Author for correspondence:
Yimen G. Araya-Ajoy
e-mail: yimencr@gmail.com

Royal Society Publishing

Characterizing behavioural 'characters': an evolutionary framework

Yimen G. Araya-Ajoy[1] and Niels J. Dingemanse[1,2]

[1]Evolutionary Ecology of Variation Group, Max Planck Institute for Ornithology, Seewiesen, Germany
[2]Behavioural Ecology, Department Biology II, Ludwig-Maximilians-University of Munich, Planegg-Martinsried, Germany

Biologists often study phenotypic evolution assuming that phenotypes consist of a set of quasi-independent units that have been shaped by selection to accomplish a particular function. In the evolutionary literature, such quasi-independent functional units are called 'evolutionary characters', and a framework based on evolutionary principles has been developed to characterize them. This framework mainly focuses on 'fixed' characters, i.e. those that vary exclusively between individuals. In this paper, we introduce multi-level variation and thereby expand the framework to labile characters, focusing on behaviour as a worked example. We first propose a concept of 'behavioural characters' based on the original evolutionary character concept. We then detail how integration of variation between individuals (cf. 'personality') and within individuals (cf. 'individual plasticity') into the framework gives rise to a whole suite of novel testable predictions about the evolutionary character concept. We further propose a corresponding statistical methodology to test whether observed behaviours should be considered expressions of a hypothesized evolutionary character. We illustrate the application of our framework by characterizing the behavioural character 'aggressiveness' in wild great tits, *Parus major*.

1. Introduction

Biologists often study phenotypic evolution assuming that phenotypes consist of a set of quasi-independent units or parts that have been shaped by selection to accomplish a particular function [1,2]. Consequently, the success of evolutionary research programmes depends to a large degree on whether such functional units have been properly characterized. For this reason, evolutionary biologists have developed an appealing conceptual framework (detailed below), in which these functional units are called 'evolutionary characters' [3]. Notably, despite the importance of labile characters in mediating interactions between organisms and their environment [4], they have not been fully integrated into this framework. This is in part because labile characters (e.g. behaviour) vary both between and within individuals; previous implementations have instead primarily focused on fixed phenotypes (e.g. structural size). In this paper, we expanded this framework to integrate (any) multi-level structure and illustrate its application by characterizing behavioural phenotypes. We introduce a definition of 'behavioural characters' and propose a general methodology that enables empirical testing of novel hypotheses concerning the question of whether observed behaviours can be considered expressions of a hypothesized evolutionary character.

Central to our framework is the concept of evolutionary characters, which can be defined as parts of an organism that exhibit causal coherence in their expression and play a causal role in a biological process [3]. This definition has two important characteristics. First, its causal coherence refers to a set of inter-related mechanisms that are involved in the character's expression and makes it quasi-independent from other characters [5]. This 'modularity' is what enables the character to respond adaptively to selection [6]. Second, its explicit link to a biological process implies that a character is a 'functional unit' used by an organism for a particular task. An evolutionary phenotypic
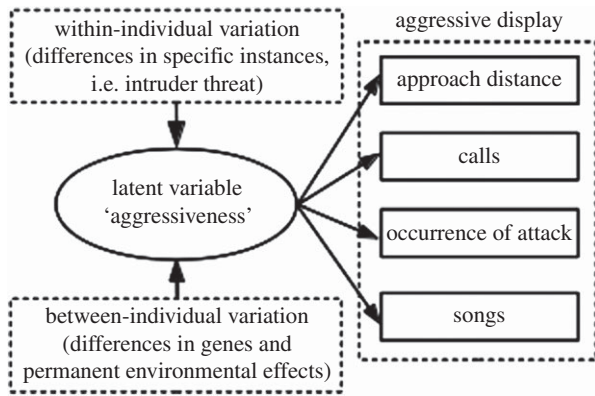
2

rspb.royalsocietypublishing.org   Proc. R. Soc. B 281: 20132645

Figure 1. Diagram of the multi-level evolutionary character concept applied to avian agonistic behaviour. The behavioural character 'aggressiveness' is represented as a latent variable affecting the expression of observed behaviours (calls, approach distance, occurrence of attack and songs). The hypothesized expression of the latent variable is plastic within the same individual, as it varies as a function of environmental conditions (top-left), but also differs between individuals owing to genetic and environmental effects specific to the individual (lower-left). Consequently, expressed behaviours are correlated in a similar fashion between versus within individuals.

module or 'character' is thus composed of several elements that are functionally related [7]. Characters are themselves in turn hierarchically structured, where a functional unit can be considered a part of a higher level unit [8]. For example, one can consider the human hand as a character that is composed of five fingers and is used to grab objects and use tools. Each finger has a specific function and can be considered a character by its own, but because all fingers need to be used as a coherent functional unit when using tools, they must be tightly correlated in terms of length, shape and neurological underpinning [9]; therefore, fingers of the same hand respond as a unit to selective forces and can be viewed as expressions of the same character. We propose to apply this general logic to behaviour and define behavioural characters by the causal coherence underlying their expression and the function that they accomplish for the organism.

We illustrate our behavioural character concept using aggressiveness displayed by territorial male great tits, *Parus major* (figure 1). We view aggressiveness as a behavioural character that dictates how an organism responds to agonistic interactions. Great tits express a wide array of behaviours during such encounters [10] that jointly execute a specific function: displacing intruders. We therefore *a priori* visualize aggressiveness as an unobserved—statistically called 'latent'—variable that affects multiple behaviours used in aggressive displays (visualized in figure 1 by arrows connecting the latent variable with the expressed behaviours). For example, during highly aggressive interactions, male great tits respond to a conspecific intrusion by calling while approaching and attacking if the intruder does not withdraw. By contrast, during less aggressive interactions, males sing from far away rather than calling and approaching close. Proximately, this functional coherence is owing to common mechanisms affecting the expression of all behaviours of the display (i.e. through pleiotropic effects of genetic or environmental factors; [11]). This common (neurological or physiological) pathway enables different behaviours to be expressed as a functional unit. It is this proximate mechanism that evolves in response to selection and that represents

the character [12]. We note that the terms phenotypic 'character' versus 'trait' are used interchangeably in the evolutionary literature. Traits are sometimes defined directly as observable variables that are biologically relevant; here, we simply call measured quantities 'observable variables' and refer to 'characters' as the inferred theoretical entities underlying the expression of functionally related observable variables. This borrows from the statistical and psychological literature where a distinction is made between attributes that are directly measurable versus those reflecting underlying unobservable quantities [13,14]. We thus propose that behavioural characters represent unmeasured 'latent' variables that can be inferred from the expression patterns of behavioural observables.

Other fields of biology, especially human personality research, have a long and productive history of studying behaviour using latent variables [14]. Our approach is distinctly different because we explicitly address the issue of how one might integrate behavioural variation between individuals (cf. 'animal personality' [15]) and variation within individuals (cf. adaptive 'individual plasticity' [16]) when studying these latent variables (behavioural characters) from an evolutionary perspective (detailed further in the Discussion). We will continue our worked example to explain this unique aspect of our approach. If there is a latent variable (aggressiveness) affecting the expression of the different agonistic behaviours, it will cause between-individual and within-individual correlations between the agonistic behaviours (cf. [17]). On the one hand, between-individual differences in aggressiveness owing to genetic differences or early-life experiences (visualized in figure 1 by the lower dashed box with arrows pointing to the latent variable) will result in between-individual correlations among behaviours of the display. Aggressive individuals should, for example, on average have high values for call rate as well as higher tendency to approach intruders. On the other hand, within-individual plastic responses to environmental changes should result in correlated changes in all behaviours of the display within the same individual (visualized in figure 1 by the upper dashed box with arrows pointing to the latent variable), resulting in within-individual correlations. If an individual increases its level of aggressiveness, its call rate should increase and it should approach the intruder closer. Decomposition and comparison of behavioural correlations within versus between individuals therefore provides clues about whether a common underlying mechanism might underpin behavioural variation at different levels.

The behavioural character concept consequently comes with predictions about patterns of (co)variation between behavioural expressions of a character. First, each of the observed behaviours should show between-individual variation (i.e. non-zero repeatability) and part of this variation should be owing to individual differences in a latent variable, provided that the population harbours between-individual variation in the behavioural character. Second, behavioural expressions of the character should change in concert within the same individual in response to environmental change ('integration of plasticity'; [18]), provided that the behavioural character is plastic within individuals. Third, similar non-zero behavioural correlations are expected between versus within individuals provided that the character also varies at both levels. Fourth, correlations between expressed behavioural observables should be the same in different environments in which the character is expressed

(e.g. breeding versus non-breeding contexts), and significant cross-environment correlations should exist if the same mechanism (character) affected the expression of behavioural observables in different environments [11]. Finally, a character should be quasi-independent from other characters to respond to selection as a unit [6]. Behavioural expressions of a character should therefore show some degree of independence from other behavioural characters.

We illustrate our thesis by analysing four behaviours that great tit males use when confronted with a territorial intrusion. We tested the hypothesis that these four behavioural observables were expressions of the behavioural character 'aggressiveness'. To do so, each male was subjected to a 'standardized territorial intrusion' four times per breeding season (year): twice during the egg-laying period of its social mate, at which time intrusions should increase risk of paternity loss [19] and consequently elicit a relatively aggressive response and twice when its social mate was incubating the clutch, at which time intrusions should not increase perceived risk of paternity loss and consequently elicit less of an aggressive response. We tested whether the data supported the hypothesis that the behavioural observables were indeed expressions of the same character ('aggressiveness').

We performed a four-step data analysis: we first ran univariate analyses where, for each of the behavioural observables separately, we estimated the amount of variance between and within individuals, as well as the level of behavioural plasticity with respect to breeding context. We expected that all behavioural observables would have a repeatable component and show a plastic response to the relative perceived threat of the intruder, provided that they represented expressions of the same repeatable but plastic behavioural character. Therefore as the second step, we quantified correlations between the different behavioural observables, asking whether they were correlated as hypothesized at each hierarchical level (i.e. between and within individuals) and in each environment (i.e. during laying and incubation). The integration of behavioural observables across environments was investigated using a character state approach [20] and assessed by testing whether correlations within environments (breeding contexts) and across environments ('cross-environment' correlations; [20]) were consistent with the presence of a single common underlying mechanism. As the third step, we statistically evaluated the amount of support for the presence of a context-general latent variable. Finally, we asked whether 'aggressiveness' constituted a quasi-independent module by evaluating whether it was distinct from other presumed aspects of risk-taking behaviour, for example level of activity in a novel environment.

2. Material and methods

(a) Experimental protocol

We studied 12 nest box populations of great tits in southern Germany (for details, see electronic supplementary material, appendix S1). Simulated territorial intrusions (i.e. aggression tests) were performed in the breeding seasons of 2010–2012. A taxidermic mount of a male great tit was presented as a visual stimulus with a playback song as an acoustic stimulus (detailed below). In each year, each male was subjected to four aggression tests during its first breeding attempt (defined as attempts initiated within 30 days after the first egg of the year in all of the plots was

found; [21]). Each male was subjected to two simulated territorial intrusions during egg-laying (1 and 3 days after its first egg was observed) and two during incubation (1 and 3 days after clutch incubation was confirmed). Owing to logistical constraints, the interval between first and repeat trials within-breeding context was more than 2 days for 7% of the 1150 repeat tests.

Aggression tests were conducted between 7.00 and 12.00; the specific time was semirandomly assigned. The taxidermic mount was presented 1 m away from the subject's nest-box on a 1.2 m wooden pole protected by a green wire mesh (see the electronic supplementary material, figure S1). Fifteen mounts and 14 playback song stimuli (recorded from German and Dutch populations) were constructed, enabling us to test whether the assayed behaviours represented responses to great tit mounts and songs in general rather than responses to their specific characteristics [22]; one mount and one song (broadcasted with a Samsung U5 Digital Audio Player connected to a Radioshack Mini Amplifier) were randomly allocated to each test. One of 25 observers performed the observation at a distance of 15 m.

Following the onset of a focal test, we recorded the behaviour of the focal male for a period of 3 min after it had entered a 15 m radius around the box. The observer counted the number of calls and songs, estimated the minimum distance to the mount ('approach distance') and noted whether the subject attacked the mount (jumping on the wire mesh of the mount; 'occurrence of attack'). (Descriptive statistics of each observable (cf. mean, range and standard deviation) are given in the electronic supplementary material, table S1.) For ease of interpretation, approach distance was multiplied by -1 (i.e. higher values represented a more aggressive response) in all the statistical analyses. Subjects that did not arrive within 15 min were scored as non-responsive. We performed 657 tests in 2010, 652 in 2011 and 937 in 2012, reflecting yearly breeding densities. Male identity was known for 1593 tests; in 1285 (80%) of these tests, the male responded. Analyses were based on these 1285 aggression tests, representing 365 unique (i.e. ringed) males. The number of responsive tests varied between males depending on number of years present and number of responses: 10 tests (n males $= 1$), 9 ($n = 3$), 8 ($n = 11$), 7 ($n = 17$), 6 ($n = 22$), 5 ($n = 21$), 4 ($n = 80$), 3 ($n = 104$) 2 ($n = 66$), 1 ($n = 40$).

(b) Statistical analyses

(i) Univariate mixed-effect models

We modelled variation in each of the agonistic behaviours separately as a function of (fixed effects) breeding context (laying versus incubation), test sequence within-breeding context (first versus second trial), year (2010, 2011, 2012) and time of the day (measured as minutes after sunrise and expressed as the deviation from the average time of all tests). Random intercepts were included for the identity of the observer ($n = 25$ levels), population ($n = 12$), playback song ($n = 14$), taxidermic mount ($n = 15$) and subject male ($n = 365$). We used the following error structure: approach distance was square root transformed and modelled with Gaussian errors, number of songs and calls (untransformed) modelled with Poisson errors and occurrence of attack (yes/no) with binomial errors. Adjusted repeatabilities were subsequently calculated as the between-individual variance divided by the sum of the between-individual and the residual variance [23].

(ii) Multi-variate mixed-effect models

Between- and within-individual correlations were estimated by fitting the assayed behaviours (approach distance, calls, songs and occurrence of attack) as four response variables into a single multi-variate mixed-effect model with random intercepts for individual identity. Further fixed or random effects were not included because our univariate analyses revealed that their effects were of minor importance (see Results and table 1). Breeding context

Table 1. Sources of variation in four agonistic behaviours based on simulated territorial intrusion experiments applied to great tits in southern Germany. (Estimates were derived, separately for each agonistic behaviour, from univariate mixed-effect models with random intercepts for individual (1–365), population (1–12), observer (1–25), taxidermic model (1–15) and playback song identity (1–14). Breeding context (laying versus incubation), test sequence within-breeding context (first versus second), time of day and year (2010, 2011, 2012) were fitted as fixed effects (n = 1285 tests). We give point estimates for each fixed (β; mean) and random (σ^2; variance) parameter, as well as adjusted repeatabilities, with their 95% CI.)

	calls	approach distance[a]	occurrence of attack	songs
fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
intercept[b]	−1.05 (−2.00, −0.38)	−2.49 (−2.65, −2.14)	−1.89 (−3.54, −0.56)	1.92 (1.68, 2.16)
breeding context	−1.96 (−2.36, −1.44)	−0.61 (−0.71, −0.47)	−2.37 (−3.04, −1.49)	0.47 (0.34, 0.59)
sequence	−0.43 (−0.94, −0.07)	−0.10 (−0.21, 0.03)	−0.33 (−0.85, 0.26)	0.11 (0.00, 0.25)
time of day	−0.23 (−0.51, −0.02)	0.07 (−0.01, 0.12)	0.00 (−0.01, 0.00)	−0.01 (−0.08, 0.05)
year 2011	1.06 (0.39, 1.88)	0.11 (−0.15, 0.30)	0.69 (−0.35, 1.44)	−0.11 (−0.29, 0.16)
year 2012	1.10 (0.33, 1.78)	−0.05 (−0.33, 0.16)	0.02 (−1.28, 0.59)	−0.21 (−0.41, 0.06)
random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
individual	3.40 (2.13, 4.95)	0.40 (0.32, 0.56)	2.20 (0.01, 5.43)	0.40 (0.26, 0.51)
population	0.01 (0.00, 1.52)	0.00 (0.00, 0.08)	0.01 (0.00, 0.48)	0.00 (0.00, 0.09)
observer	0.00 (0.00, 0.28)	0.04 (0.00, 0.09)	0.00 (0.00, 0.32)	0.01 (0.00, 0.07)
model	0.00 (0.00, 0.37)	0.00 (0.00, 0.03)	0.00 (0.00, 0.67)	0.00 (0.00, 0.02)
song	0.06 (0.00, 0.12)	0.00 (0.00, 0.01)	0.00 (0.00, 0.17)	0.00 (0.00, 0.04)
residual[c]	10.19 (8.55, 12.06)	1.10 (1.00, 1.20)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
repeatability	r (95% CI)	r (95% CI)	r (95% CI)	r (95% CI)
	0.21 (0.14, 0.29)	0.29 (0.22, 0.35)	0.38 (0.13, 0.61)	0.25 (0.20, 0.32)

[a]Approach distance was multiplied by −1 prior to analysis.
[b]Reference categories for fixed effects were set to 'laying' (breeding context), '1st' (sequence), 2010 (year) and population mean time of the day.
[c]Residual error distributions were binomial (occurrence of attack), Gaussian (approach distance) or Poisson (calls, songs).

strongly affected all of the behaviours (table 1) but was not included in the model because we wanted the within-individual covariance matrix to capture all sources of within-individual plasticity. Behaviour-specific error structure was applied as detailed above. Notably, the within-individual variance of 'occurrence of attack' was fixed to one because it is not estimable for binary data [13]; within-individual correlations with this variable should consequently be treated with caution. Exclusion of this response variable did not change our general findings (see electronic supplementary material, table S2b).

Within- and cross-breeding context correlations were estimated at the between-individual level by treating each of the four behavioural observables as a distinct response variable for each breeding context (e.g. 'songs during laying' and 'songs during incubation'), resulting in a multi-variate mixed-effect model with eight response variables and random intercepts for individual identity. We consequently estimated, within the same model, between-individual correlations within *and* across breeding contexts. This model estimated 28 between-individual correlations (six within-context correlations among all four behavioural observables × 2 contexts + 16 across-context correlations). Within-individual cross-context covariances were non-estimable (because the two breeding contexts cannot be experienced at the same time) and were therefore constrained to zero [17]; further fixed or random effects were not included (detailed above).

To assess whether the behaviours were correlated as expected according to the behavioural character concept, we compared the similarity between the posterior distributions (defined below) of pairwise correlations between versus within individuals and between laying versus incubating, using the 'overlapping coefficient' [24]. We further applied Mantel tests to assess whether the two matrices differed in correlation structure.

(iii) Structural equation modelling

We applied structural equation modelling (a statistical technique that includes confirmatory factor analysis as a special case) to the between-individual covariance matrix derived from the mixed-effect model with eight response variables (detailed above). We evaluated relative support for each of four *a priori* considered scenarios (based upon their relative AIC-values): (i) the absence of any latent variable (figure 2a); (ii) the presence of a single latent variable affecting all behaviours in both contexts (figure 2b); (iii) the presence of two context-specific latent variables (figure 2c) and (iv) the presence of two correlated but context-specific latent variables (figure 2d).

(iv) Quasi-independence of behavioural modules

We tested for quasi-independence of the hypothesized aggressiveness module by assessing whether the four agonistic behaviours (occurrence of attack, approach distance, calls and songs) were correlated with another observed behaviour, the individual's level of activity when placed into a novel environment (see [25] and electronic supplementary material, appendix S2). We estimated the correlations between the four hypothesized behavioural expressions of the character aggressiveness and activity in a novel environment by fitting them all as response variables into a multi-variate mixed-effect model with random intercepts for individual identity (1–277).

(v) Parameter estimation methods

We used R statistical environment v. 3.0.2 for all statistical analyses [26]. Mixed-effect models were fitted using Monte Carlo Markov chains in the MCMCglmm package [27], which retrieves posterior distributions of estimated parameters. We subsequently
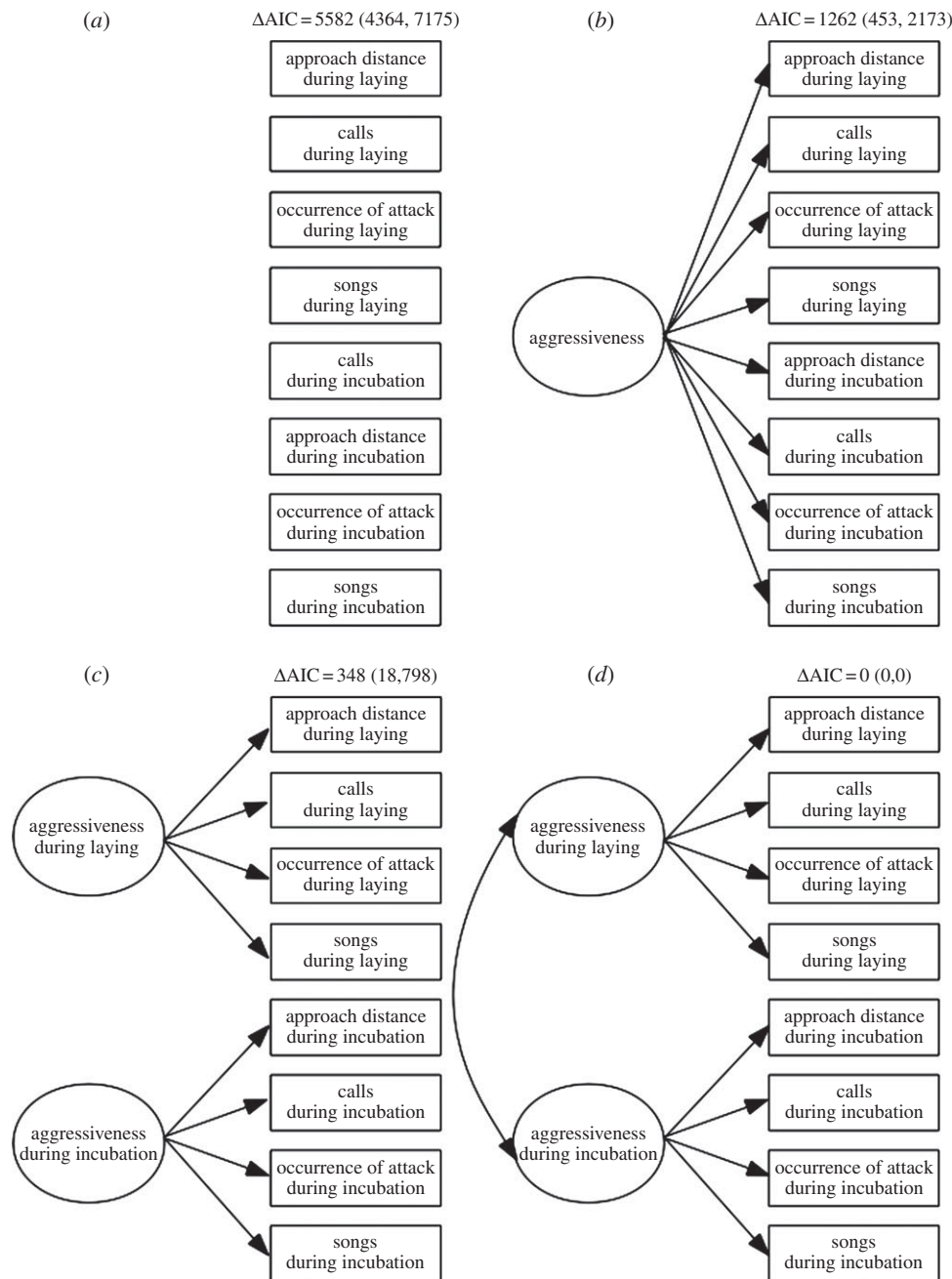
Figure 2. Four models (hypotheses) explaining covariance structure among agonistic behaviours assayed during laying and incubation in wild great tits. Model (a) proposes a scenario where each combination of observables and breeding stage is underpinned by a separate factor (the null model); model (b) hypothesizes a common factor ('module') underpinning all observables regardless of breeding context, whereas model (c) hypothesizes a separate module for each breeding context; model (d) expands upon this scenario by hypothesizing that those modules are themselves submodules influenced by a common factor.

calculated the mode and 95% credible interval (CI) for each parameter. This Bayesian approach allows for uncertainty to be appropriately carried forward to follow-up analyses [28]. Structural equation models were fitted with the 'sem' package [29]. Model implementation and procedures used for taking forward uncertainty from one analysis to the next are detailed in the electronic supplementary material, appendix S3.

3. Results

(a) Sources of variation in behavioural observables

A substantial part of phenotypic variation in each of the observed agonistic behaviours was explained by differences between individuals. CIs for repeatability were never close to zero, implying strong support for the presence of between-individual variation. Adjusted repeatability ranged between

0.21 and 0.38 (table 1). All behavioural observables changed with breeding context (table 1): individuals produced more calls, sang less, approached closer and were more likely to attack during laying compared with incubation (see electronic supplementary material, table S1). Effects of time of day, test sequence or year were not supported, except for calls that differed among years and decreased with time of day and sequence within-breeding context (table 1). The identity of the observer, mount or playback song explained little variation, if any at all (table 1).

(b) Between- versus within-individual correlations

Our mixed-effect model with four response variables (see Material and methods) provided strong support for non-zero correlations among all behavioural observables at the
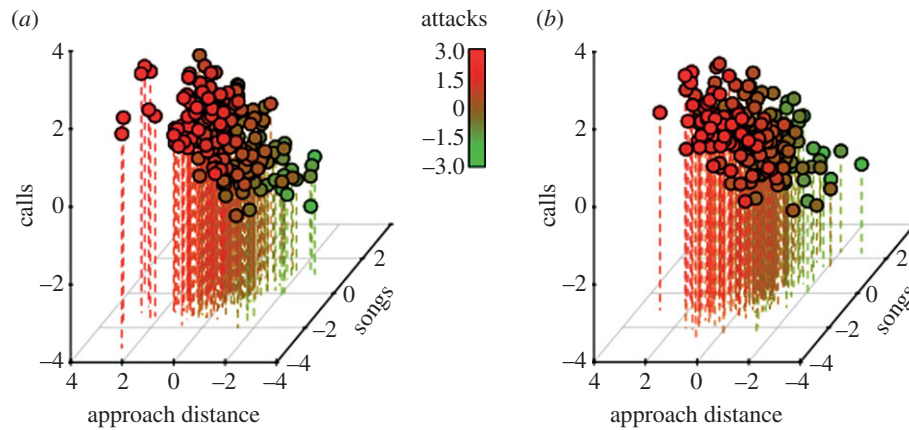
Figure 3. (a) Visual representation of between-individual correlations (plotted are the individuals' average values as deviations from the population mean value) and (b) within-individual correlations (plotted are the observations represented as deviations from individual mean values) between agonistic behaviours observed in wild great tits in Bavaria (Germany). Values were simulated from the correlation matrix estimated from a multi-variate mixed-effect model with random intercepts for individual identity (1–365) and calls, approach distance, occurrence of attack and songs fitted as response variables. We plot here predicted values on their untransformed latent scale.

between-individual level (figure 3a; electronic supplementary material, table S2). Individuals that on average (across all observations) approached the mount relatively closely also called at relatively high rates, produced fewer songs and were more likely to attack the model compared with individuals that on average did not approach closely. Within-individual correlations showed the same pattern (figure 3b; electronic supplementary material, table S2): during observations where an individual approached the dummy relatively closely, it would also call relatively much but sing relatively little compared with observations of the same individual where it approached less closely. These findings imply that the assayed agonistic behaviours changed in concert as hypothesized. Posterior distributions of pairwise correlations within versus between individuals overlapped substantially (see electronic supplementary material, table S2), providing strong support for the hypothesis that behavioural correlations did not differ between levels. This was confirmed by matrix-wide statistical comparisons (Mantel test: r (95% CI) = 0.88 (0.76–0.96)). Taken together, these findings support the hypothesis that the same latent variable (character) affected the expression of the agonistic behavioural observables within versus between individuals.

(c) Between-individual correlations within- versus across-breeding contexts

Signs and magnitudes of between-individual within-breeding context correlations were very similar for the two breeding contexts (table 2b): posterior distributions overlapped considerably (see electronic supplementary material, table S3). For example, the correlation between calls and approach distance was (point estimate (95% CI)) 0.67 (0.53, 0.77) during laying and 0.53 (0.35, 0.67) during incubation (overlap: 0.32). This similarity was confirmed by matrix-wide statistical comparisons (Mantel test: r (95% CI) = 0.98 (0.91–0.99)).

Most behavioural observables showed 'significant' positive between-individual cross-breeding context correlations (i.e. most CIs did not overlap zero; table 2a). In other words, individuals that had relatively high average values during laying also had relatively high average values during incubation, suggesting that the same behavioural observable was proximately underpinned by the same mechanism when expressed in different contexts. Upper CI

nevertheless never included 1.00 (calls: 0.72; approach distance: 0.60; occurrence of attack: 0.65; songs: 0.62), implying that their between-individual variances were also shaped—though only partly—by context-specific proximate factors [19]. Crossbreeding context correlations between different behavioural observables were of the same sign as their within-context counterparts, but the former correlations were less strong (table 2b), again suggesting some level of context-specific expression of between-individual variance (i.e. hierarchical structure) in the presumed behavioural character.

(d) Structural equation modelling

Cross-context correlations were substantial but their within-context counterparts were tighter (table 2), implying context-specific but correlated submodules affecting the expression of the behavioural observables (cf. model (d) in figure 2). Our comparison of four a priori considered structural equation models supported this interpretation: model (d) was the single best-supported model; the upper 95% CI of its ΔAIC value did not overlap with the lower 95% CI of other models (figure 2). The presumed context-specific submodules (cf. latent variables) were, as expected, positively correlated (r (95% CI): 0.59 (0.24, 0.78); figure 4).

(e) Quasi-independence of behavioural modules

The four observed agonistic behaviours were, as expected, not associated with activity in a novel environment. There was very little statistical support for correlations between observables that were a priori hypothesized expressions of the character 'aggressiveness' and activity in a novel environment. (see electronic supplementary material, table S4).

4. Discussion

This paper proposed an approach for the inclusion of 'labile characters' into the evolutionary character framework [3] and introduced a corresponding statistical methodology to test whether labile observables can be considered expressions of a hypothesized evolutionary character. We used the labile behavioural character 'aggressiveness' in great tits as a worked example to show that the character concept has novel

Table 2. Between-individual correlations (*r*) between four agonistic behaviours within- and across-breeding contexts (laying versus incubation). (Estimates were derived from a cross-environment multi-variate mixed-effect model where each of four agonistic behaviours (calls, approach distance, occurrence of attack and songs) was fitted as a separate response variable for each breeding context (i.e. eight response variables), with random intercepts for individual identity. (*a*) Between-individual correlations between the same agonistic behaviour across the two contexts; (*b*) between-individual correlations between two different agonistic behaviours within breeding contexts (i.e. both behaviour 1 and 2 are measured within the same context) and across breeding contexts (i.e. behaviour 1 is measured during laying and behaviour 2 instead during incubation or *vice versa*). We give point estimates for each parameter with their 95% CI.)

	within-breeding context correlations		cross-breeding context correlations	
context 1 – context 2 behaviour 1 – behaviour 2	laying – laying *r* (95% CI)	incubation – incubation *r* (95% CI)	laying – incubation *r* (95% CI)	incubation – laying *r* (95% CI)
(*a*) same behaviour				
calls – calls	—	—	0.58 (0.34, 0.72)	
approach – approach	—	—	0.51 (0.31, 0.60)	
attack – attack	—	—	0.34 (−0.14, 0.65)	
songs – songs	—	—	0.45 (0.32, 0.62)	
(*b*) different behaviours				
calls – approach	0.67 (0.53, 0.77)	0.53 (0.35, 0.67)	0.32 (0.12, 0.49)	0.34 (0.12, 0.50)
calls – attack	0.72 (0.51, 0.83)	0.67 (0.34, 0.88)	0.28 (−0.16, 0.61)	0.27 (0.03, 0.50)
calls – songs	−0.80 (−0.88, −0.71)	−0.73 (−0.82, −0.62)	−0.37 (−0.54, −0.20)	−0.73 (−0.83, −0.62)
approach – attack	0.86 (0.78, 0.91)	0.64 (0.46, 0.78)	0.34 (−0.03, 0.63)	0.44 (0.22, 0.61)
approach – songs	−0.54 (−0.67, −0.39)	−0.28 (−0.44, −0.11)	−0.18 (−0.32, 0.03)	−0.34 (−0.48, −0.10)
attack – songs	−0.56 (−0.68, −0.40)	−0.44 (−0.66, −0.17)	−0.15 (−0.32, 0.08)	−0.31 (−0.60, 0.08)

predictions that are empirically testable when applied to multi-level phenotypes. Explicit to this framework is (i) that characters should be defined *a priori* as latent variables that affect functionally correlated observables, (ii) that if just one observable was measured, it would not be possible to validate whether it did reflect the character of interest, (iii) that both variation between and within individuals should explicitly be acknowledged and incorporated and (iv) that functionally unrelated observables also need to be measured to test for the quasi-independence of an hypothesized character from other ones.

(a) Novel predictions

Our multi-level implementation of the character concept introduced novel predictions that concern specific variance components [17] of observables. First, all labile observables that are *a priori* hypothesized expressions of a labile character should logically contain between-individual variance if the character itself contains between-individual variation. A statistical outcome where some but not all hypothesized expressions of a character showed between-individual variation would suggest that the hypothesis was false. Second, all labile observables should respond in concert to variation in the environment if they belong to a functional unit [11]. If one of the observables would not show a plastic response to a specific environmental gradient while others did, they would not all be expressions of the same character. Third, labile observables should correlate similarly at all hierarchical levels at which the latent variable varied. In summary, the characteristic multi-level nature of labile characters will enable researchers to test predictions that have not previously been considered in evolutionary character theory. We applied this logic to the between versus within individual level, but it would equally apply to others (e.g. between versus within

populations; [30]). Assessment of similarity in between-individual correlation structure when comparing contexts (cf. laying versus incubation in our worked example) constitutes another test of the same idea.

(b) Empirical testing of predictions

In our worked example, we defined aggressiveness as a latent variable affecting the expression of behaviours used in aggressive interactions. We subsequently tested whether four behaviours used in aggressive interactions were indeed expressions of this labile character. We found between-individual differences in all assayed behaviours (table 1) that were partly attributable to hypothesized latent variables (figure 4). These observables were all plastic in a coordinated way as expected based upon level of intruder threat (effect of breeding context: table 1). This suggests a common underlying proximate mechanism that makes the aggressive display a functional unit. Patterns of correlation within and between individuals agreed with this interpretation: all expressed behaviours were associated, and in a very similar way, within and between individuals (figure 3). Sign and magnitude of the between-individual correlations also did not differ between breeding contexts (table 2), despite substantial levels of cross-context plasticity (table 1). Furthermore, an individual's typical value for a focal behaviour was repeatable across breeding contexts (i.e. cross-environment correlations were positive; table 2), supporting the notion of a common context-independent mechanism affecting all agonistic behaviours. At the same time, between-individual within-context correlations were somewhat tighter than their cross-context counterparts (table 2), implying partial context-specific modularity (figure 4). Those modules were positively correlated, implying that they represented submodules of an overarching context-independent latent
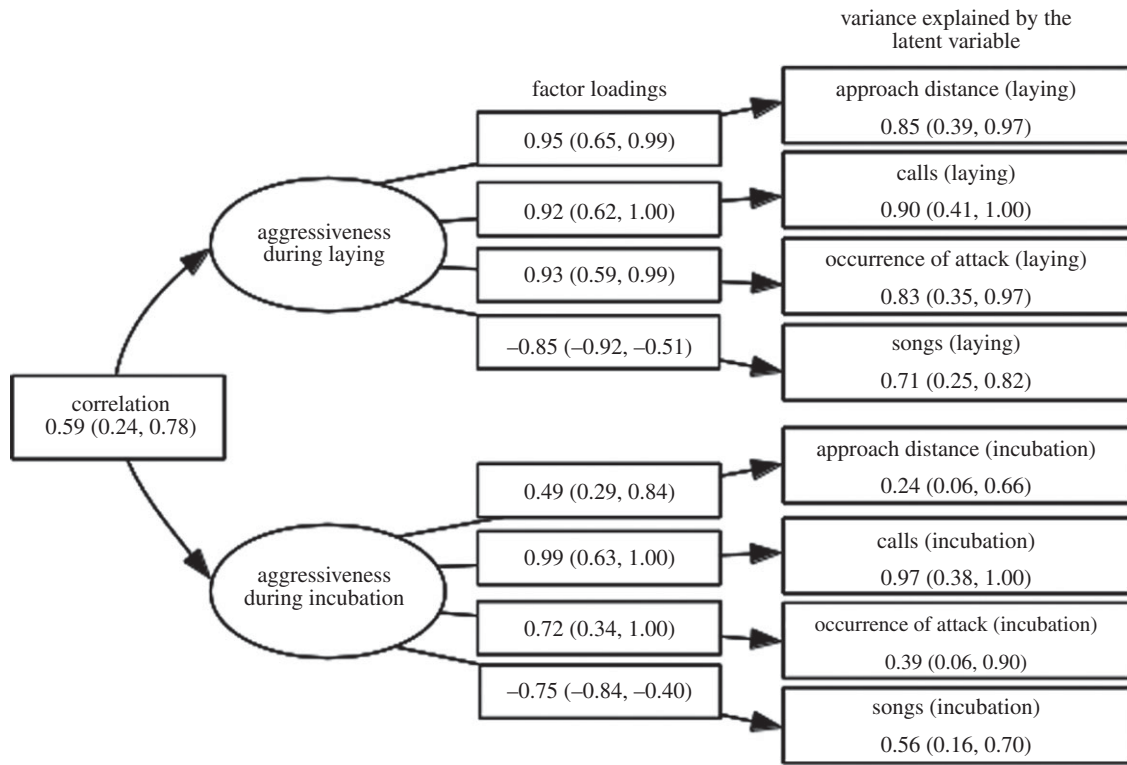
Figure 4. Parameter estimates of the structural equation model that best fitted our data.

variable affecting the aggressive display. This finding shows that the expression of the latent variable itself had a hierarchical structure, illustrating the level of detail about the structure of labile characters that can be derived by applying this framework. Furthermore, we showed that activity in a novel environment—a presumed observable expression of 'risk-taking behaviour' in non-social contexts [25]—was not significantly correlated to any of the agonistic behaviours (see electronic supplementary material, table S4), implying that the behavioural character 'aggressiveness' indeed represented a quasi-independent behavioural module at least with respect to this observed behaviour but potentially also from other risky behaviours in general.

(c) Why study behavioural characters?

The study of behaviour has a long history in fields of evolutionary biology (cf. animal behaviour and behavioural ecology), with research programmes focusing on a diverse array of topics, such as proximate causation, development and function of behaviour [31]. The proposed application of the evolutionary character concept in the study of behaviour will, in our opinion, greatly help researchers in deciding whether observed behaviours do or do not quantify the 'characters' that correspond to those for which adaptive theory has been developed. For example, theory predicts that between-individual variation in future fitness expectations can explain between-individual variation in 'risky behaviour' [32]. Tests of theory would involve manipulation of state-variables to quantify whether an individual's risky behaviour changed in the direction predicted by theory. However, the validity of the empirical test would hinge critically on whether the assayed behaviour did indeed represent a risky behaviour. Researchers may thus inappropriately interpret empirical tests of a given theoretical model because they did not measure the target character [33,34]. The usefulness of the proposed

framework is further illustrated by our empirical example: if we had only measured the amount of songs produced as a proxy of aggressiveness, we could have arrived at the conclusion that the more songs produced the more aggressive was the response. Our empirical example implied that more aggressive displays were, in contrast, characterized by a lower—not higher—number of produced songs (figure 4).

Other fields of biology have, notably, been pioneers in some elements of our proposed approach. Specifically, human personality psychology has a long history of focusing on latent variables in the study of behaviour [14], where techniques such as the 'multi-trait multi-method' approaches [35] are commonly used to examine the validity of measurements of latent variables. Nevertheless, key characteristics of behavioural characters, such as within-individual variation owing to adaptive responses to the environment (i.e. 'individual plasticity') and its multi-variate extension (i.e. 'integration of plasticity' [18]), are not fully embedded in human personality research. The treatment of within-individual variation as personality 'signatures' [36,37] in psychology does not, in our reading, appear to be based on evolutionary principles. By contrast, within-individual variation owing to adaptive individual plasticity represents a key concept in evolutionary biology [16]. A possible reason for this mismatch could be the prevailing type of experimental design in human psychology [37], where individuals (or their peers) are typically—though not always—subjected to questionnaires that asks about the subject's typical behaviour (i.e. average, long-term response) in a diverse range of (social and non-social) situations. Our proposed approach would instead require *repeated* exposure to the same questionnaire (over an environmental gradient), such that within- and between-individual (co)variances can be estimated explicitly. Fully integrating multi-level (co)variation in characterizing labile characters would, in our view, represent a very fruitful expansion in both evolutionary biology and human psychology research.

(d) The hierarchical structure of behavioural characters

The hierarchical nature of behaviour and other labile phenotypes represents a key aspect of the evolutionary character framework [3]. At the lowest level of the hierarchy are the behaviours observed in a particular context; those may represent expressions of a lower order character. Two of such lower order characters were evident in the great tit dataset (i.e. 'aggressiveness during laying' and 'aggressiveness during incubation'; figure 4). If such lower order characters represented evolutionary modules with partial—though not full—overlap in function, they should in turn be partly underpinned by a higher order (i.e. context general) character. Indeed, the positive correlation between the two context-specific latent variables supported the existence of such a higher order character (figure 4), which we might (objectively) call 'aggressiveness during the reproductive season'. One could readily extend this approach by including agonistic behaviours expressed in other contexts, for example those expressed outside the reproductive season (e.g. winter dominance interactions). This would yield insight in the generality versus (seasonal) specificity of aggressiveness as a character.

Inclusion of observed behaviours expressed in related but functionally distinct contexts would help to reveal the existence of higher order behavioural characters. For example, a higher order behavioural character representing 'willingness to take risk' might modulate lower order characters, such as aggressiveness, anti-predator boldness and exploratory tendency. An analogy in human psychology—a field that fully acknowledges hierarchical structuring—would be that 'orderliness', 'achievement striving' and 'cautiousness' are all part of a broad factor known as the personality axis 'conscientiousness' [38]. We illustrated this idea empirically by testing whether or not aggressiveness and activity in a novel environment were associated (see electronic supplementary material, table S4). This was not the case, implying that aggressiveness during the reproductive season was quasi-independent of activity in a novel environment, suggesting that the postulated higher order character did not exist in this case. Nevertheless, even if distinct modules (characters) would underpin behaviour in different functional contexts, correlations among them may be observed (cf. 'behavioural syndromes'; [39]). Functionally unrelated behavioural characters might also share proximate mechanism owing to the redundancy in expression pathways [6] resulting in an overarching modularity driven by constraints in the architecture of behaviour rather than functional coherence.

(e) The adaptive nature of behavioural characters

As detailed in this paper, the functional coherence that defines a 'behavioural character' comes with predictions about the (multi-level) structure of behavioural (co)variation. Implicit to the framework is also the adaptive nature of modules, an assumption that can be tested empirically. Specifically, if the organism indeed benefits from functional units in the execution of a particular task (e.g. grabbing objects in our example of the human hand), we explicitly expect natural selection to favour correlations ('correlational selection'; [40]) between the expressed observables (i.e. length of the five fingers). In the case of aggressiveness, we would thus expect strong correlational selection to act on the agonistic behaviours during egg production when ineffective displays, for example calling but not approaching, might have important fitness costs (for example, risk of paternity loss). While awaiting formal phenotypic selection analyses applied to our data, the structure of behaviour is in

line with this notion: our point estimates of behavioural correlations were tighter during egg laying compared with incubation (table 2), and the latent variable 'aggressiveness during laying' explained more variance in the agonistic behaviours than its counterpart during incubation (figure 4).

(f) Estimating behavioural character values

Researchers are continuously faced with the challenge of which behavioural data to incorporate in their analyses. What guidelines might one apply once the behavioural character concept has empirically been confirmed? We see two options. First, one could calculate a composite score derived from the structure of the latent variable. In the electronic supplementary material, appendix S4, we detail how an individual's score for the latent variable might be calculated (see also the electronic supplementary material, figure S2). This represents a more appropriate version of the traditionally recommended usage of composite scores from PCA within behavioural ecology [41], while having the advantage of (i) being able to deal with missing data [13] and (ii) avoiding failure to acknowledge the statistical non-independence of repeated measures data [42]. Unfortunately, the usage of such latent scores for further analyses is without doubt more complex and in some circumstances may demand large sample sizes [17,42]. Researchers might, therefore, alternatively use a single observable that closely predicts the behavioural character under study. Of course, such an approach would represent a less precise way of quantifying the character, but would also, logistically and technically, be less challenging. No matter which approach is chosen, it is important to acknowledge the distinction between behavioural characters and the behavioural observables. In some cases, the observable will accurately reflect the target behavioural character, though observables may represent expressions of multiple characters. Above all, we recommend that behavioural characters are defined explicitly in reference to a specific biological process and that behavioural observables should thus be labelled as objectively as possible. Doing so would help to avoid subjectivity in studying behavioural characters [43].

5. Conclusion

Our proposed framework attempts to unite advances in different fields of research in the study of characters. Our framework integrates cross-disciplinary research paradigms, including the study of latent variables in human psychology, the multi-level approach in the study of labile characters in behavioural and evolutionary ecology and the conceptualization of phenotypic organization in evolutionary biology. Such a holistic framework will enhance our ability to characterize the structure of behaviour, and other labile characters, and place it firmly in the realm of evolutionary biology.

10

rspb.royalsocietypublishing.org *Proc. R. Soc. B* **281**: 20132645

References

1. Gould SJ, Lewontin RC. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)

2. Pigliucci I, Kaplan I. 2000 The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. *Trends Ecol. Evol.* **15**, 66–70. (doi:10.1016/S0169-5347(99)01762-0)

3. Wagner GP. 2001 *The character concept in evolutionary biology*. San Diego, CA: Academic Press-Elsevier Science.

4. Duckworth RA. 2008 The role of behavior in evolution: a search for mechanism. *Evol. Ecol.* **23**, 513–531. (doi:10.1007/s10682-008-9252-6)

5. Wagner GP, Stadler PF. 2003 Quasi-independence, homology and the unity of type: a topological theory of characters. *J. Theor. Biol.* **220**, 505–527. (doi:10.1006/jtbi.2003.3150)

6. Wagner GP, Pavlicev M, Cheverud JM. 2007 The road to modularity. *Nat. Rev. Genet.* **8**, 921–931. (doi:10.1038/nrg2267)

7. Schwenk K. 2001 Functional units and their evolution. In *The character concept in evolutionary biology* (ed. G Wagner), pp. 156–198. San Diego, CA: Academic Press.

8. Wake MH. 2008 Organisms and organization. *Biol. Theory* **3**, 213–223. (doi:10.1162/biot.2008.3.3.213)

9. Susman RL. 1979 Comparative and functional morphology of hominoid fingers. *Am. J. Phys. Anthropol.* **50**, 215–236. (doi:10.1002/ajpa.1330500211)

10. Krebs JR. 1982 Territorial defence in the great tit: do residents always win? *Behav. Ecol. Sociobiol.* **11**, 185–194. (doi:10.1007/BF00300061)

11. Pigliucci M. 2003 Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecol. Lett.* **6**, 265–272. (doi:10.1046/j.1461-0248.2003.00428.x)

12. Houle D. 2001 Characters as the units of evolutionary change. In *The character concept in evolutionary biology* (ed. GP Wagner), pp. 109–140. San Diego, CA: Academic Press.

13. Gelman A, Hill J. 2007 *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

14. Bollen KA. 2002 Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.* **53**, 605–634. (doi:10.1146/annurev.psych.53.100901.135239)

15. Dingemanse NJ, Kazem AJN, Réale D, Wright J. 2010 Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* **25**, 81–89. (doi:10.1016/j.tree.2009.07.013)

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

17. Dingemanse NJ, Dochtermann NA. 2013 Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**, 39–54. (doi:10.1111/1365-2656.12013)

18. Pigliucci M. 2001 Characters and environments. In *The character concept in evolutionary biology* (ed. G Wagner), pp. 365–390. San Diego, CA: Academic Press.

19. Birkhead T, Atkin L, Møller A. 1987 Copulation behaviour of birds. *Behaviour* **101**, 101–138. (doi:10.1163/156853987X00396)

20. Roff DA. 1997 *Evolutionary quantitative genetics*. New York, NY: Chapman & Hall.

21. Nicolaus M, Tinbergen JM, Bouwman KM, Michler SPM, Ubels R, Both C, Kempenaers B, Dingemanse NJ. 2012 Experimental evidence for adaptive personalities in a wild passerine bird. *Proc. R. Soc. B* **279**, 4885–4892. (doi:10.1098/rspb.2012.1936)

22. Hulbert SH. 1984 Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**, 187–211. (doi:10.2307/1942661)

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

24. Inman H, Bradley E. 1989 The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Stat. Theory Methods* **18**, 3851–3874. (doi:10.1080/03610928908830127)

25. Stuber EF, Araya-Ajoy YG, Mathot KJ, Mutzel A, Nicolaus M, Wijmenga JJ, Mueller JC, Dingemanse NJ. 2013 Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behav. Ecol.* **24**, 1092–1098. (doi:10.1093/beheco/art035)

26. R Core Team. 2012 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

27. Hadfield JD. 2010 MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *J. Stat. Softw.* **33**, 1–36.

28. Ellison AM. 2004 Bayesian inference in ecology. *Ecol. Lett.* **7**, 509–520. (doi:10.1111/j.1461-0248.2004.00603.x)

29. Fox J. 2006 Structural equation modeling with the sem package in R. *Struct. Equ. Model.* **13**, 465–486. (doi:10.1207/s15328007sem1303_7)

30. Armbruster W, Schwaegerle K. 1996 Causes of covariation of phenotypic traits among populations. *J. Evol. Biol.* **9**, 261–276. (doi:10.1046/j.1420-9101.1996.9030261.x)

31. Owens IPF. 2006 Where is behavioural ecology going? *Trends Ecol. Evol.* **21**, 356–361. (doi:10.1016/j.tree.2006.03.014)

32. Wolf M, van Doorn GS, Leimar O, Weissing FJ. 2007 Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**, 581–584. (doi:10.1038/nature05835)

33. Watanabe NM, Stahlman WD, Blaisdell AP, Garlick D, Fast CD, Blumstein DT. 2012 Quantifying personality in the terrestrial hermit crab: different measures, different inferences. *Behav. Process.* **91**, 133–140. (doi:10.1016/j.beproc.2012.06.007)

34. Carter AJ, Marshall HH, Heinsohn R, Cowlishaw G. 2012 How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. *Anim. Behav.* **84**, 603–609. (doi:10.1016/j.anbehav.2012.06.015)

35. Campbell DT, Fiske DW. 1959 Convergent and discriminant validation by the multitrait-mulit method matrix. *Psychol. Bull.* **56**, 81–105. (doi:10.1037/h0046016)

36. Mischel W, Shoda Y. 1995 A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychol. Rev.* **102**, 246–268. (doi:10.1037/0033-295X.102.2.246)

37. Penke L, Denissen J, Miller G. 2007 The evolutionary genetics of personality. *Eur. J. Pers.* **587**, 549–587. (doi:10.1002/per.629)

38. Jang KL, McCrae RR, Angleitner A, Riemann R, Livesley WJ. 1998 Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *J. Pers. Soc. Psychol.* **74**, 1556–1565. (doi:10.1037/0022-3514.74.6.1556)

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

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

41. McGregor PK. 1992 Quantifying responses to playback: one, many, or composite multivariate measures? In *Playback studies of animal communication* (ed. PK McGregor), pp. 79–96. New York, NY: Plenum Press.

42. Budaev SV. 2010 Using principal components and factor analysis in animal behaviour research: caveats and guidelines. *Ethology* **116**, 472–480. (doi:10.1111/j.1439-0310.2010.01758.x)

43. Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2012 Animal personality: what are behavioural ecologists measuring? *Biol. Rev. Camb. Philos. Soc.* **88**, 465–475. (doi:10.1111/brv.12007)