individual variance (or 'animal personality') in risk-taking behaviour may evolve (reviewed by Dingemanse & Wolf, 2010; Wolf & Weissing, 2010). Early explanations centred on selection favouring alternative life-histories associated with their specific behavioural adaptations, particularly in the context of pace-of-life: risk-takers would live fast but die young (Réale et al., 2010; Stamps, 2007; Wolf et al., 2007). Contemporary explanations reserve a more central role for ecological variation (Dammhahn et al., 2018; Montiglio et al., 2018; Mouchet et al., 2021; Wright et al., 2019). Fluctuating selection induced by ecological variability has come to the foreground, in part, because the mere existence of life-history trade-offs cannot explain the maintenance of variation (Stearns, 1992). A specifically appealing explanation is that fast (vs. slow) life-histories are favoured when a focal population is below (vs. at) carrying capacity; variation in the intensity of competition should consequently induce correlational selection favouring the adaptive integration of life-history, morphology, behaviour and their physiological underpinnings (Wright et al., 2019). The idea is that selection for early reproduction, or large clutch sizes per reproductive attempt (Araya-Ajoy et al., 2018), is favoured when competition is relaxed. Under such conditions, large risk-takers in particular would be able to both monopolize and exploit the available resources required to produce relatively large clutches (Wright et al., 2019). A key component of this ‘fast’ lifestyle is a decreased investment in self-maintenance, which would ultimately result in a shorter lifespan or earlier onset of reproductive senescence (Moiron et al., 2020).

The general prediction that risk-takers are heavier and/or larger has been confirmed by a recent meta-analysis (Niemelä & Dingemanse, 2018). Previous work on great tits fully aligned with this meta-analytic result: among males, larger individuals were both heavier, and more explorative and aggressive (Moiron et al., 2019). Simultaneously, the more explorative and aggressive individuals—though larger (thus heavier)—were relatively lean: their body mass was relatively low for their size (Moiron et al., 2019). These findings make sense, as such relatively lean individuals do poorly when breeding densities are high and competition for resources is intense (Both et al., 1999). In the same population that Moiron et al. (2019) studied, the more explorative great tits also produced larger clutches per reproductive attempt, and showed an earlier onset of reproductive senescence (Araya-Ajoy et al., 2016; Dingemanse, Moiron, et al., 2020). Altogether, these findings suggest, first, that great tits with a faster pace-of-life trade-off investments in self-maintenance towards current reproduction and, second, that a suite of morphological, behavioural and life-history traits are integrated as part of pace-of-life syndromes.

Various empirical studies imply that physiological traits are also integrated into such behaviour-morphology syndromes. For example, in great tits, the less explorative (Carere & Van Oers, 2004), and in blue tits, the less aggressive (Class & Brommer, 2020), individuals have higher breathing rates. This particular behaviour-physiology syndrome is thought to exist because less (vs. more) explorative birds have a reactive (vs. proactive) stress physiology (Coppens et al., 2010; Groothuis & Carere, 2005), of which breathing rate is a phenotypic indicator (Carere & Van Oers, 2004). The stability of this integration between risk-taking behaviours and physiology is, however, debated as behaviour-physiology correlations may be labile and vary with age, sex or environmental context (Class & Brommer, 2015; Kluen et al., 2022; Krams et al., 2014). This lack of consensus calls for studies investigating whether the integration of behaviour, physiology and morphology is replicable over—rather than specific to—species, populations or other characteristics of the study model.

We tackle this question by testing whether individual blue and great tit males and females of the same and different populations share the same pattern of phenotypic integration. We repeatedly measured a suite of behavioural (exploration), physiological (breathing rate) and morphological traits (body mass, tarsus length, wing length, bill length) on all individuals breeding in two populations, and did so for multiple years. We then estimated trait correlations among-individuals separately for each unique combination of species, population, and sex (hereafter called ‘datasets’). Blue tits were not studied in the Starnberg population, and we thus compared six (rather than eight) datasets. Following Moiron et al. (2019), we use structural equation models (SEMs) to test for the existence of (a) a latent variable (LV) driving observable expressions of size (body mass, tarsus length, wing length, bill length) and other phenotypic traits (exploration and breathing rate), (b) a size-independent (i.e. size-corrected) syndrome between the non-morphological traits (exploration behaviour and breathing rate) and (c) a size-independent effect of body mass on the non-morphological traits (i.e. an effect of body mass not attributable to size). Next, we made use of meta-analytic approaches to specifically ask which of the structural paths included in the SEMs were overall supported (vs. unsupported), and whether the strength and direction of each focal path was the same (vs. different) between our six datasets (Figure 1).

2 | MATERIALS AND METHODS

2.1 | Study area

The data were collected in two separate study areas. The first area consisted of six plots situated in the Forstenrieder Park (southwest of Munich, Germany; 48°2′49″N, 11°27′40″E; Figure S1). The vegetation of the park consisted of mixed deciduous forest, composed primarily of beech *Fagus sylvatica* and oak *Quercus robur*. Plots were established during the autumn of 2018 (A, B, C, D; Figure S1) and 2019 (E, F; Figure S1). All boxes were placed in a ~50-m grid, except for 15 pairs of boxes in plots A, B and F in 2020; plots A, B and D in 2021; and A, C and D in 2022, which were placed ~5 m apart from each other. Differences in the number as well as distance of nest boxes across study plots were part of another study for which we manipulated densities of blue and great tit nest boxes (Table S1). All boxes had inner margins of $14.6 \times 12 \times 25\,\mathrm{cm}^3$ but differed in the diameter of the entry hole: 32 mm (‘great tit boxes’) vs. 26 mm (‘blue tit boxes’). For the current study, only data between 2020 through 2022 were used as blue tits were monitored only from the second year onward.

The second area consisted of 12 plots situated between the Ammersee and Starnbergersee (south of Starnberg, Germany;

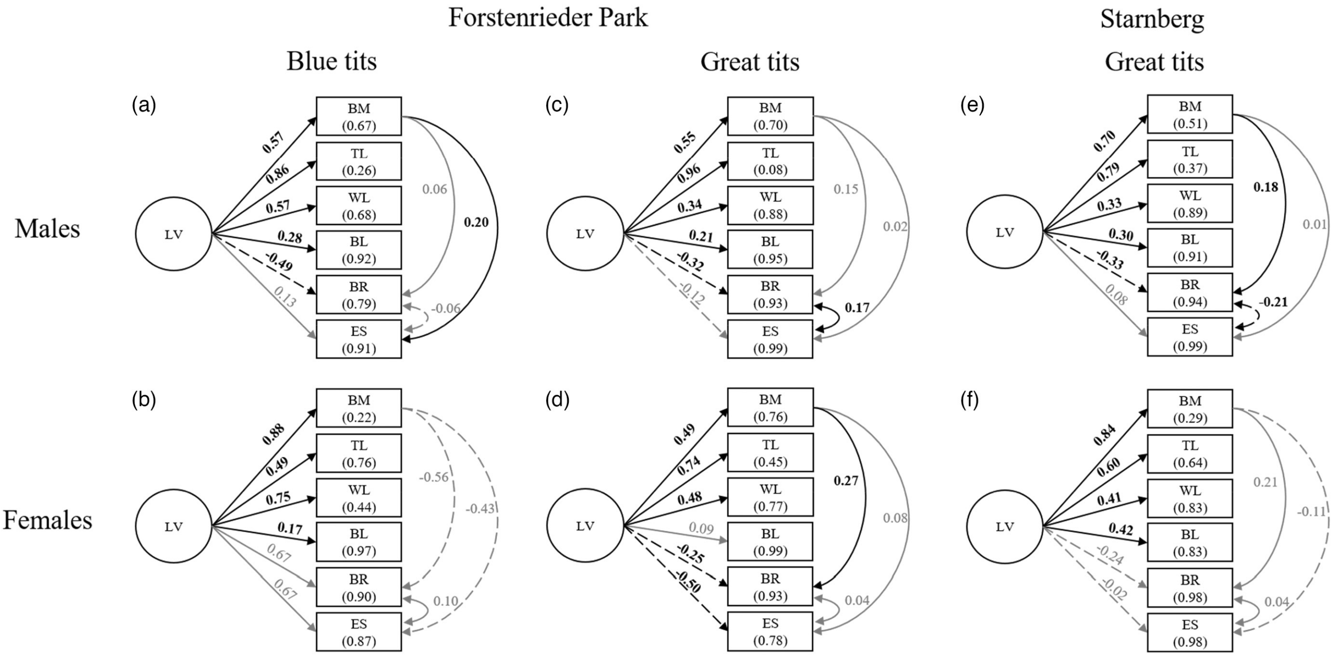

FIGURE 1 Visualization of the structural equation model (SEM) fitted for each combination of species, population, and sex. Shown are the point estimates for each of nine structural paths connecting behaviour, physiology, and morphology among-individuals in each dataset. Numbers correspond to standardized path coefficients (single-headed arrows) or residual among-individual correlations (double-headed arrows). Solid versus dashed arrows are positive versus negative estimates, respectively. Residual variances are printed inside boxes. Black versus grey arrows are statistically significant ($p < 0.05$) versus non-significant ($p > 0.05$) estimates, respectively. See Table S2 for the uncertainty (SE) and statistical significance of each path. Panels (a–d) show the SEM for blue tit males (a) and females (b), and great tit males (c) and females (d) of the Forstenrieder Park population. Panels (e–f) show the SEM for great tit males (e) and females (f) of the Starnberg population. BL, bill length; BM, body mass; BR, breathing rate; ES, exploration score; LV, latent variable; TL, tarsus length; WL, wing length.

47°58′N, 11°14′E; hereafter called Starnberg) that were monitored from 2010 through 2019. Each plot fitted 50 nest boxes in a regular grid with ~50 m between boxes. For general descriptions of this study area, see Nicolaus et al. (2015). In Starnberg, only great tit boxes were used, that is, only data for great tits were collected, which explains why our analyses compared phenotypic integration among six (rather than eight) unique combinations of species, populations and sexes (Figure 1).

2.2 | Data collection

In both study areas, we inspected nest boxes biweekly from the beginning of April to record lay date, clutch size and incubation date. Fourteen days following incubation onset, we inspected nest boxes daily until egg hatching (day 0). We captured (and ringed if not previously banded) both parents (day 10–12) with a spring trap set inside the nest box and assayed their exploration behaviour in a cage test (see Section 2.3). Immediately after the exploration test, the bird was removed from the cage and its breathing rate recorded (Holtmann & Dingemanse, 2022). To do so, the focal bird was held in a ringer's grip laying on its back with its breast and belly visible to the observer (Kluen et al., 2014). In both populations, the number of breaths were counted for a 30s-period. In Starnberg, this procedure was repeated a second time but this data are not used here as we did not collect a repeat measure in the Forstenrieder Park. We further measured body mass, tarsus length, wing length and bill length. In Forstenrieder Park, we measured body mass using a digital scale (±0.01 g) and wing length as maximum chord (Svensson, 1992). In Starnberg, we measured body mass using a Pesola spring balance (±0.25 g) and wing length as the feather length of the 3rd outermost primary (Jenni & Winkler, 1989). Previous studies have shown that these two alternative measures of wing length are highly correlated, thus measuring the same trait (Jenni & Winkler, 1989). We determined the breeder's sex based on the presence/absence of a brood patch (both species), and on the size of the black breast band (great tits) or the hue of crown feathers (blue tits). We aged breeders (first-year breeder vs. older) using plumage characteristics (Jenni & Winkler, 1994). Following all measurements, we released the focal breeder near its nest box.

2.3 | Exploration behaviour

Exploration behaviour was assayed in a small cage representing a smaller-scale field version of the classic novel environment test (Dingemanse et al., 2002; for details of the cage setup, see Stuber et al., 2013). Prior to recording exploration behaviour, the focal individual was placed in a side compartment of the cage for 30 s to allow habituation. We then opened the side door of the compartment to release the bird into the cage without handling, and immediately video-recorded its behaviour for 2 min with a camera (Panasonic HC-V100) placed 1 m from the cage while the observer was out of sight. In Starnberg, videos were later scored using JWatcher v.1.0 (Blumstein et al., 2006) whereas in Forstenrieder Park, we used Boris v.7.9.8 (Friard & Gamba, 2016). To do so, we divided the cage into 12 sections consisting of six sections on the front mesh, three floor sections and three perches (detailed in fig. 1 in Stuber et al., 2013). Finally, we calculated exploration scores as the number of movements made between the 12 sections within the first 2 min. This behaviour is positively correlated with anti-predator boldness (Stuber et al., 2013) and aggression (Moiron et al., 2019) and was thus viewed as a measure of risk-taking behaviour (cf. Carter et al., 2013).

As part of another experiment in Forstenrieder Park, some individuals (*n* = 60 blue tits and *n* = 92 great tits) stayed in the side compartment for up to 120 s (year 2020 only). We analysed a subset of randomly allocated individuals within pairs for which one breeding pair member spent 30 s and the other 120 s in the side compartment (*n* = 6 blue tit pairs and *n* = 7 great tit pairs), which showed no statistically significant effect of the amount of time spent in the side compartment (see Text S1).

All procedures were approved by the Regierung von Oberbayern, Bayern, Germany (permission number: ROB-55.2-2532.Vet_02-17-215). All applicable international, national and/or institutional guidelines for the care and use of animals were followed.

2.4 | Data selection and statistical analysis

Great tits normally produce second clutches (Tinbergen, 1987; Verhulst, 1998) but blue tits rarely do so (Dhondt, 2012). To facilitate comparisons between the two species, we therefore only analysed data from first broods. We defined first broods as those produced within 30 days after the first egg of each species was laid (within each species within a given population) (Nicolaus et al., 2015). Moreover, we only included first broods for which the identity of the female was known as this helped avoid pseudoreplication caused by replacement broods of unknown females (Starnberg: 1664 out of 2222 great tit broods; Forstenrieder Park: 280 out of 397 great tit and 210 out of 269 blue tit broods). Furthermore, Moiron et al. (2019) included aggressiveness in their SEM-analyses of phenotypic integration in male great tits, but we did not because aggressiveness was not assayed in females of either population. The lack of this data for one sex meant that we would not be able to execute our aim (formally comparing the same SEMs across the six unique combinations of species, population, and sex; Figure 1).

2.5 | Multivariate mixed-effects modelling

We partitioned variances and covariances among individuals, plots, and field observers for each of the six traits by fitting a multivariate mixed-effects model with the six traits (body mass, tarsus length, wing length, bill length, breathing rate, and exploration score) as the response variables. Response variables were scaled (mean-centred and expressed in SD-units) prior to analyses and modelled with a Gaussian error distribution. We performed six models, one for each unique combination of species (blue vs. great tits), population (Forstenrieder Park vs. Starnberg) and sex (male vs. female). We included data from 334 adult blue tits (400 observations) and 402 adult great tits (511 observations) from Forstenrieder Park, and 2123 adult great tits (3226 observations) from Starnberg. Each model included year as a fixed-effect factor (Forstenrieder Park: *n* = 3; Starnberg: *n* = 10) to account for year-specific temporal effects. Random intercepts were fitted for individual identity, plot (Forstenrieder Park: *n* = 6; Starnberg: *n* = 12) and field observer identity (Forstenrieder Park: *n* = 21; Starnberg: *n* = 56). Following visual inspection of the data, we removed three data points for body mass, two for tarsus length, twelve for wing length and four for bill length of the Forstenrieder Park dataset. Additionally, we removed one data point for bill length of the Starnberg dataset. As detailed in Text S2, those measurements were all outside the natural range of the focal species, and thus considered faulty data rather than extreme phenotypes. All models were run following the Bayesian framework using the function 'MCMCglmm' of the MCMCGLMM package v.2.34 (Hadfield, 2010) in R v.4.1.2 (R Core Team, 2020). The models sampled 650,000 iterations, with a burning interval of 150,000 and a thinning interval of 500. For all multivariate mixed-effects models, we specified an inverse-gamma prior (V = diag(6), nu = 1.002) for the residuals and a parameter expanded prior (V = diag(6), nu = 6, alpha.mu = diag(0, 6), alpha.V = diag(6)*1000) for the random effects.

2.6 | Structural equation models

We estimated the strength, sign and standard error (SE) of each of the nine structural paths associated with each SEM (Figure 1). As model input, we used the estimated among-individual correlation matrix obtained from the multivariate mixed-effects models, where the sample size was assumed to equate the number of individuals. We achieved this by using the 'sem' function of the SEM package v.3.1-15 (Fox, 2006) in R v.4.1.2 (R Core Team, 2020). This procedure was executed separately for each of our six unique combinations of species, population and sex (Figure 1).

2.7 | Meta-analysis

We performed two types of meta-analyses. First, we fitted intercept-only random-effects meta-analyses. We did so separately for each of the nine structural paths that the SEM estimated for each trait (Figure 1). We did so using the 'rma' function of the METAFOR package v.3.81-1 (Viechtbauer, 2010) in R v.4.1.2 (R Core Team, 2020). The response variable in these meta-analyses was the SEM's point estimate of each focal path, where each of our six unique combinations of species, population and sex provided one data point. Sampling variance was calculated as the squared of the SE (Nakagawa et al., 2022) associated with each focal SEM estimate. We fitted random intercepts for dataset (i.e. the unique combination of species, population, and sex; $n=6$). Estimates were considered statistically significant when the 95% confidence intervals (CIs) did not overlap zero. We report I^2 as an estimate of the relative heterogeneity for each of the intercept-only random-effects meta-analyses and corresponds to the unexplained variation across effect sizes—in our case across the six unique combinations of species, population, and sex—that was not attributable to their differences in sample size (i.e. sampling variance). I^2 values of 25%, 50%, and 75% are considered low, moderate, and high levels of heterogeneity, respectively (Higgins et al., 2003).

Second, we performed uni-moderator random-effects meta-regressions for each response variable (see above) that could explain some of the heterogeneity found. Our moderators were species, population and sex. We calculated χ^2-based significance tests to compare the levels of each moderator, and further report values of R^2_{marginal} as indicators of the amount of heterogeneity (I^2) explained by the moderator in each of the random-effects meta-regressions.

3 | RESULTS

We found strong evidence for the existence of a latent variable affecting our observable expressions of size. For each of the six unique combinations of species, population and sex, all four morphological traits (body mass, tarsus length, wing length and bill length) were positively and statistically significantly connected to the latent variable (Figure 1; Table S2). The only exception was bill length, where the standardized effect of the latent variable was positive in all but significantly supported in only five of the six datasets. The global (meta-analytic) effect sizes for the standardized path coefficient connecting the latent variable to body mass, tarsus length, wing length and bill length, were, respectively, all positive and strongly supported (due to 95% CIs not overlapping zero) (Table 1; Figure 2). This implies that heavier animals had longer tarsi, longer wings and longer bills across combinations of species, population and sex.

In five of the six datasets, the latent variable negatively affected breathing rate (Figure 1; Table S2). The global (meta-analytic) effect size for this standardized effect of the latent variable on breathing rate was negative and strongly supported (due to 95% CIs not overlapping zero) (Table 1; Figure 2). The strength of this path did not differ between datasets as our global meta-analysis showed no statistical support for heterogeneity (Table 1). Larger animals thus generally breathed slower. In five of the six datasets, the residual effect of body mass on breathing rate independent of effect of size was, by contrast, positive. This implies that animals that were relatively heavy for their size, also breathed relatively fast. The global (meta-analytic) effect size for this standardized size-independent effect of body mass on breathing rate was also positive and strongly supported (due to 95% CIs not overlapping zero) (Table 1; Figure 2). The strength of this path did not differ between datasets as there was no statistical support for heterogeneity (Table 1). These results altogether imply that larger animals generally breathe slower and that animals that are heavy for their size generally breathe faster.

The latent variable, however, did not consistently affect exploration score. This link was significantly heterogeneous, while the global (meta-analytic) effect did not differ from zero (due to 95% CIs overlapping zero) (Table 1; Figure 2). In other words, in some datasets, the larger animals were more explorative, while in other datasets, they were less explorative (Figure 1). There was also no overall support for a residual effect of body mass on exploration score that was independent of the effect of size, and neither was this link heterogeneous (Table 1). That is, animals that were relatively heavy for their size were not more/less explorative. Indeed, only in one out of six datasets was this link statistically supported (Figure 1; Table S2).

Finally, we found strong evidence for heterogeneity in the correlation between behaviour and physiology that was attributable to their respective covariances with size, while the overall

TABLE 1 Global (meta-analytic) effect sizes of each of nine hypothesized structural paths with 95% confidence interval (CI), 95% predictability interval (PI) and among-dataset heterogeneity (I^2). Single-headed arrows (→) represent standardized path coefficients; double-headed arrows (↔) represent correlations.

Path	Global effect size (95% CI)	95% PI	Heterogeneity I^2	χ^2_5	p
Latent variable → Body mass	0.68 (0.56, 0.81)	(0.38, 0.99)	83.65%	25.91	<0.001
Latent variable → Tarsus length	0.72 (0.60, 0.85)	(0.42, 1.03)	84.90%	29.39	<0.001
Latent variable → Wing length	0.47 (0.35, 0.60)	(0.17, 0.77)	86.50%	27.26	<0.001
Latent variable → Bill length	0.26 (0.16, 0.36)	(0.03, 0.49)	77.90%	25.10	<0.001
Latent variable → Breathing rate	−0.32 (−0.40, −0.23)	(−0.40, −0.23)	0.00%	7.64	0.18
Latent variable → Exploration score	−0.04 (−0.26, 0.17)	(−0.54, 0.46)	82.10%	23.99	<0.001
Body mass → Breathing rate	0.17 (0.10, 0.25)	(0.10, 0.25)	0.10%	5.39	0.37
Body mass → Exploration score	0.03 (−0.04, 0.09)	(−0.04, 0.09)	0.00%	6.47	0.26
Breathing rate ↔ Exploration score	0.00 (−0.11, 0.12)	(−0.27, 0.28)	87.00%	51.24	<0.001

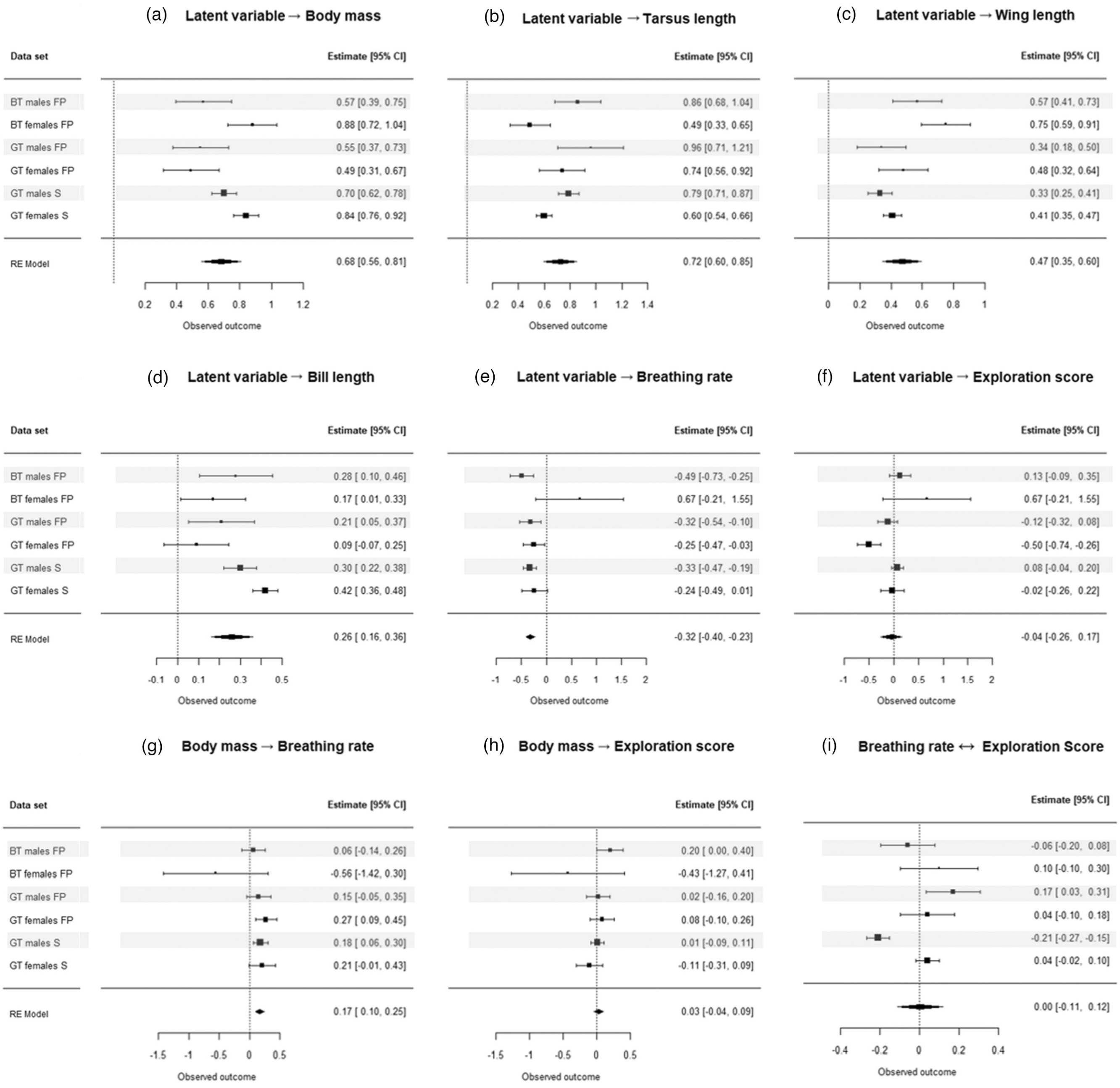

FIGURE 2 Forest plots of the global (meta-analytic) effects for standardized paths coefficients (a–h) and correlations (i) with 95% confidence intervals (for model estimates see Table 1). BT, blue tit; F, female; FP, Forstenrieder Park; GT, great tit; M, male; RE Model, global effect size; S, Starnberg.

(meta-analytic) effect size was not different from zero (due to 95% CIs overlapping zero) (Table 1). This was because this size-corrected correlation between breathing rate and exploration score was significantly positive in some but significantly negative in other datasets (Figure 1; Table S2).

In summary, we found strong and consistent support for both size- and condition-related physiology, as larger animals breathed slower, but animals heavy for their size breathed faster. By contrast, relationships between size and exploration behaviour, and between physiology and exploration behaviour were strongly heterogeneous and not overall supported.

The uni-moderator random-effects meta-regressions (Tables S3–S5) provided no support for main effects of species, population, or sex on integration between size and the non-morphological traits (exploration score and breathing rate). These factors neither explained variation in the effect of mass on the non-morphological traits independent of size, nor in the covariance between non-morphological traits independent of size. The heterogeneity between datasets in integration between size and exploration behaviour, or between exploration behaviour and breathing rate (Table 1) was therefore attributable to other dataset-specific factors of unknown origin. These analyses did, by contrast, support effects of species, population and sex on how the latent variable affected certain observable expressions of size. Specifically, wing length was more closely reflective of the latent variable ('size') in blue compared to great tits (Table S3), bill length was more closely reflective of the latent variable in Forstenrieder Park compared to Starnberg (Table S4), and tarsus length was more closely reflective of the latent variable in males compared to females (Table S5).

4 | DISCUSSION

This study was designed to test whether patterns of phenotypic integration between morphology, behaviour and physiology were generally representative of tit populations vs. specific to species, population or the sex under study. We therefore assessed and compared patterns of phenotypic integration among species (blue tits vs. great tits), populations (Forstenrieder Park vs. Starnberg), and sexes (males vs. females). We found that certain components, specifically, the integration of physiological and morphological traits, was highly conserved across datasets. This conservation occurs when the strength of the trait correlations is qualitatively similar between datasets (Dingemanse, Barber, et al., 2020). In all cases, (i) morphological traits were positively correlated, which supported the notion that each represented an 'observable' reflection of the animals' size. Each dataset further exhibited strong support for the existence of (ii) size-dependent physiology and (iii) size-independent effect of body mass on breathing rate. Larger animals breathed slower, but animals that were heavy for their size breathed faster. Our analyses implied that these structures of phenotypic integration were generally representative rather than specific to characteristics of the study species, population or sex.

The structure of phenotypic trait integration was thus shared among species, populations and sexes, but this was only true for the integration of physiology and morphology. Any links with behaviour were, by contrast, dataset-specific. In some datasets, we found evidence for size-dependent behaviour. Similarly, in some datasets, behaviour and physiology were structured into syndromes independent of size or size-corrected body mass. Importantly, in both cases, the characteristic pattern of phenotypic integration found in one specific combination of species, population and sex, was not predictive of patterns of phenotypic integration in other datasets. This was because neither pattern of phenotypic integration was overall supported, while both patterns of phenotypic integration varied qualitatively and significantly across datasets. Our uni-moderator random-effects meta-regressions implied that this was not because patterns of behavioural integration differed between species, population or sexes. This suggests that population differences in study design (see Section 2) did not underpin this form of heterogeneity and, thus, that ecological conditions specific to unique combinations of species, populations and/or sexes are likely drivers of heterogeneous patterns of phenotypic integration between behaviour and morphology, and behaviour and physiology observed in natural tit populations. The specific ecological drivers of the correlational selection pressures that might shape the evolution of these diverse patterns of integration between behaviour and other phenotypic traits now requires study. For example, differences in the strength of competition or risk of predation specific to a unique combination of species, population, or sex might explain this variation. Identifying such drivers goes beyond the scope of this study, as this would require many replicates species and populations for each sex as well as quantitative information on ecological variation between these major biological categories. For other patterns of phenotypic integration, studies of the role of ecological variation may perhaps not be warranted. Specifically, the size-independent relationship between body mass and exploration was neither heterogeneous nor supported overall. This conclusion, notably, assumes that our specific study species and populations were more broadly representative of the average tit species and population.

Patterns of (variation in) phenotypic integration are ultimately underpinned by (variation in) the relative contribution of genetic and environmental effects on trait correlations (Searle, 1961). Our finding that correlations between observable expressions of size (tarsus length, bill length, wing length and body mass) were overall highly preserved suggests that patterns of genetic correlations or early-life environmental effects are largely shared among species, populations, and sexes. However, we did find some differences in how specific components of morphology was integrated differently across species (wing length), populations (bill length) or sexes (tarsus). A next step would thus be to unravel the quantitative genetics background of size- and condition-dependent physiology and to study whether the pattern of genetic integration of these traits facilitates or constraints adaptive evolution (Dochtermann & Dingemanse, 2013). This would ultimately require not just an understanding of both the quantitative genetics underpinning (e.g. Dingemanse, Barber, et al., 2020) but also insights into the patterns of (correlational) selection in the wild (Lande & Arnold, 1983; Schluter, 1996).

This combination of quantitative genetics and selection studies will be particularly fruitful when applied to the integration of (exploration) behaviour. From a proximate perspective, there are multiple non-exclusive explanations for why the phenotypic integration between behaviour and morphology, and behaviour and physiology varied among datasets. This is evident when inspecting the mathematical equation demonstrating how different variance components contribute to observed correlations (detailed in Text S3). Specifically, the among-individual correlation r_I between two focal traits that we used as input in our SEMs can differ between datasets because they differ in the amount of additive genetic variance, the influence of permanent environmental effects, genetic trait correlations or permanent environmental trait correlations. Genetic correlations can result from selection on trait integration ('correlational selection') causing linkage-disequilibrium. Differences in trait correlations among species, populations or sexes may thus reflect variation in selection pressures (Roff & Fairbairn, 2012). Permanent environmental correlations, by contrast, can occur when the two focal traits are underpinned by different—yet correlated—environmental factors, and thus change when correlations between environmental factors are different between unique combinations of species, populations or sexes. These examples imply that a productive next step in furthering our understanding of differences in phenotypic integration among datasets is to estimate the variance components that shape observed correlation structures (Kruuk, 2004). In stickleback fish, for example, behavioural correlations were stronger in populations living in the presence of predators (Dingemanse et al., 2010) but genetic correlations did not differ between the types of population, neither were they affected by experience with predators during ontogeny (Dingemanse, Barber, et al., 2020). This implied that differences in behavioural correlations were caused by differences in permanent-environmental correlations between predator-naïve and predator-sympatric populations (Dingemanse, Barber, et al., 2020). Such approaches might also be applied to understand why allometric relationships (the integration of observable expression of size) differed somewhat between species, populations and sexes.

Our study demonstrates the value of study replication in behavioural ecology. We asked whether the same biological patterns of phenotypic integration characterized data collected from the same species and sex across populations, and from different species within the same population. Our analyses demonstrated that certain aspects of phenotypic integration were common to all datasets, which implies that such patterns likely also characterize blue and great tit males and females from other populations. The ability to draw such conclusions is an important benefit of replication studies. Similarly, we have demonstrated that the sign and strength of the link with exploration behaviour is highly dataset-specific and not due to main effects of species, population, or sex. In particular, based on a focal study of blue or great tits, we cannot predict how animal personality is related to structural size or body condition in other populations of those species if populations differ in ecology. An important obstacle in interpreting meta-analytical differences and similarities is that datasets often differ in methodology (Nakagawa & Santos, 2012). In our case, body mass was measured with different devices (digital scale vs. Pesola spring balance) and wing length was measured slightly different (maximum chord vs. feather length, see Section 2), which could explain why their respective links with the latent variable reflective of 'size' were somewhat heterogeneous across datasets. However, we did not find population differences in path coefficients involving body mass or wing length. This finding corroborates our assumption that our alternative measures of wing length measured the same trait, as did our alternative measures of body mass. We otherwise used exactly the same methodology to measure our physiological and behavioural traits. Another obstacle in comparing published datasets is that estimates obtained from the literature are typically derived from differently structured models. We avoided this problem by fitting the same structure to all six datasets (Figure 1). This approach, however, does constrain our ability to compare our findings with previously published results coming from different analyses. Specifically, in male great tits from the Starnberg population, Moiron et al. (2019) found evidence for size- and condition-dependent behaviour. Our analyses of the same dataset implied that neither link was present. The difference in results of analyses that were based on the very same dataset can likely be explained by our decision to use breathing rate in the current study rather than aggression as in Moiron et al. (2019). We took this decision because the former was measured for all datasets (i.e. sexes) whereas the latter was not (see Section 2). This could mean that among-individual correlations between exploration behaviour and size and between exploration behaviour and size-independent body mass vary depending on whether correlations with aggression versus breathing rate are accounted for. This underlines the need for methodical robustness in analyses as an insightful tool for testing the generality of published results.

5 | CONCLUSIONS

This replication study has demonstrated that populations of great and blue tits generally show patterns of phenotypic integration characterized by size-dependent and condition-dependent physiology among-individuals: large individuals breathed slower, and individual that were heavy for this size breathed faster. The pattern of integration between behaviour and size, behaviour and size-corrected body mass, and physiology and behaviour differed between datasets but was not overall supported. This implies that variability in trait integration should be expected in the context of state-dependent personality variation, and that future studies should focus on identifying the proximate and functional drivers of this variation in phenotypic integration of behaviour and morphology in the wild. A fruitful next step would be to apply our approach to study jointly variation in phenotypic integration and ecology in large numbers of (published and unpublished) datasets, such as those becoming increasingly available through large-scale collaborations such as the SPI-Birds network (Culina et al., 2021).

AUTHOR CONTRIBUTIONS

Irene Gaona-Gordillo, Niels J. Dingemanse and Benedikt Holtmann conceived the study; Irene Gaona-Gordillo, Benedikt Holtmann, Alexia Mouchet, Alexander Hutfluss and Niels J. Dingemanse performed fieldwork; Irene Gaona-Gordillo and Benedikt Holtmann coordinated fieldwork and managed the database; Irene Gaona-Gordillo and Benedikt Holtmann extracted exploration data from videos; Irene Gaona-Gordillo analysed the data with input from Benedikt Holtmann and Niels J. Dingemanse; Alfredo Sánchez-Tójar advised on meta-analyses and reproducible coding; Irene Gaona-Gordillo and Niels J. Dingemanse wrote the manuscript with input of all other authors. All authors gave final approval for publication.

ACKNOWLEDGEMENTS

We thank all members and students of the Behavioural Ecology Group for help with data collection. We are grateful to Stefanie Grosser, Sebastian Höhna and Sakie Komada for their assistance during study plot set-up. We thank Johannes Steigner and the Bayrischen Staatsforsten for support and permission to conduct research on their grounds. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

Irene Gaona-Gordillo was supported by a Forschungsstipendien of the Deutscher Akademischer Austauschdienst (DAAD; grant number 91731707). The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) funded Benedikt Holtmann (HO 6288/1-1) and Alexia Mouchet, Alexander Hutfluss and Niels J. Dingemanse (grant no. DI 1694/1-1).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw data and scripts are available via Zenodo at https://doi.org/10.5281/zenodo.7874729 (Gaona-Gordillo et al., 2023).

ORCID

Irene Gaona-Gordillo https://orcid.org/0000-0001-5437-5754
Benedikt Holtmann https://orcid.org/0000-0002-2995-7274
Alexia Mouchet https://orcid.org/0000-0001-5572-4688
Alexander Hutfluss https://orcid.org/0000-0001-8514-6452
Alfredo Sánchez-Tójar https://orcid.org/0000-0002-2886-0649
Niels J. Dingemanse https://orcid.org/0000-0003-3320-0861

REFERENCES

Araya-Ajoy, Y. G., Bolstad, G. H., Brommer, J., Careau, V., Dingemanse, N. J., & Wright, J. (2018). Demographic measures of an individual's 'pace of life': Fecundity rate, lifespan, generation time, or a composite variable? *Behavioral Ecology and Sociobiology*, *72*, 1–14. https://doi.org/10.1007/s00265-018-2477-7

Araya-Ajoy, Y. G., Kuhn, S., Mathot, K. J., Mouchet, A., Mutzel, A., Nicolaus, M., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2016). Sources of (co)variation in alternative siring routes available to male great tits (*Parus major*). *Evolution*, *70*, 2308–2321. https://doi.org/10.1111/evo.13024

Armbruster, W. S., Pélabon, C., Bolstad, G. H., & Hansen, T. F. (2014). Integrated phenotypes: Understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130245. https://doi.org/10.1098/rstb.2013.0245

Armbruster, W. S., & Schwaegerle, K. E. (1996). Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology*, *9*, 261–276. https://doi.org/10.1046/j.1420-9101.1996.9030261.x

Barber, I., & Dingemanse, N. J. (2010). Parasitism and the evolutionary ecology of animal personality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 4077–4088. https://doi.org/10.1098/rstb.2010.0182

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

Blumstein, D. T., Daniel, J. C., & Evans, C. S. (2006). *JWatcher 1.0: An introductory user's guide*. 211.

Both, C., Visser, M. E., & Verboven, N. (1999). Density-dependent recruitment rates in great tits: The importance of being heavier. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 465–469. https://doi.org/10.1098/rspb.1999.0660

Brodie, E. D. (1992). Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. *Evolution*, *46*, 1284–1298. https://doi.org/10.1111/j.1558-5646.1992.tb01124.x

Brodie, E. D. (1993). Homogeneity of the genetic variance-covariance matrix for antipredator traits in two natural populations of the garter snake *Thamnophis ordinoides*. *Evolution*, *47*, 844–854. https://doi.org/10.1111/j.1558-5646.1993.tb01238.x

Brodie, E. D., Moore, A. J., & Janzen, F. J. (1995). Visualizing and quantifying natural selection. *Trends in Ecology & Evolution*, *10*, 313–318. https://doi.org/10.1016/S0169-5347(00)89117-X

Carere, C., & Van Oers, K. (2004). Shy and bold great tits (*Parus major*): Body temperature and breath rate in response to handling stress. *Physiology & Behavior*, *82*, 905–912. https://doi.org/10.1016/j.physbeh.2004.07.009

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, *88*, 465–475. https://doi.org/10.1111/brv.12007

Class, B., & Brommer, J. (2020). Contrasting multilevel relationships between behavior and body mass in blue tit nestlings. *Behavioral Ecology*, *31*, 702–710. https://doi.org/10.1093/beheco/araa014

Class, B., & Brommer, J. E. (2015). A strong genetic correlation underlying a behavioural syndrome disappears during development because of genotype–age interactions. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20142777. https://doi.org/10.1098/rspb.2014.2777

Coppens, C. M., De Boer, S. F., & Koolhaas, J. M. (2010). Coping styles and behavioural flexibility: Towards underlying mechanisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 4021–4028. https://doi.org/10.1098/rstb.2010.0217

Culina, A., Adriaensen, F., Bailey, L. D., Burgess, M. D., Charmantier, A., Cole, E. F., Eeva, T., Matthysen, E., Nater, C. R., Sheldon, B. C., Sæther, B. E., Vriend, S. J. G., Zajkova, Z., Adamík, P., Aplin, L. M., Angulo, E., Artemyev, A., Barba, E., Barišić, S., … Visser, M. E. (2021). Connecting the data landscape of long-term ecological studies: The SPI-birds data hub. *The Journal of Animal Ecology*, *90*, 2147–2160. https://doi.org/10.1111/1365-2656.13388

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*, 734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x

Dammhahn, M., Dingemanse, N. J., Niemelä, P. T., & Réale, D. (2018). Pace-of-life syndromes: A framework for the adaptive integration of behaviour, physiology and life history. *Behavioral Ecology and Sociobiology*, *72*, 1–8. https://doi.org/10.1007/s00265-018-2473-y

Dhondt, A. A. (2012). *Interspecific competition in birds*. Oxford University Press.

Dingemanse, N. J., Barber, I., & Dochtermann, N. A. (2020). Non-consumptive effects of predation: Does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? *Ecology Letters*, *23*, 107–118. https://doi.org/10.1111/ele.13413

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

Dingemanse, N. J., Dochtermann, N., & Wright, J. (2010). A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Animal Behaviour*, *79*, 439–450. https://doi.org/10.1016/j.anbehav.2009.11.024

Dingemanse, N. J., Moiron, M., Araya-Ajoy, Y. G., Mouchet, A., & Abbey-Lee, R. N. (2020). Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? *The Journal of Animal Ecology*, *89*, 601–613. https://doi.org/10.1111/1365-2656.13122

Dingemanse, N. J., & Wolf, M. (2010). Recent models for adaptive personality differences: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 3947–3958. https://doi.org/10.1098/rstb.2010.0221

Dochtermann, N. A. (2011). Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution*, *65*, 1814–1820. https://doi.org/10.1111/j.1558-5646.2011.01264.x

Dochtermann, N. A., & Dingemanse, N. J. (2013). Behavioral syndromes as evolutionary constraints. *Behavioral Ecology*, *24*, 806–811. https://doi.org/10.1093/beheco/art002

Filazzola, A., & Cahill, J. F. (2021). Replication in field ecology: Identifying challenges and proposing solutions. *Methods in Ecology and Evolution*, *12*, 1780–1792. https://doi.org/10.1111/2041-210X.13657

Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling*, *13*, 465–486. https://doi.org/10.1207/s15328007sem1303_7

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*, 1325–1330. https://doi.org/10.1111/2041-210X.12584

Gaona-Gordillo, I., Holtmann, B., Mouchet, A., Hutfluss, A., Sánchez-Tójar, A., & Dingemanse, N. J. (2023). Are animal personality, body condition, physiology and structural size integrated? A comparison of species, populations and sexes, and the value of study replication. *Dryad Digital Repository*, https://doi.org/10.1111/1365-2656.13966

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*, 1213–1235. https://doi.org/10.1007/s10682-012-9589-8

Groothuis, T. G. G., & Carere, C. (2005). Avian personalities: Characterization and epigenesis. *Neuroscience and Biobehavioral Reviews*, *29*, 137–150. https://doi.org/10.1016/j.neubiorev.2004.06.010

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22. https://doi.org/10.18637/jss.v033.i02

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Holtmann, B., & Dingemanse, N. J. (2022). Strong phenotypic trait correlations between mating partners do not result from assortative mating in wild great tits (*Parus major*). *Journal of Evolutionary Biology*, *35*, 552–560. https://doi.org/10.1111/jeb.13908

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. https://doi.org/10.1111/1365-2435.12779

Jenni, L., & Winkler, R. (1989). The feather-length of small passerines: A measurement for wing-length in live birds and museum skins. *Bird Study*, *36*, 1–15. https://doi.org/10.1080/00063658909476996

Jenni, L., & Winkler, R. (1994). *Moult and aging of european passerines*. Academic Press.

Jennions, M. D., & Møller, A. P. (2002). Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 43–48. https://doi.org/10.1098/rspb.2001.1832

Kelly, C. D. (2006). Replicating empirical research in behavioral ecology: How and why it should be done but rarely ever is. *The Quarterly Review of Biology*, *81*, 221–236. https://doi.org/10.1086/506236

Kelly, C. D. (2019). Rate and success of study replication in ecology and evolution. *PeerJ*, *2019*, e7654. https://doi.org/10.7717/peerj.7654

Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., & Carlson, S. M. (2012). Synthetic analyses of phenotypic selection in natural populations: Lessons, limitations and future directions. *Evolutionary Ecology*, *26*, 1101–1118. https://doi.org/10.1007/s10682-012-9563-5

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, *157*, 245–261. https://doi.org/10.1086/319193

Klingenberg, C. P. (2008). Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 115–132. https://doi.org/10.1146/annurev.ecolsys.37.091305.110054

Klingenberg, C. P. (2014). Studying morphological integration and modularity at multiple levels: Concepts and analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 33–35. https://doi.org/10.1098/rstb.2013.0249

Kluen, E., Rönkä, K., & Thorogood, R. (2022). Prior experience of captivity affects behavioural responses to 'novel' environments. *PeerJ*, *10*, e13905. https://doi.org/10.7717/peerj.13905

Kluen, E., Siitari, H., & Brommer, J. E. (2014). Testing for between individual correlations of personality and physiological traits in a wild bird. *Behavioral Ecology and Sociobiology*, *68*, 205–213. https://doi.org/10.1007/s00265-013-1635-1

Krams, I. A., Vrublevska, J., Sepp, T., Abolins-Abols, M., Rantala, M. J., Mierauskas, P., & Krama, T. (2014). Sex-specific associations between nest defence, exploration and breathing rate in breeding pied flycatchers. *Ethology*, *120*, 492–501. https://doi.org/10.1111/eth.12222

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 873–890. https://doi.org/10.1098/rstb.2003.1437

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*, 1210. https://doi.org/10.2307/2408842

Moiron, M., Araya-Ajoy, Y. G., Mathot, K. J., Mouchet, A., & Dingemanse, N. J. (2019). Functional relations between body mass and risk-taking behavior in wild great tits. *Behavioral Ecology*, *30*, 617–623. https://doi.org/10.1093/beheco/ary199

Moiron, M., Laskowski, K. L., & Niemelä, P. T. (2020). Individual differences in behaviour explain variation in survival: A meta-analysis. *Ecology Letters*, *23*, 399–408. https://doi.org/10.1111/ele.13438

Montiglio, P. O., Dammhahn, M., Dubuc Messier, G., & Réale, D. (2018). The pace-of-life syndrome revisited: The role of ecological conditions and natural history on the slow-fast continuum. *Behavioral Ecology and Sociobiology*, *72*, 1–9. https://doi.org/10.1007/s00265-018-2526-2

Mouchet, A., Cole, E. F., Matthysen, E., Nicolaus, M., Quinn, J. L., Roth, A. M., Tinbergen, J. M., van Oers, K., van Overveld, T., & Dingemanse, N. J. (2021). Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, 1–6. https://doi.org/10.1073/pnas.2024994118

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*, 4–21. https://doi.org/10.1111/2041-210X.13724

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*, 1–6. https://doi.org/10.1186/s12915-015-0196-3

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*, 1253–1274. https://doi.org/10.1007/s10682-012-9555-5

Nicolaus, M., Mathot, K. J., Araya-Ajoy, Y. G., Mutzel, A., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2015). Does coping style predict optimization? An experimental test in a wild passerine bird. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20142405. https://doi.org/10.1098/rspb.2014.2405

Niemelä, P. T., & Dingemanse, N. J. (2018). On the usage of single measurements in behavioural ecology research on individual differences. *Animal Behaviour*, *145*, 99–105. https://doi.org/10.1016/j.anbehav.2018.09.012

Pigliucci, M., & Preston, K. (2004). *Phenotypic integration: Studying the ecology and evolution of complex phenotypes*. Oxford University Press.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Réale, 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*, 4051–4063. https://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*, 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Roff, D. A., & Fairbairn, D. J. (2012). A test of the hypothesis that correlational selection generates genetic correlations. *Evolution*, *66*, 2953–2960. https://doi.org/10.1111/j.1558-5646.2012.01656.x

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, *50*, 1766–1774. https://doi.org/10.1111/j.1558-5646.1996.tb03563.x

Searle, S. R. (1961). Phenotypic, genetic and environmental correlations. *Biometrics*, *17*, 474. https://doi.org/10.2307/2527838

Siepielski, A. M., Dibattista, J. D., & Carlson, S. M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, *12*, 1261–1276. https://doi.org/10.1111/j.1461-0248.2009.01381.x

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*, 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P. O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state-behaviour feedbacks: A review and guide for empiricists. *Trends in Ecology & Evolution*, *30*, 50–60. https://doi.org/10.1016/j.tree.2014.11.004

Sinervo, B., & Svensson, E. (2002). Correlational selection and the evolution of genomic architecture. *Heredity*, *89*, 329–338. https://doi.org/10.1038/sj.hdy.6800148

Stamps, J. A. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. *Ecology Letters*, *10*, 355–363. https://doi.org/10.1111/j.1461-0248.2007.01034.x

Stearns, S. (1992). *The evolution of life histories*. Oxford University Press.

Stuber, E. F., Araya-Ajoy, Y. G., Mathot, K. J., Mutzel, A., Nicolaus, M., Wijmenga, J. J., Mueller, J. C., & Dingemanse, N. J. (2013). Slow explorers take less risk: A problem of sampling bias in ecological studies. *Behavioral Ecology*, *24*, 1092–1098. https://doi.org/10.1093/beheco/art035

Svensson, L. (1992). *Identification guide to european passerines* (4th ed.). British Trust for Ornithology.

Tinbergen, J. M. (1987). Costs of reproduction in the great tit: Intraseasonal costs associated with brood size. *Ardea*, *75*, 111–122.

Verhulst, S. (1998). Multiple breeding in the great tit, II. The costs of rearing a second clutch. *Functional Ecology*, *12*, 132–140. https://doi.org/10.1046/j.1365-2435.1998.00165.x

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor. *Journal of Statistical Software*, *36*, 1–48.

Wilson, E. (1998). *Consilience: The unity of knowledge*. Little, Brown & Company.

Wolf, M., Van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, *447*, 581–584. https://doi.org/10.1038/nature05835

Wolf, M., & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 3959–3968. https://doi.org/10.1098/rstb.2010.0215

Wright, J., Bolstad, G. H., Araya-Ajoy, Y. G., & Dingemanse, N. J. (2019). Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. *Biological Reviews*, *94*, 230–247. https://doi.org/10.1111/brv.12451

Yang, Y., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Koricheva, J., Jennions, M. D., Parker, T. H., Lagisz, M., & Nakagawa, S. (2023). Publication bias impacts on effect size, statistical power, and magnitude (type M) and sign (type S) errors in ecology and evolutionary biology. *BMC Biology*, *21*, 1–20. https://doi.org/10.1186/s12915-022-01485-y

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1: Distribution of the study plots (A, B, C, D, E, and F) in the Forstenrieder Park. This study population was located southwest of Munich, Germany (48°2′49″N, 11°27′40″E).

Table S1: Subplot specific nest box density treatments applied in the Forstenrieder Park for three consecutive years. The four unique combination of density treatments (L: low vs. H: high, for blue tits: BT boxes and great tits: GT boxes) are shown thrice per year among the plots (A–F). Additionally, the table shows the total number of nest boxes per subplot and plot.

Text S1: We tested if the time spent in the side compartment of the exploration cage might influence the exploration scores.

Text S2: After visual inspection of the data from the blue tit adults of the Forstenrieder Park, we removed two observations of body mass (15.20 and 15.38 g; natural range 9.2–12.4 g; Blondel et al., 2002; Kullberg et al., 2002; Perrier et al., 2018), two observations of tarsus length (18.6 and 18.9 mm; maximum value 18.4 mm in 15 days old nestlings; Nord & Nilsson, 2011), eight values of wing length (>71 mm; maximum value found 69 mm but measured to the nearest of 1 mm; Furness & Furness, 2016), and two observations of bill length (10.5 and 10.8 mm; natural range 6.4–10.3 mm; Blondel et al., 2002; Perrier et al., 2018).

Table S2: For each unique combination of species, population, and sex, we ran a structural equation model that estimated all nine hypothesized paths. The full model was fitted using the among-individual correlation matrix among all six mean- and variance-standardized traits. We printed each path's standardized estimate, SE, and *p*-value. BL, bill length; BM, body mass; BR, breathing rate; ES, exploration score; LV, latent variable; TL, tarsus length; WL, wing length. Bold values represent paths that were statistically significant ($p < 0.05$). We present two types of parameters: standardized path coefficients (→) and correlations (↔).

Table S3: Estimates, 95% confidence intervals (CI) and the amount of heterogeneity (R^2_{marginal}) explained by the species (moderator) for each of the nine hypothesized structural paths.

Table S4: Estimates, 95% confidence intervals (CI) and the amount of heterogeneity (R^2_{marginal}) explained by the population (moderator) for each of the nine hypothesized structural paths.

Table S5: Estimates, 95% confidence intervals (CI) and the amount of heterogeneity (R^2_{marginal}) explained by the sex (moderator) for each of the nine hypothesized structural paths.

Text S3: From a proximate perspective, there are multiple non-exclusive explanations for why the phenotypic integration between behaviour and morphology, and behaviour and physiology, varied significantly among datasets.

How to cite this article: Gaona-Gordillo, I., Holtmann, B., Mouchet, A., Hutfluss, A., Sánchez-Tójar, A., & Dingemanse, N. J. (2023). Are animal personality, body condition, physiology and structural size integrated? A comparison of species, populations and sexes, and the value of study replication. *Journal of Animal Ecology*, *92*, 1707–1718. https://doi.org/10.1111/1365-2656.13966