Covariance reaction norms:
A flexible method for estimating complex environmental effects on trait (co)variances

Jordan Scott Martin[1]

jordanscott.martin@eawag.ch

[1]Fish Ecology and Evolution
Eawag, Swiss Federal Institute of Aquatic Science and Technology
Seestrasse 79, CH-6047 Kastanienbaum, Switzerland

Abstract

Estimating quantitative genetic and phenotypic (co)variances is crucial for investigating evolutionary ecological phenomena such as developmental integration, life history tradeoffs, and niche specialization, as well as for describing selection and predicting multivariate evolution in the wild. While most studies assume (co)variances are fixed over short timescales, environmental heterogeneity can rapidly modify the variation of and associations among organisms' traits. Here I synthesize prior random regression and double hierarchical animal models to develop a novel covariance reaction norm (CRN) model for detecting how trait (co)variances respond to complex (i.e., continuous, multivariate, and potentially nonlinear) environmental change, even in the absence of repeated individual measurements or experimental breeding designs. After introducing the CRN model, I validate its implementation in Stan, demonstrating unbiased Bayesian inference. I then apply the model to long-term field data on cooperation among meerkats (*Suricata suricatta*). I find nonlinear effects of group size on the genetic (co)variances of cooperative behaviors, leading to increased social niche specialization among foraging and pup feeding versus babysitting tasks in larger groups. Multivariate gene-by-environment interactions are also observed in response to age, sex, and dominance status. R code and a tutorial are provided to aid empiricists in applying CRN models to their own datasets.

Keywords: GxE, PxE, plasticity, context-dependent, social evolution, eco-evo

Introduction

Accurately estimating phenotypic and quantitative genetic (co)variances is essential for understanding multivariate evolution in the wild. For instance, quantifying the (co)variances of thermoregulatory traits and growth rates is crucial for explaining differential patterns of population adaptation and divergence in response to climate change (de la Mata et al., 2022; Oomen & Hutchings, 2022; Schaum et al., 2022). Empirical estimates of covariance between life history traits are also critical for testing theoretical models of putative tradeoffs (negative covariances) between growth, maintenance, survival, or reproduction (Haave-Audet et al., 2022; Chang et al., 2023), which are hypothesized to constrain the direction and rate of adaptive evolution (Stearns, 1989; Roff, 1996). Positive genetic covariances may instead accelerate adaptation across environments, such as in red flour beetles (*Tribolium castaneum*), where selection for drought resistance has been found to indirectly select for greater heat resistance via a correlated genetic response (Koch et al., 2020). Estimating phenotypic (co)variances is similarly important for addressing various challenges in evolutionary ecology, such as distinguishing between repeatable and stochastic patterns of trait selection in the wild (Damián et al., 2020; Dingemanse et al., 2021; Martin, 2021), testing theoretical models of developmental integration and niche specialization (Damián et al., 2020; Rolian, 2020; Martin et al., 2023), as well as for making evolutionary predictions in systems undergoing rapid environmental change or exhibiting processes of non-genetic inheritance, such as cultural learning and niche construction (Danchin & Wagner, 2010; Fogarty & Wade, 2022).

For polygenic and environmentally responsive traits, the quantitative genetic **G** matrix and phenotypic **P** matrix can be used to describe these multivariate (co)variances and predict their evolutionary consequences (Lande, 1979; Lande & Arnold, 1983). Various quantities derived from **G** and **P** matrices have also long been of interest in evolutionary genetics and ecology, such as covariance tensors and principal components (Schluter, 1996; Aguirre et al., 2014) for comparing divergence across populations (McGlothlin et al., 2018; Royauté et al., 2020), or canonical axes (Phillips & Arnold, 1989; Blows & Brooks, 2003) for describing (non)linear selection on correlated phenotypes (Nussey et al., 2007; Dingemanse & Dochtermann, 2013; Brommer et al., 2019). Multivariate, multilevel regression models (also known as mixed effects, hierarchical, or random regression models) are well-established in the literature and widely applied for empirically estimating **G** and **P** matrices (e.g. Nussey et al., 2007; Dingemanse & Dochtermann, 2013; Brommer et al., 2019). Multivariate animal models—a specific form of generalized multilevel regression model—are particularly useful for quantitative genetic analysis, as they can take full advantage of naturally occurring, continuous variation in genetic relatedness and environmental conditions across subjects (Kruuk, 2004; Wilson et al., 2010). This allows the animal model to provide greater flexibility and robustness for describing heritable (co)variation in wild populations, in comparison to classical methods that rely on the assumptions of balanced breeding experiments or specific kin-class comparisons (Kruuk & Hadfield, 2007). Building on the well-established animal model, the present paper develops flexible extensions for predicting variation in **G** and **P** matrices attributable to continuous, nonlinear, and multivariate environmental effects.

Motivation for a novel method

Despite longstanding theoretical interest in and empirical evidence for the micro- and macroevolutionary stability of **G** and **P** matrices (Björklund, 1996; Estes & Arnold, 2007; Henry & Stinchcombe, 2023; McGlothlin et al., 2018), genetic and phenotypic (co)variances can also change rapidly across space and time, as individuals face continuously varying environmental conditions that predictably shape the expression and selection of their traits (**Fig. 1**). For example, previous research across a wide range of taxa (e.g. lizards, Yewers et al., 2017; Wittman et al., 2021; flies, Carvalho &

Mirth, 2015; frogs, Lofeu et al., 2017; mice, vom Saal, 1979; Huber et al., 2017; and primates, Montoya et al., 2013; Grebe et al., 2019) has shown that endocrine activity and the resulting hormonal milieu experienced during both prenatal and postnatal development exhibit dose-dependent effects on the integration (positive genetic covariance) of various morphological and behavioral phenotypes in adult organisms (**Fig. 1a**). As another example, consider that classic theoretical models (van Noordwijk & de Jong, 1986) predict associations among life history traits to be contingent on the relative importance of among-individual differences in resource acquisition versus allocation. As a consequence, spatial or temporal heterogeneity in factors such as resource availability are expected to cause continuous variation in the genetic effects acting to constrain (negative genetic covariance, i.e. tradeoffs) or facilitate (positive genetic covariance) ongoing adaptation (Mats Björklund, 2004; Mats Björklund & Gustafsson, 2015; Haave-Audet et al., 2022); **Fig. 1b**). Similarly, continuous fluctuations in selection are expected to occur when the fitness effects of quantitative traits vary across functional contexts, as described by changes in the covariance between relative fitness and phenotype (Russell Lande, 1976). In many fish, for instance, large body size reduces predation risk and promotes greater mating and reproductive success (Barneche et al., 2018; Uusi-Heikkilä, 2020); however, commercial harvesting of fish also tends to target larger individuals (Sharpe & Hendry, 2009; Heino et al., 2015), facilitating continuous shifts in the strength and direction of selection on size as a function of the intensity of local harvesting (**Fig. 1c**). Both theory (Bonner, 2004; Jeanson et al., 2007) and extensive empirical study (e.g. Karsai & Wenzel, 1998; Thomas & Elgar, 2003; Ferguson-Gow et al., 2014; Ulrich et al., 2018) have also demonstrated that division of labor can emerge spontaneously during colony growth in eusocial species, with workers exhibiting generalist phenotypes at small group sizes (average positive phenotypic covariance among tasks) but shifting toward specialist phenotypes as group size increases (negative phenotypic covariance; **Fig. 1d**). Each of these specific cases is likely subject to further multivariate environmental interactions, due to e.g. antagonistic effects among hormones (Trumble et al., 2015; Qi et al., 2019), feedbacks between resource availability and competition (Lankau, 2011; Koutsidi et al., 2024), fluctuating selection on body size as a function of local sex ratios and predator densities (Uusi-Heikkilä, 2020; Jusufovski & Kuparinen, 2020), as well as the role of colony age structure in shaping division of labor (Huang & Robinson, 1996; Enzmann & Nonacs, 2021).

These dynamic and multivariate patterns of genotype-by-environment (GxE), phenotype-by-environment (PxE), and fitness-by-environment interaction can be formally quantified by changes in **P** and **G** matrices across contexts. Current multivariate animal models are particularly well suited for characterizing discrete changes in trait (co)variances due to categorical environmental effects, such as experimental conditions (e.g., solitary versus group housing) and developmental stages (e.g. juvenile versus adult) or discretely binned environmental covariates from the field (e.g. high versus low quality habitats). This is typically achieved through a so-called character state approach, where separate models are fit for trait expression in each discrete environmental state and individuals' additive genetic (breeding) values are allowed to correlate across models (Via & Lande, 1985; Lynch & Walsh, 1998). However, as argued above, environmental effects on **P** and **G** matrices will often reflect continuous, multivariate, and potentially nonlinear processes that are challenging to describe with character state models (**Fig. 1, 2a**). These complex dynamics can be interpolated post-hoc from estimates across discrete states (see Mitchell & Houslay, 2021 for a detailed treatment). However, this strategy will often require prohibitively large sample sizes for accurate inference of complex environmental effects, due to discretizing the problem into at least $k = s\frac{p(p+1)}{2}$ distinct and independently estimated (co)variance terms, where p is the number of phenotypes and s is the number of states necessary to effectively approximate the underlying function (which may be very large for multivariate environments, **Fig. 2a**). When appropriate data is available, heritable variation

in plasticity due to GxE effects can also be quantified. This requires specifying $k = \frac{sp(sp+1)}{2}$ genetic covariances between character states across environments in a full model. Genetic correlations < 1 across environmental states usually indicate heritable variation in plasticity due to GxE interactions (Mitchell & Houslay, 2021). Consequently, while the character state model is extremely useful for systems experiencing a small number of environmental states, it will tend have reduced statistical power for detecting complex functional relationships in more heterogeneous environments. Outside of controlled experiments, artificial binning of naturally occurring continuous variation will reduce statistical power for detecting true effects, while also increasing the risk of false positives and downwardly biasing effect sizes (e.g. Cohen, 1983; MacCallum et al., 2002). Qualitative inferential biases can also arise from insufficient sampling of discrete states in the presence of nonlinear and/or multivariate environments (**Fig. 2a**).

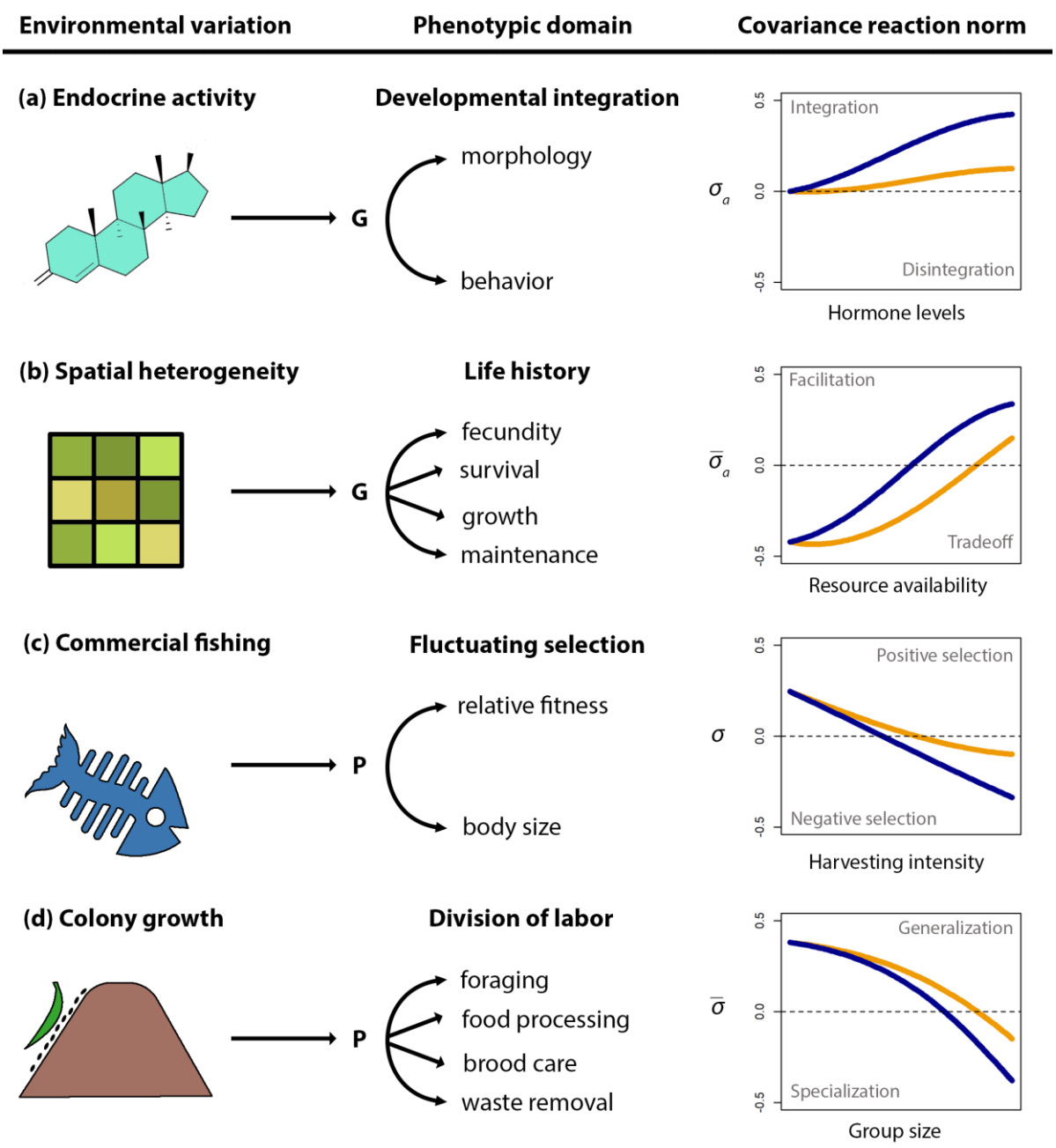
Mathematically complementary reaction norm models (de Jong, 1995; Lynch & Walsh, 1998; Nussey et al., 2007) can be used to more directly and parsimoniously describe such continuous processes, taking full advantage of available environmental information with much fewer parameters. Multilevel models with random individual intercepts and slopes (or at any other hierarchical level of interest) are often termed random regression models in biology (Henderson, 1982), and they provide one common and well-established approach to the estimation of reaction norms, including continuous patterns of GxE and PxE under specific study designs. For instance, when experimental breeding is used to observe relatives across a continuous environmental gradient, such as in a full-sib, half-sib design with dams nested in sires (Falconer & Mackay, 1996), a random regression animal model can be used to estimate genetic slopes quantifying how character state (co)variances continuously change across the distinct environments experienced by siblings. However, these breeding designs may only be practical for a subset of species with desirable properties for experimental study, such as relatively small body sizes, short life spans, sessility or small home ranges, and simple mating systems, or those with extensive infrastructure and resource investment due to their role in biomedical, agricultural, or livestock applications. Given the large sample sizes necessary to achieve appropriate balancing of relatives across multivariate environments, these designs also generally rely on discretization of the environment or manipulation of a single environmental gradient, greatly simplifying the ecological reality experienced by natural populations. It is, therefore, unfeasible to use this as a general approach for studying multivariate patterns of GxE, which are likely to occur for many labile behavioral, physiological, and morphological traits (**Fig. 2b**). Indeed, many of the most pertinent multivariate causes of GxE and PxE relevant for explaining development and adaptation in contemporary populations may simply be unfeasible and/or unethical to experimentally control, such as the interacting effects of predation risk, resource scarcity, climate change, and anthropogenic disturbance on wild populations.

Random regression models can also be applied in the absence of appropriate breeding designs when many repeated individual-level measurements are available (Nussey et al., 2007). For instance, consider a scenario where the genetic or phenotypic (co)variance between behavior and morphology increases as function of age and local resource availability. A field study design allowing for repeated observations of the same individuals across ages and levels of resource availability could then be used to estimate a random regression model with individual intercepts and slopes, which could in turn be used to calculate continuous changes in phenotypic and/or genetic (co)variance between behavior and morphology across environments. However, doing so would rely on the assumption that the (co)variance between these random intercepts and slopes is itself constant across environments. If, for example, the variation of and correlation among individuals' intercepts and slopes also changes continuously as a function of age and resource availability, e.g. if younger individuals show more

variable and genetically integrated responses to local resource availability, a standard random regression model will not accurately predict the magnitude of GxE or PxE across environments. A typical solution in this case would be discretize age and estimate separate age class-specific (co)variances matrices of individuals' intercepts and slopes, falling prey to the same limitations of discretization discussed above for character state approaches. Discretization can be avoided using interaction effects, such as by estimating random slopes for the effect of age x resource availability on both behavior and morphology, but this strategy requires repeated sampling designs that will often be unrealistic and burdensome, particularly for field studies, when quantifying multivariate environmental causes of GxE and PxE (**Fig. 2b**). For instance, the (co)variance between behavior and morphology may also vary continuously as a function of interactions between age, body size, conspecific density, and resource availability. In the general case, a research team will need to collect sufficient repeated individual measurements to estimate $k = \frac{vp(vp+1)}{2}$ free parameters in a (co)variance matrix, where p is the number of traits and v is the number of individual-level parameters (intercepts and slopes) describing all environmental effects of interest. Such matrices can quickly grow quite large, even in simple cases such as a 2nd-order polynomial for two phenotypes, which requires estimating $k = 78$ free individual-level parameters (**Fig. 2b**). Statistically identifying and reliably estimating such large matrices of random slopes on high-order interactions will simply be unfeasible for most empirical datasets (Matuschek et al., 2017).

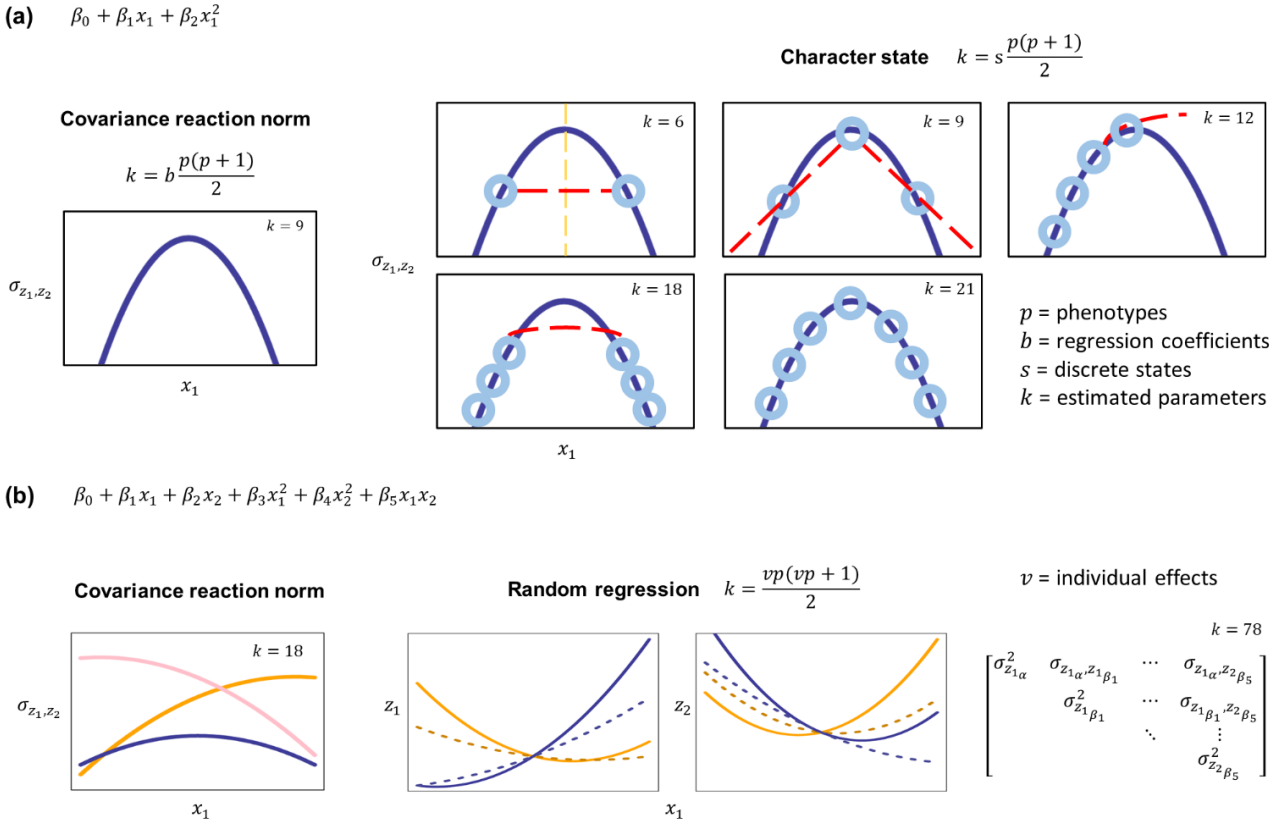
Overcoming the limitations discussed above will greatly improve empiricists' ability to understand complex environmental effects on the development and evolution of complex traits. Therefore, to address this challenge, I here introduce a 'covariance reaction norm' (CRN) approach for estimating continuous, multivariate, and potentially nonlinear environmental effects on trait (co)variances, building on and generalizing beyond standard models currently used in the literature for investigating GxE and PxE. This is accomplished by synthesizing character state and random regression approaches with a broader class of multilevel regression models, which includes so-called double hierarchical animal models as a special case. After formally outlining this CRN model, I subsequently validate this model for empirical application with simulation-based calibration (Talts et al., 2018), and then demonstrate its utility through a worked empirical example using long-term field data on cooperative behavior among meerkats (*Suricata suricatta*). Accompanying code and a guided tutorial for implementation of CRN models in the R statistical environment (R Core Team, 2023) using the Stan statistical programming language (Carpenter et al., 2017) can be found on Github (see **data availability**).

Figure 1. Examples of empirical applications for covariance reaction norm models.



Footnote. Four simplified examples (**a-d**) are shown of phenotypic domains (middle column) where continuous environmental variation (left column) is likely to cause continuous changes in quantitative genetic (**G**; top rows) and phenotypic (**P**; bottom rows) trait covariances, as formally described by hypothetical covariance reaction norms (CRNs; right column) quantifying patterns of continuous GxE and PxE across environmental states. Orange lines indicate potential interactions due to multivariate patterns of GxE and PxE, where the effect of one environmental gradient on trait (co)variation changes as a function of another environmental factor. See the main text for a detailed description of each scenario and Eq. 2-3 for a formal description of how such CRNs can be empirically estimated.

Figure 2. Challenges in estimating nonlinear and multivariate GxE interactions.



Footnote. Examples are shown of complex environmental effects on the covariance between two traits z_1 and z_2, demonstrating that even in simple cases the CRN model will generally require less free parameters k to accurately describe population patterns of GxE and PxE than standard approaches in the literature. (**a**) A nonlinear effect of a single continuous environment x_1 on the covariance between two traits, where $\sigma_{z_1,z_2} = \beta_0 + \beta_1 x_1 + \beta_2 x^2$. The k needed to detect this expected relationship, without prior knowledge of whether effects occur on trait variances or correlations, are shown for the CRN model (left) in comparison to a character state approach (right), where a varying number of discrete environmental states (light blue circles) are used to interpolate the underlying continuous function (dark blue curve). Red lines indicate biased interpolation resulting from insufficient sampling of the environment: discretizing to a high and low state (yellow line) results in detecting no change (top-left); sampling low, mid, and high results in failing to detect nonlinearity, under- or overpredicting change at different levels of the environment (top-center); failing to sample sufficiently high (or low) environments leads to predicting linear or monotonic change (top-right); and sampling only high and low environments leads to predicting a non-existent plateau (bottom-left). If sufficient sampling is done of the entire environmental range (bottom-center), the curve can be accurately interpolated, but at the cost of needing to independently estimate more than twice as many parameters as the CRN model. (**b**) A nonlinear interaction between two continuous environments x_1 and x_2, where $\sigma_{z_1,z_2} = \beta_0 + \beta_1 x_1 + \beta_2 x + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$. This requires $k = 18$ parameters to characterize with the CRN, assuming no prior knowledge. Interpolating such processes is very challenging with a character state approach but can be accomplished with a random regression model, where individual-level intercepts and slopes are estimated for both traits across environments. The solid and dashed lines show two individuals' hypothetical RNs for x_1 across two levels of x_2 (blue and orange). In this case, interpolating the population average function without prior knowledge requires over 4x as many parameters in comparison to the CRN.

Covariance reaction norms

Quantitative genetic analysis

The animal model is a multilevel regression model that allows for partitioning random quantitative genetic effects **G** and environmental effects on phenotypes. Extensive prior work has provided detailed overview of the animal model and its various extensions (e.g. Nussey et al., 2007; Wilson et al., 2010; Thomson et al., 2018; Martin & Jaeggi, 2022). Therefore, I focus herein on a highly simplified presentation of the animal model to highlight novel extensions, as well as to avoid detailed discussion of general issues in regression analysis such as the inclusion of various kinds of fixed and random effects. A multivariate animal model can be specified for each of *p* Gaussian phenotypes $\left[\boldsymbol{z_1^\top}, \ldots, \boldsymbol{z_p^\top}\right]^\top$ measured for *n* individuals by

$$\begin{bmatrix} g_{z_1}(\boldsymbol{z_1}) \\ \vdots \\ g_{z_p}(\boldsymbol{z_p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{\alpha_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{\alpha_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (1.1)$$

The functions $g_{z_1}, \ldots, g_{z_p}$ are link functions (e.g. identity, log, logit, atanh, sqrt) that can be used to appropriately specify non-Gaussian measurements on a latent linear scale. Linear predictors for these measurements are estimated with an *n* x *b* matrix **X** for *b* continuous and/or discrete covariates (e.g. local density, age, sex, resource abundance, seasonal precipitation and temperature, etc.), and $\left[\boldsymbol{\beta_1^\top}, \ldots, \boldsymbol{\beta_p^\top}\right]^\top$ are *b* x 1 vectors of trait-specific fixed effect sizes including global intercepts. After adjusting for these effects, the model estimates trait-specific additive genetic (breeding) values $\left[\boldsymbol{\alpha_1^\top}, \ldots, \boldsymbol{\alpha_p^\top}\right]^\top$ and residual environmental values $\left[\boldsymbol{\epsilon_1^\top}, \ldots, \boldsymbol{\epsilon_p^\top}\right]^\top$. Further genetic effects due to dominance or epistasis can also be parameterized when relevant for the goals of the analysis, along with any other random intercepts or slopes of interest. If repeated individual-level measurements are available, residuals can also be further partitioned into permanent and stochastic environmental components.

Trait (co)variances due to additive genetic and residual effects are assumed to be approximated by multivariate normal distributions

$$\begin{bmatrix} \boldsymbol{a_1} \\ \vdots \\ \boldsymbol{a_p} \end{bmatrix} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G} \otimes \boldsymbol{A}); \begin{bmatrix} \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{\epsilon_p} \end{bmatrix} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \quad (1.2)$$

With the **G** matrix being scaled using the Kronecker product $\otimes$ by a relatedness matrix **A** that quantifies pairwise relatedness among subjects, calculated using standard pedigree methods or molecular approaches. This basic animal model structure assumes that phenotypic (co)variances described by the **G** matrix are constant across subjects, adjusted for any other fixed and random effects predicting phenotypic means. The goal is now to relax this assumption by also allowing for fixed effects due to continuous or discrete environmental factors to also predict variation in trait (co)variances.

Predicting genetic (co)variances

The **G** matrix can be parameterized using genetic variances σ_a^2 and correlations r_a such that

$$\boldsymbol{G}: \begin{bmatrix} \sigma_{a_1}^2 & \cdots & \sigma_{a_1,p} \\ & \ddots & \vdots \\ & & \sigma_{a_p}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{a_1}^2 & \cdots & r_{a_1,p}\sigma_{a_1}\sigma_{a_p} \\ & \ddots & \vdots \\ & & \sigma_{a_p}^2 \end{bmatrix} \quad (1.3)$$

Here the genetic covariances $\sigma_{a_{1,p}} = r_{a_{1,p}}\sigma_{a_1}\sigma_{a_p}$ are given by the product of genetic correlations and standard deviations (square roots of the genetic variances). Note that bold symbols are used to distinguish vectors and matrices from scalars. Separating out the scale of variation σ_a^2 for each variable from their standardized associations r_a is crucial for further expanding the model, as environmental factors may exhibit independent effects on the variances and correlations of traits, which would otherwise be confounded together through direct prediction of the covariance. This parameterization also provides a straightforward solution to ensuring the positive definiteness of the G matrix during model estimation, as described further below (see *computational efficiency*).

With **Eq. 1.3**, the basic animal model can now be expanded to a covariance reaction norm (CRN) model by using link functions to predict how genetic variances and correlations change in response to the same matrix $\boldsymbol{X}$ of environmental covariates used to predict phenotypic means (or a relevant subset of these predictors). Using the subscript (X_n) to denote the **G** matrix predicted from a CRN in the environmental context measured for subject *n*

$$\begin{bmatrix} g_{z_1}(\boldsymbol{z_1}) \\ \vdots \\ g_{z_p}(\boldsymbol{z_p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{\alpha}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{\alpha}_{(X)_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} \boldsymbol{a}_{(X)_1} \\ \vdots \\ \boldsymbol{a}_{(X)_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(X)} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(X_n)}: \begin{bmatrix} \sigma^2_{a(X_n)_1} & \cdots & r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{a(X_n)_p} \end{bmatrix}$$

$$\begin{bmatrix} \log\left(\boldsymbol{\sigma^2_{a(X)_1}}\right) \\ \vdots \\ \log\left(\boldsymbol{\sigma^2_{a(X)_p}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\sigma^2_1}} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\sigma^2_p}} \end{bmatrix};\ \begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{r_{a(X)_{1,2}}}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{r_{a(X)_{p-1,p}}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{r_1}} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{r_{p-1,p}}} \end{bmatrix}$$

Rather than defining a single genetic variance and set of correlations for each response variable, as in the standard animal model (**Eq. 1**), the CRN animal model predicts *n* **G** matrices $\boldsymbol{G}_{(X)} = \left(\boldsymbol{G}_{(X_1)}, \ldots, \boldsymbol{G}_{(X_n)}\right)$ each composed of context-specific genetic variances $\boldsymbol{\sigma^2_{a(X)_p}} = [\sigma^2_{a(X_1)_p}, \ldots, \sigma^2_{a(X_n)_p}]'$, and correlations $\boldsymbol{r_{a(X)_{1,p}}} = [r_{a(X_1)_{1,p}}, \ldots, r_{a(X_n)_{1,p}}]'$. There are as many unique **G** matrices as the number of unique multivariate contexts defined by the environmental covariates in **X**, yet the prediction of these matrices only requires estimating a much smaller set of CRN parameters. The log and inverse hyperbolic tangent link functions are respectively used to infer these trait-specific parameters (additive fixed effects, including global intercepts) defined on the transformed linear scale of genetic variances $\left[\boldsymbol{\beta}^\top_{\boldsymbol{\sigma^2_1}}, \ldots, \boldsymbol{\beta}^\top_{\boldsymbol{\sigma^2_p}}\right]^\top$ and genetic correlations $\left[\boldsymbol{\beta}^\top_{\boldsymbol{r_{1,2}}}, \ldots, \boldsymbol{\beta}^\top_{\boldsymbol{r_{p-1,p}}}\right]^\top$. Note that the link function $\operatorname{atanh}(r) = \operatorname{logit}\left(\frac{r+1}{2}\right)/2$ extends the logit transformation defined for probability scale values to the scale of correlation coefficients. The variance and correlation parameters of the CRN may also include coefficients for more flexible non-parametric and generalized additive functions, such as splines or Gaussian processes (Pedersen et al., 2019; Riutort-Mayol et al., 2022), which are useful for capturing environmental effects such as spatiotemporal autocorrelation that are difficult to estimate with standard polynomials.

In the general case, there will be bp CRN parameters for genetic variances and $b\frac{p(p-1)}{2}$ parameters for the genetic correlations, where b is the number of columns in $\boldsymbol{X}$ (regression coefficients), resulting in $k = b\frac{p(p+1)}{2}$ total free parameters. In comparison to current methods, the CRN model is expected to greatly reduce the number of parameters required to estimate continuous

changes in trait (co)variances in the presence of nonlinear effects and multivariate interactions (**Fig. 2**). Given that $\boldsymbol{X}$ can include binary or categorical predictors, it is important to also note that the CRN straightforwardly generalizes the character state approach to more complex cases involving, for example, a combination of interacting continuous and discrete environmental factors.

Any non-zero fixed effects predicting $\boldsymbol{G}_{(\boldsymbol{X})}$ provide evidence for gene-by-environment (GxE) interaction. In general, however, direct interpretation of these CRN fixed effect sizes will be challenging due to the distinct scales of link functions used for genetic variances and correlations. Therefore, once the model is estimated, I encourage researchers to use model predictions from **Eq. 2** for more directly visualizing and quantifying total environmental effects on the more intuitive scales of genetic variances, correlations, and covariances, where $\sigma_{a(X_n)_{1,p}} = r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p}$. A worked example is provided below. When relevant, the same approach outlined above can be taken to predict continuous and/or discrete effects on residual or permanent environmental (co)variances.

Prediction of trait variances as a function of continuous and/or discrete variables is often called a double hierarchical model (Lee & Nelder, 2006; Rönnegård et al., 2010). The CRN can, therefore, be conceptualized as a form of double hierarchical animal model flexibly extended for multivariate prediction of both genetic variances and correlations. The term "double hierarchical" can be somewhat confusing, however, given that any distributional parameter could be modeled as a function of covariates, giving rise to the possibility of triple, quadruple, etc. hierarchical models of non-Gaussian responses. Therefore, I emphasize that the CRN is principally a multilevel model, as this is a more general class extending beyond the double hierarchical models applied in prior literature.

Random regression CRN

When repeated individual measures are available or a proper breeding experiment has been implemented, random individual-level slopes can be introduced to the model, so that the CRN describes changes in the (co)variances of the intercepts and slopes governing RNs of trait means. For instance, empiricists may be interested in testing theoretical predictions of how the genetic integration between individuals' mean trait value and plasticity to the environment changes across developmental or social contexts (Kraft et al., 2006; Stamps et al., 2018; Dingemanse et al., 2020; Bucklaew & Dochtermann, 2021; Martin et al., 2023). A random regression CRN can be implemented under a proper breeding design for detecting GxE and/or with repeated measurements, where individuals' breeding values for environmental slopes can be estimated from observations of related individuals' trait values across at least two or more environmental states. To do so, new vectors and matrices need to be introduced: *v*i x 1* vectors $\boldsymbol{u}$ for each phenotype containing *v* random effects (intercepts and slopes) for *i* individuals, and *n* x *v*i* block diagonal design matrices $\boldsymbol{W}$ indexing repeated measurements and scaling the *v* random effects for *i* individuals across *n* total measurements of each phenotype. Note that I use $\boldsymbol{W}$ rather than $\boldsymbol{Z}$ to avoid confusion of this random effect matrix with the vector of phenotypic measures $\boldsymbol{z}$. The random regression CRN is given by

$$\begin{bmatrix} g_{z_1}(\boldsymbol{z_1}) \\ \vdots \\ g_{z_p}(\boldsymbol{z_p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{Wu}_{(\boldsymbol{X})_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{Wu}_{(\boldsymbol{X})_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} \boldsymbol{u}_{(\boldsymbol{X})_1} \\ \vdots \\ \boldsymbol{u}_{(\boldsymbol{X})_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(\boldsymbol{X})} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(\boldsymbol{X_n})}: \begin{bmatrix} \sigma^2_{\alpha(X_n)_1} & \cdots & r_{a_{X_{I1}}(X_n)_1, b_{X_{Ib}}(X_n)_p}\sigma_{a_{X_{I1}}(X_n)_1}\sigma_{b_{X_b}(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{\beta_b(X_n)_p} \end{bmatrix}$$

Note that the design matrix $\boldsymbol{W} = \text{blockdiag}(\boldsymbol{X_{v1}}, \ldots, \boldsymbol{X_{vi}})$ is a block diagonal matrix containing repeated observations of individuals 1 to *i* from the subset of *v* columns in the full environmental matrix $\boldsymbol{X}$ over which individual intercepts $\boldsymbol{a}_{(X)_p}$ and slopes $\boldsymbol{\beta}_{\boldsymbol{1}(X)_p}, \ldots, \boldsymbol{\beta}_{\boldsymbol{v-1}(X)_p}$ are defined in the model for trait *p*. The process of prediction for the elements in $\boldsymbol{G}_{(X)}$ is equivalent to **Eq. 2**, though the total number of parameters to estimate in a full random regression CRN model expands to $k = b\frac{vp(vp+1)}{2}$, where b is the number of environmental CRN parameters and v is the number of individual effects (random intercept + $v - 1$ random slopes).

Phenotypic analysis

Empirical studies may lack the genetic information necessary to estimate **Eq. 2-3** or otherwise be principally interested in estimating phenotypic (co)variances. Without genetic data or repeated measurements, among- and within-individual patterns of phenotypic (co)variance will be confounded together, potentially biasing evolutionary predictions with measurement error and ephemeral environmental effects (Dingemanse et al., 2021; J. Martin, 2021). However, if multiple measurements are made on the same subjects across time, as with the random regression CRN introduced above, then repeatable among-individual differences in phenotype, due to both genetic variation and permanent environmental effects, can be effectively partitioned from stochastic variation using individual-level random effects. **Eq. 3** can be straightforwardly modified to produce a phenotypic CRN, described by a simplified multivariate normal distribution

$$\begin{bmatrix} \boldsymbol{u}_{(X)_1} \\ \vdots \\ \boldsymbol{u}_{(X)_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{P}_{(X)}\right);\ \boldsymbol{P}_{(X_n)}: \begin{bmatrix} \sigma^2_{(X_n)_1} & \cdots & r_{(X_n)_{1,p}}\sigma_{(X_n)_1}\sigma_{(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{(X_n)_p} \end{bmatrix} \quad (4)$$

where the phenotypic random effects $\left[\boldsymbol{\mu}^{\top}_{(x)_1}, \ldots, \boldsymbol{\mu}^{\top}_{(x)_p}\right]^{\top}$ are now assumed to be independently distributed among individuals. As with the quantitative genetic model, $\boldsymbol{P}_{(X_n)}$ is a matrix of among-individual phenotypic (co)variances predicted in response to the environmental context of measurement *n* for subject *i*, as determined by CRN fixed effect parameters for phenotypic variances $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_1}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_p}}\right]^{\top}$ and correlations $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{r_{1,2}}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{r_{p-1,p}}}\right]^{\top}$ estimated on transformed scales, equivalently to **Eq. 2**. Any non-zero fixed effects predicting $\boldsymbol{P}_{(X)}$ provide evidence for phenotype-by-environment (PxE) interactions. See Bliard, Martin et al. (2024) for detailed discussion and applications of bivariate phenotypic CRNs to detect life history tradeoffs under multiple sampling regimes common in population ecology.

Statistical implementation

Bayesian inference in Stan

The CRN model (**Eq. 2-3**) cannot currently be estimated using standard statistical software packages for multivariate animal models and multilevel models more generally, due to a lack of in-built functionality for expressing elements of covariance matrices as generalized linear predictors. Fortunately, however, the extremely flexible Stan statistical programming language can be used to construct bespoke animal models of desired complexity within a Bayesian inferential framework, facilitating general estimation of CRNs models using cutting-edge Markov Chain Monte Carlo (MCMC) methods (Hoffman & Gelman, 2011; Nishio & Arakawa, 2019; Martin & Jaeggi, 2022). Detailed discussion of contemporary Bayesian statistics is beyond the scope of this paper. However, I encourage readers to consult some of the excellent primers available on Bayesian data analysis (e.g. Gelman et al., 2013, 2020; McElreath, 2020) for thorough introductions, including extensive tips and

suggestions for key decisions such as the choice of priors, model validation and comparison, variable selection, and the interpretation of posterior estimates. As a general rule of thumb, I suggest using weakly regularizing priors when estimating CRN models, to reduce the risk of inferential bias while promoting efficient model convergence (Lemoine, 2019; McElreath, 2020). Despite it still being common to see thinning of MCMC chains reported in the literature, note that this is generally unnecessary (Link & Eaton, 2011).

Computational efficiency

This subsection covers formal details on efficient implementation of CRN models in Stan, which can be safely overlooked by empiricists without impeding interpretation or practical implementation. Prediction of large covariance matrices is computationally burdensome in a Bayesian framework, even with the use of appropriately regularizing priors and efficient MCMC algorithms, because the probability of observing a permissible (i.e. positive-definite) covariance or correlation matrix declines rapidly with increasing dimensionality of the matrix (Dean & Majumdar, 2008). Estimation of the CRN model with three or more traits can, therefore, be best achieved through use of a mathematically equivalent but more computationally efficient reparameterization of the $\boldsymbol{G}_{(X)}$ and $\boldsymbol{P}_{(X)}$ matrices than is described by the standard parameterization presented in **Eq. 2-4**.

Firstly, the *p* x *p* correlation matrix $\boldsymbol{R_a}$ containing all genetic (or phenotypic) correlations for *p* phenotypes can be decomposed using a Cholesky factorization such that

$$\boldsymbol{R_a} = \boldsymbol{L_R}\boldsymbol{L_R}^{\top} \tag{5}$$

where $\boldsymbol{L_R}$ is a lower-triangular matrix with unit length rows and positive diagonal elements. These assumptions reduce the number of free parameters necessary for calculating $\boldsymbol{R_a}$, as the diagonal elements of $\boldsymbol{L_R}$ are determined by the off-diagonal elements of each row. Therefore, estimating $\boldsymbol{L_R}$ and subsequently deriving $\boldsymbol{R_a}$ using **Eq. 5** improves computational time of the model (Stan Development Team, 2023). Following previous work on the prediction of covariance matrices (Lewandowski et al., 2009; Bloome & Schrage, 2021), computational efficiency can then be further increased by decomposing $\boldsymbol{L_R}$ into a vector $\boldsymbol{\omega}$ of length $\frac{p(p-1)}{2}$ containing the canonical partial correlations constitutive of all unique lower-triangular elements in this matrix. The canonical partial correlations in $\boldsymbol{\omega}$ are of the same sign as their corresponding elements in $\boldsymbol{L_R}$, but their magnitudes represent residual correlations between corresponding row and column variables after regressing both on all prior occurring row variables. In the general case, the canonical partial correlation ω_u, where $u = \frac{2cp-c^2+2r-3c-2}{2}$ is the vector element corresponding to unique lower-triangular Cholesky factor $L_{R[r,c]}$ at row *r* and column *c*, is given by

$$\omega_u = \begin{cases} L_{R[r,c]}, & \text{if } c = 1 < r \\ L_{R[r,c]} \,/\, \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \tag{6.1}$$

such that the Cholesky factor can in turn be derived from ω_u by

$$L_{R[r,c]} = \begin{cases} \omega_u, & \text{if } c = 1 < r \\ \omega_u * \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \tag{6.2}$$

This general decomposition strategy can be adapted for the CRN model by extending each element in the vector $\boldsymbol{\omega}$ to its own vector of context-specific canonical partial correlations. Using the same approach developed above (**Eq. 2-4**), continuous environmental effects can then be specified and estimated more efficiently as predictors of the transformed canonical partial correlations

$$\begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})_1}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})\frac{p(p-1)}{2}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\omega}_1} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\omega}_{\frac{p(p-1)}{2}}} \end{bmatrix} \quad (7)$$

Applying the inverse link function $\tanh()$ and using Eq. 6.2 to calculate Cholesky factorized matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$, the original context-specific correlation matrices can then be derived $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ and subsequently applied to generate model predictions for estimating environmental effects on a more familiar scale. It is important to reiterate that the proposed implementation in Stan (Eq. 5-7) ensures the positive definiteness of the resulting correlation matrices $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ predicted by the CRN. Given that environmental effects are specified separately for trait correlations and variances in the CRN model (Eq. 1.3), the context-specific (co)variance matrices $\boldsymbol{G}_{(\boldsymbol{X})}$ derived from context-specific correlation matrices $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ will necessarily be positive definite.

Covarying environmental predictors can reduce the efficiency and accuracy of CRN parameter estimation. To reduce the effects of collinearity, the CRN fixed effects $\boldsymbol{\beta}_{\boldsymbol{\sigma}^2}$ and $\boldsymbol{\beta}_{\boldsymbol{\omega}}$ (or $\boldsymbol{\beta}_{\boldsymbol{r}}$) can also be more efficiently estimated using a so-called thin QR factorization of the **X** matrix (Harville, 1997). This involves decomposing the predictor matrix $\boldsymbol{X} = \boldsymbol{Q}^*\boldsymbol{R}^*$ into an orthogonal matrix $Q^* = Q\sqrt{n-1}$ and upper-triangle matrix $R^* = \frac{R}{\sqrt{n-1}}$, estimating trait-specific regression coefficients using the orthogonal vectors $\boldsymbol{Q}^*\boldsymbol{\beta}^*$, and then returning regression coefficients appropriately scaled to the original data scale of $\boldsymbol{X}$ using $\boldsymbol{\beta} = \boldsymbol{R}^{*-1}\boldsymbol{\beta}^*$. The QR decomposition increases efficiency by reducing posterior correlations during model sampling that would otherwise result from covariation among predictors.

Finally, the Cholesky matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$ can also be used to more efficiently predict individuals' context-specific additive genetic values from the CRN model. Following previous work by (Martin & Jaeggi, 2022), this can be accomplished using a matrix normal sampling distribution (Dutilleul, 1999), which extends the vectorized multivariate normal distribution to the sampling of multivariate normally distributed matrices. Using a n x p matrix $\boldsymbol{Z}_{\boldsymbol{G}}$ of standardized individual-level additive genetic deviations (i.e. z-scores of breeding values), a lower-triangular Cholesky decomposition $\boldsymbol{L}_{\boldsymbol{A}}$ of the relatedness matrix, and a diagonal matrix $\boldsymbol{S}_{\boldsymbol{a}(\boldsymbol{X}_n)} = \operatorname{diag}\left(\left[\sigma_{a(X_n)_1}, \ldots, \sigma_{a(X_n)_p}\right]\right)$ of context-specific genetic standard deviations, an *n* x *p* matrix of context-specific genetic values for each phenotype can be predicted by

$$\left[\boldsymbol{a}_{(X_n)_1}, \ldots, \boldsymbol{a}_{(X_n)_p}\right] = \mathbf{L}_{\mathbf{A}}\boldsymbol{Z}_{\boldsymbol{G}}\left(\mathbf{S}_{\mathbf{a}(\mathbf{X}_\mathbf{n})}\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X}_n)}\right)^{\top} \sim \text{Matrix Normal}\left(\mathbf{0}_{\text{nxp}}, \boldsymbol{A}, \boldsymbol{G}_{(\boldsymbol{X}_n)}\right) \quad (8)$$

$$\rightarrow \operatorname{vec}\left(\left[\boldsymbol{a}_{(X_n)_1}, \ldots, \boldsymbol{a}_{(X_n)_p}\right]\right) \sim \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{G}_{(\boldsymbol{X}_n)} \otimes \boldsymbol{A}\right)$$

Easy-to-use R functions are provided (see **data availability**) to straightforwardly facilitate computational gains from Eq. 5-8 while also generating more intuitive model estimates and predictions with respect to the standard parameterization of the CRN model (Eq. 2-4).

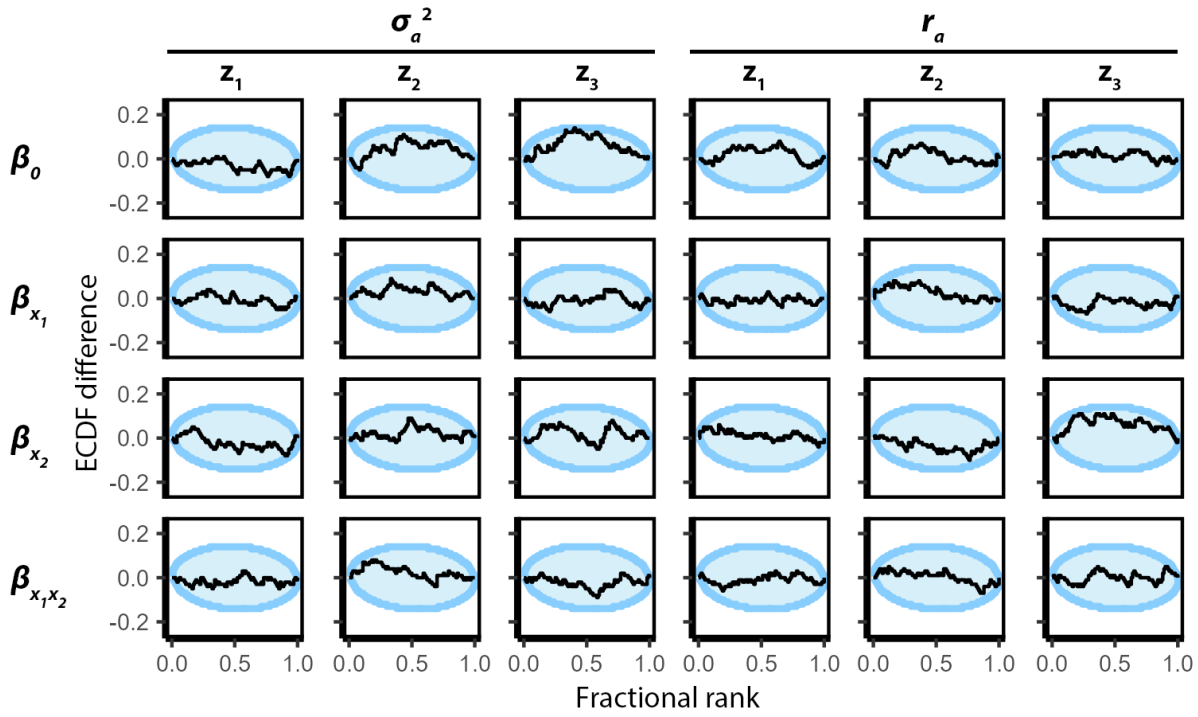
Model validation

To provide a general validation of the proposed model, I conducted a simulation-based calibration (SBC) procedure to assess whether the quantitative genetic CRN (Eq. 2) is an unbiased Bayesian estimator. Note that the phenotypic CRN (Eq. 4) is simply a variant of the quantitative genetic model with independent random effects and thus does not require additional validation. SBC is a procedure for assessing the performance of a Bayesian algorithm across a broad range of possible parameter values generated from the prior distributions of a generative model (see Talts et al., 2018 for further

details). This approach removes the need for arbitrarily picking a limited range of effect sizes for assessing performance and reduces the risk of missing unexpected sources of bias in uninvestigated regions of parameter space. Visual inspection of the correspondence between the generative prior distributions (expected values) and subsequent posterior distributions (inferred values) estimated during SBC is used to detect sources of bias, such as overdispersion in the estimator or inconsistent performance for extreme values.

100 datasets were simulated for SBC under very minimal sampling conditions of 200 individuals with a single measurement of 3 traits. Measurements were taken across environments characterized the interaction between 10 measured values of two continuous covariates (e.g. monthly temperatures, ages, plot densities). Parameter values were generated using standard weakly regularizing priors (Lemoine, 2019; McElreath, 2020), such that $\boldsymbol{\beta} \sim N(0,1)$ for RN fixed effects determining phenotypic means and genetic (co)variances, and $\boldsymbol{R_\epsilon} \sim \mathrm{LKJ}(10)$ for residual correlation matrices with fixed $\sigma_\epsilon = 1$ residual standard deviations. Relatedness matrices were simply positive-definite correlation matrices simulated from $\boldsymbol{A} \sim \mathrm{LKJ}(1)$. Posteriors for each dataset were estimated using 2000 MCMC samples across 4 chains with 500 samples each for warmup. Results from the SBC analysis showed that the distributions of inferred parameter values were congruent with the distributions of expected parameter values across the CRN fixed effects predicting genetic (co)variances (**Fig. 2**), with a 0.95+ probability that posterior inferences were not systematically upwardly or downwardly biased from the true values used to generate the data. This provides strong evidence that the proposed Bayesian estimator provides unbiased inference of CRNs even under conditions of very minimal sampling effort and a reasonably broad range of effect sizes. It is important to emphasize that these results concern bias per se in estimates of expected values and do not quantify the statistical uncertainty or power of hypothesis tests for detecting these effects. Achieving high levels of power and low levels of uncertainty will generally require much larger sample sizes, as it is the case for any quantitative genetic analysis. Simulations functions are provided (see **data availability**) to aid researchers in carrying out a priori power analyses for effect and sample size ranges of interest.

Figure 3. Simulation-based calibration of the CRN model.



Footnote. Results are shown for SBC analysis of 100 simulated datasets of 3 traits under minimal sampling conditions (N = 200 / 10 environmental contexts) generated from prior distributions defined over the parameters of the quantitative genetic CRN model (**Eq. 2**). The CRN contained four parameters for each genetic variance (σ_α^2) and correlation (r_α): β_0 for the trait-specific intercepts, β_{x_1} and β_{x_2} for the main effects of two continuous and independently distributed environments, and $\beta_{x_1x_2}$ for the interaction effect of these continuous environments. Plots show the difference (y-axis) between the empirical cumulative density functions (ECDFs) for CRN parameters from the generative prior distributions used to simulate datasets and the ECDFs of the estimated posterior distributions across datasets. This difference is shown by the black line and plotted as a function of the relative fractional rank (x-axis) of the simulated values in comparison to inferred values. Blue ellipses show regions providing 0.95+ probability of uniformity between the ECDFs of the simulated and estimated parameter distributions, providing support for a well-calibrated model without systematic bias (Talts et al., 2018). Therefore, while stochastic fluctuations are expected at computationally efficient sample sizes, black lines should remain within the blue ellipses across fractional ranks if the model generates unbiased posterior estimates of parameter values, with respect to the prior simulated values. Consistent deviations of the black line beyond the blue ellipse provide statistical evidence of bias in the region of parameter space indicated by the fractional ranks. For instance, if a model systematically underestimates parameter values, we expect the black lines to peak outside the blue ellipses at high fractional ranks, indicating that prior values were systematically larger than inferred estimates.

Worked example: social niche specialization in meerkats

To demonstrate the utility of the proposed framework, I applied a CRN model to analyze an openly available dataset from a long-term study (Houslay et al., 2021) on the heritability of three cooperative behaviors (babysitting, pup feeding and foraging, and vigilant guarding/sentinel activity) in wild meerkats (**Fig. 4a**). The goal of the analysis was to estimate the interactive effects of age, sex, and dominance status on the genetic (co)variance of these cooperative behaviors, as well as to investigate whether group size has a negative effect on genetic correlations. Prior work suggests that cooperative task generalization decreases while specialization subsequently increases in larger social groups, due to synergistic fitness benefits among individuals who benefit from investing more time performing distinct and complementary behaviors in larger groups (e.g. Bonner, 2004; Jeanson et al., 2007; Ulrich et al., 2018; Martin et al., 2023). If so, we would expect to observe positive genetic correlations among cooperative behaviors in small groups, but negative genetic correlations in large groups (**Fig. 1d**). Accordingly, fluctuations in group size within organisms' lifetimes may select for social plasticity in cooperative behavior to track these shifting fitness optima across social groups (de Jong, 1995; Martin et al., 2023), leading to the evolution of a group size dependent CRN and GxE in the expression of different tasks. Meerkats engage in extensive cooperative breeding, defense, and foraging in groups of variable size and composition (Clutton-Brock et al., 2001), providing a valuable system to further investigate these predictions.

Using only data of individuals with measures available for all three behaviors in the study of Houslay et al. (2021), the total sample size for the analysis was 1560 pedigreed individuals with 6751 (babysitting), 6461 (pup feeding), and 11532 (guarding/sentinel activity) total observations. I simplified certain components of the animal models employed by these authors to focus attention on the CRN, using only the covariates (age, sex, dominance status, group size) that were available for all traits and were identified as important for understanding mean phenotypic differences in the meerkats' behavior. Additional random effects were included for each trait to capture individual-level permanent environmental effects, group identity during observation, breeding season, and individual-by-season interactions. The three phenotypes were modeled using binomial (half-days observed babysitting/total days) and Poisson (count of pup feeding and minutes in sentinel activity) distributions. Following **Eq. 2** and using the computational strategy explained in **Eq. 5-8**, the same environmental covariates used to predict phenotypic means were also used to predict potential changes in quantitative genetic (co)variances among cooperative behaviors. Consider that from the perspective of a gene, organismal attributes such as sex, age, and dominance (serving as proxies for various attendant changes in hormonal activity, social experiences, etc.) are just as much aspects of 'the environment' potentially modulating its expression as more exogenous factors like group size (Service & Rose, 1985; Via & Lande, 1985; Pigliucci, 1996; Elgart et al., 2022; Martin et al., 2023). These covariates also allowed for appropriately testing the independent (age, sex, and dominance adjusted) effect of group size on genetic correlations among cooperative behaviors. A coding tutorial accompanying this worked example is provided on Github (see **data availability**).

Results

The CRN analysis uncovered continuous changes in the genetic variances and correlations of meerkats' cooperative behaviors in response to the interactive effects of age, dominance status, and sex, as well as the nonlinear effects of group size, providing clear evidence for GxE shaping the **G** matrix across environments. These effects are visualized as CRNs in **Fig. 4b-c** and summarized quantitatively in **Table 1**. Firstly, considering genetic variances, increasing age was strongly associated with greater genetic variance in babysitting behavior (BS), while age had weaker and more uncertain effects on the genetic variance of foraging and pup feeding (FD) and vigilant guarding behavior (GD). This indicates

that heritable individual differences in BS are expected to increase across the lifespan, independently of sex and dominance status. Sex did not have a main effect on the genetic variance of any traits, while dominance status had moderate to strong positive effects on the genetic variance of FD and GD. Changes in dominance status were, therefore, a primary driver of changes in the magnitude of heritable individual differences in cooperative behaviors (personality). Dominant individuals showed greater genetic variation than subordinates in their magnitude of FD and GD. Multivariate interactions also occurred between age, sex, and dominance. Genetic variance in BS reduced in response to the interaction of age and sex with dominance, while genetic variance in GD increased as a function of the interaction between age and dominance as well as the three-way interaction among age, sex, and dominance.

Environmental variation was also associated with changes in the genetic correlations among cooperative behaviors (**Table 1**). Among subordinates, males exhibited relatively stronger genetic correlations for BS ~ GD than females, which increased with age (**Fig. 4b**). Some evidence was found for reversed sex effects among dominant individuals, but dominance effects exhibited moderate to high statistical uncertainty overall. A clear main effect of age was observed for FD ~ BS, indicating that this genetic correlation tended to decrease across the lifespan, with older individuals being more likely to specialize in FD or BS than younger individuals. Negative age effects were also estimated for FD~BS and BS~GD but with greater statistical uncertainty. Group size decreased both FD~BS and FD~GD, independently of age, sex, and dominance effects, with more uncertainty in the positive effect of group size on BS~GD. Evidence was also found for a positive quadratic effect of group size on FD ~ GD, such that the negative effect was diminished for larger group sizes.

Combined effects of the multivariate environment on genetic variances and correlations generate nonlinear CRNs that are visualized in **Fig. 4b-c**. Subordinate males typically show more positive genetic (co)variances across ages than subordinate females, indicating more generalized genetic effects on and heritable individual differences in cooperative behavior. Subordinate females are in turn expected to show more negative genetic covariances among behaviors as they age (**Fig. 4b**). However, these patterns were complicated among dominant breeders. The direct effects of dominance status on genetic correlations were highly uncertain (**Table 1**) and should be interpreted cautiously, as is reflected by the much larger credible intervals for the predicted age CRNs of dominant individuals (bottom row plots in **Fig. 4b**). Independently of these effects, negative genetic covariance is expected between FD and BS in larger social groups, while a positive genetic covariance is expected between BS and GD in larger social groups (**Fig. 4c**). The genetic covariance between FD and GD is positive in small groups but declines nonlinearly and remains near to zero in average and larger than average group sizes. These results provide support for the prediction that fluctuations in group size select for plasticity in the expression of generalized versus specialized cooperative behavior across social groups. Consistent with prior research (Clutton-Brock et al., 2003), social niche specialization is not observed on average across social groups. However, the CRN model reveals that this is because small group sizes promote more positively integrated ($\sigma_a > 0$) genetic effects across cooperative behaviors, while larger group sizes promote negative genetic correlations ($\sigma_a < 0$) indicative of specialized performance of FD versus BS and GD tasks.

603 **Table 1**. Summary of CRN parameter posterior distributions.

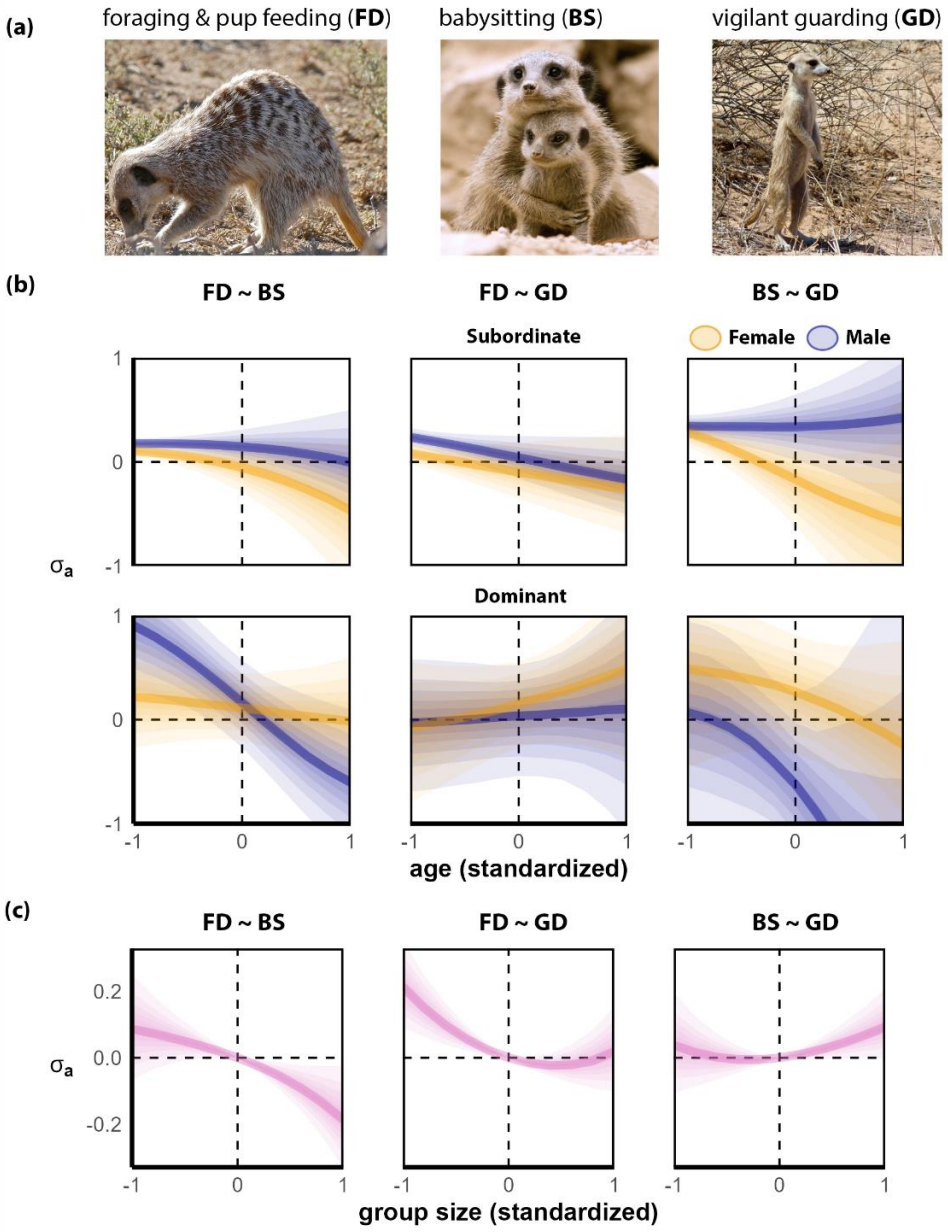
Regression coefficient	variance reaction norm $\boldsymbol{\beta}_{\sigma_\alpha^2}$		correlation reaction norm $\boldsymbol{\beta}_{r_\alpha}$	
	median	$p_{+/-}$	median	$p_{+/-}$
foraging and feeding pups (FD)			FD ~ BS	
age	0.19	0.81	-0.34	0.98
sex	0.10	0.62	0.31	0.90
dominance status	1.10	1.00	0.17	0.77
age * sex	-0.07	0.61	0.11	0.70
age * dominance	-0.10	0.64	0.25	0.79
sex * dominance	-0.36	0.84	-0.28	0.80
age * sex * dominance	-0.17	0.65	-0.65	0.93
group size	0.20	1.00	-0.12	0.98
group size^2	0.21	1.00	-0.04	0.73
babysitting (BS)			FD ~ GD	
age	0.96	1.00	-0.21	0.90
sex	-0.21	0.75	0.15	0.77
dominance status	-0.02	0.52	0.19	0.80
age * sex	-0.13	0.66	-0.01	0.52
age * dominance	-0.85	0.99	0.34	0.92
sex * dominance	0.76	0.96	-0.20	0.77
age * sex * dominance	-0.01	0.56	-0.10	0.60
group size	-0.12	0.97	-0.10	0.98
group size^2	0.08	0.87	0.11	0.99
vigilant guarding (GD)			BS ~ GD	
age	-0.19	0.94	-0.16	0.77
sex	0.12	0.77	0.32	0.96
dominance status	0.49	0.99	0.23	0.85
age * sex	-0.12	0.81	0.30	0.96
age * dominance	0.47	0.99	0.15	0.71
sex * dominance	-0.01	0.52	-0.37	0.84
age * sex * dominance	0.65	0.98	-0.13	0.60
group size	0.02	0.72	0.07	0.88
group size^2	0.05	0.94	0.05	0.79

604

605 **Footnote**. Posterior distributions of CRN parameters (regression coefficients) for the genetic variances ($\boldsymbol{\beta}_{\sigma_\alpha^2}$) and
606 genetic correlations ($\boldsymbol{\beta}_{r_\alpha}$) among three meerkat social behaviors: foraging and pup feeding (FD), babysitting (BS), and
607 vigilant guarding (GD). Posteriors are summarized by their median and the probability of a directional effect ($p_{+/-}$).
608 Note that $p_{+/-}$closer to 1 provide stronger support for a positive or negative effect, contingent on the sign of the median
609 effect size. Reference categories for sex and dominance are female and subordinate.

610

Figure 4. Multivariate CRN of cooperative behavior among meerkats.



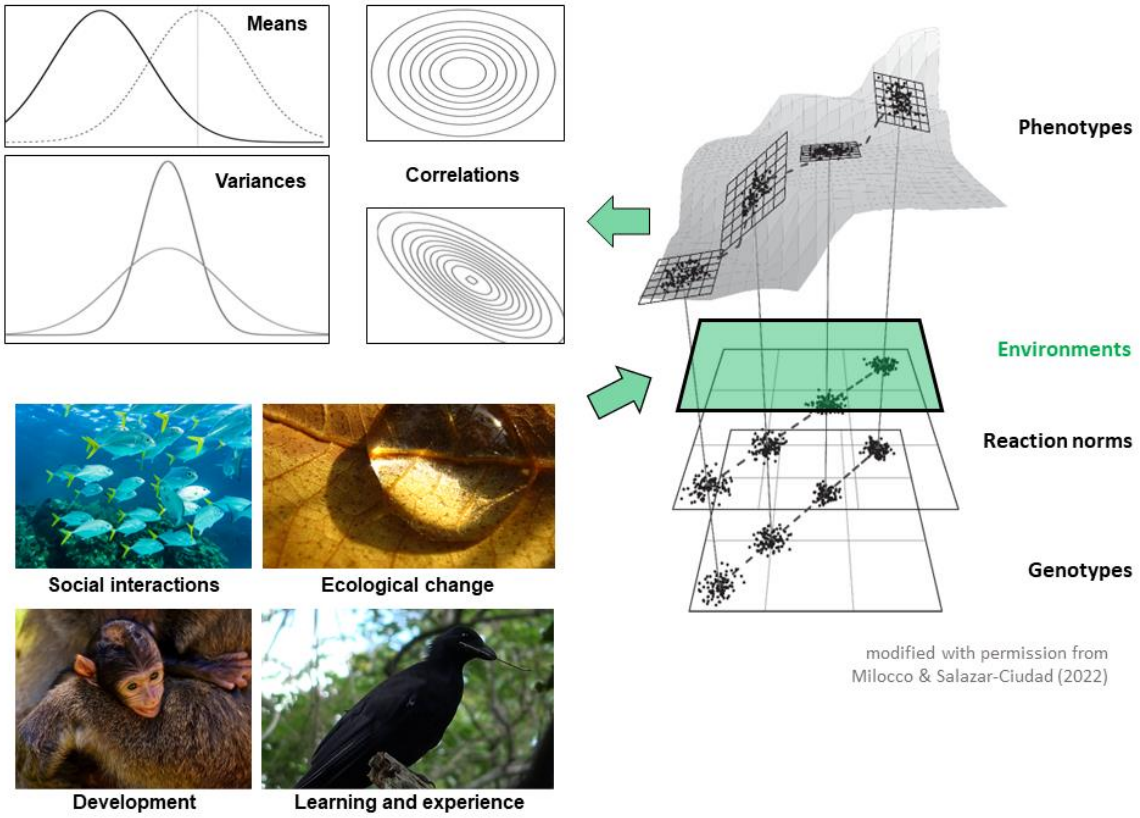
Footnote. Posterior estimates are shown for multivariate and nonlinear environmental effects on the genetic covariances σ_a among (**a**) meerkats' foraging and pup feeding (FD), babysitting (BS), and vigilant guarding (GD) behavior. Creative commons picture credit: Bernard DUPONT and Jon Pinder (Flickr). (**b**) Posterior CRNs for the interactive effects of sex (orange = female, blue = male), dominance status (top row = subordinate, bottom = dominant), and age (units of months, SD standardized) on σ_a^2 (left row = FD~BS, center = FD~GD, right = BS~GD). Shaded bands indicate 10–90% posterior CI from the darkest to lightest bands, respectively, while the dark lines indicate posterior median values. CRN slopes greater or less than zero provide evidence for GxE interactions. (**x**) CRNs for the effect of group size (units of 5, SD standardized) on σ_a, adjusted for the interactive effects of sex, age, and dominance status. Dotted vertical lines indicate the expected covariance at the average group size (0), while dotted horizontal lines indicate $\sigma_a = 0$, so that values above this line provide evidence for task generalization ($\sigma_a > 0$) and values below provide evidence for task specialization ($\sigma_a < 0$).

Conclusion

A longstanding goal unifying diverse fields of ecological and evolutionary science is to understand the role of phenotypic plasticity in the adaptation of complex traits (Via et al., 1995; Paenke et al., 2007; Hutchings, 2011; Kuzawa & Bragg, 2012; Hendry, 2016; Pfennig, 2021). While strong theoretical emphasis has been placed on understanding the role of genetic (co)variances in constraining multivariate evolution (Phillips & Arnold, 1989; Walsh & Blows, 2009; Chebib & Guillaume, 2017), it is often underappreciated that genetic and phenotypic (co)variances are themselves the product of underlying genotype- and phenotype-by-environment interactions (Service & Rose, 1985; de Jong, 1989; Pigliucci, 1996; Elgart et al., 2022; Martin et al., 2023). Modeling these dynamic environmental interactions is, therefore, a crucial but easily overlooked step in effectively explaining ongoing adaptation in a rapidly changing world (Westneat et al., 2019; Hudak & Dybdahl, 2023). Analytic tools for efficiently inferring these complex patterns have been limited, however, particularly outside of the laboratory or agricultural contexts, where organisms are exposed to continuous and multivariate patterns of spatial and temporal variation in their local microhabitats. When such environmental variation is relevant for fitness and the benefits of responding to it outweigh the costs of producing a response, adaptive plasticity is expected to evolve in trait expression (Gavrilets & Scheiner, 1993; de Jong, 1995; Haaland et al., 2021). In many cases, this plasticity will be reflected in average trait values; however, when fitness-relevant variation also occurs with respect to trait (co)variances within individuals' lifetimes (e.g. through fluctuating correlational selection, Revell, 2007; Roff & Fairbairn, 2012), adaptive plasticity can evolve in trait variances and correlations (**Fig. 1, 5**).

Important empirical efforts have been made to investigate the fluctuations in **G** and **P** matrices that result from such plasticity, as well as potentially rapid microevolution, in response to environmental heterogeneity and ongoing change in natural populations (Björklund et al., 2013; Bolund et al., 2015; Wood & Brodie, 2015). However, current character state approaches for analyzing changes in trait (co)variances rely on discretizing the environment, as well as often unrealistic sample size requirements, resulting in undesirable inferential risks (**Fig. 2a**). Random regression approaches suffer from similar considerations (**Fig. 2b**), particularly in the presence of complex, interactive environmental effects and/or systems where repeated individual measurements or experimental breeding designs across environments are not feasible. Ultimately, these constraints limit empiricists' ability to robustly infer continuous, multivariate, and potentially nonlinear environmental processes underlying GxE and PxE interactions in the wild (**Fig. 1**). The CRN model proposed here provides a validated solution (**Fig. 3**) to this challenge, extending the standard animal model (Kruuk, 2004) to increase its flexibility for describing multivariate environmental effects on all aspects of quantitative genetic expression. As demonstrated by the worked example in meerkats, building on prior research by Houslay et al. (2021), CRNs can harness the rich information in long-term field datasets to generate fresh insights into longstanding empirical questions, such as the effects of group size on social niche specialization in animal societies (**Fig. 4c**). The CRN also uncovered multivariate GxE interactions among sex, age, and dominance status (**Fig. 4b**), which would require many more parameters and larger sample sizes to effectively estimate using alternative methods (**Fig. 2**). Further application of the CRN model (**Eq. 2-4**) is, therefore, likely to enhance our understanding of the evolution and ecology of multivariate plasticity across a variety of complex phenotypes in the wild.

671 **Figure 5.** Environmental effects on the expression of multivariate phenotypes.



672

673 **Footnote**. A conceptual figure of GxE and PxE on multivariate traits, modified with permission from
674 Milocco and Salazar-Ciudad (2022). The phenotype-to-genotype map, shown here by lines connecting
675 populations of genotypes (lowest surface) to distributions of phenotypes (highest), is mediated
676 through individuals' RNs and the distribution of environments within and across generations. RNs not
677 only structure the expression of trait means, but also the variances, correlations, and (co)variances
678 among traits (i.e. CRNs). Therefore, **G** and **P** matrices describing the mapping between genetic and
679 phenotypic variation are often highly sensitive to the environmental contexts in which individuals are
680 measured (GxE and PxE interactions, indicated by green arrows). CRNs may evolve in response to
681 diverse environmental contexts such as the quality and consistency of early parental care,
682 opportunities for and costs of learning, variability and harshness of the climate, habitat degradation,
683 magnitude and predictability of local resources, the density of predators and parasites, the strengths
684 of intra and intersexual competition, social network position and mating system, food web structure,
685 etc. When such environments change (dotted lines) and developmental and/or contextual plasticity
686 has evolved in a population, trait (co)variances may rapidly respond to spatiotemporal heterogeneity
687 within and across generations (top layer planes). Mechanistically and ecologically informed CRN
688 models can be used to better predict how GxE will shape the expression and evolution of multivariate
689 traits in response to ongoing socio-eco-evolutionary dynamics. Creative commons picture credit:
690 NickJack and Alexas_Fotos (Pixabay) and Luz Adriana Villa and Corvus moneduloides (Flickr).

Acknowledgements

I owe special thanks to Thomas Houslay for openly providing and helping me to organize the meerkat data analyzed in this manuscript, as well as to the many other researchers who played a role in producing this dataset. I also extend my thanks to Michael Morrissey and the research groups of Erik Postma, Niels Dingemanse, and Adrian Jaeggi for their helpful feedback on previous versions of this manuscript.

Data availability

Guided tutorials for implementing CRNs, as well as R code for replicating the worked empirical example, are publicly available on Github at https://github.com/Jordan-Scott-Martin/covariance-reaction-norms

References

Aguirre, J. D., Hine, E., McGuigan, K., & Blows, M. W. (2014). Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity*, *112*(1), 21–29.

Barneche, D. R., Robertson, D. R., White, C. R., & Marshall, D. J. (2018). Fish reproductive-energy output increases disproportionately with body size. *Science*, *360*(6389), 642–645.

Björklund, M., Husby, A., & Gustafsson, L. (2013). Rapid and unpredictable changes of the G-matrix in a natural bird population over 25 years. *Journal of Evolutionary Biology*, *26*(1), 1–13.

Björklund, Mats. (1996). The importance of evolutionary constraints in ecological time scales. *Evolutionary Ecology*, *10*(4), 423–431.

Björklund, Mats. (2004). Constancy of the G matrix in ecological time. *Evolution; International Journal of Organic Evolution*, *58*(6), 1157–1164.

Björklund, Mats, & Gustafsson, L. (2015). The stability of the G-matrix: The role of spatial heterogeneity. *Evolution; International Journal of Organic Evolution*, *69*(7), 1953–1958.

Bliard, L., Martin, J. S., Paniw, M., Blumstein, D. T., Martin, J. G. A., Pemberton, J. M., Nussey, D. H., Childs, D. Z., & Ozgul, A. (2024). Detecting context dependence in the expression of life history trade-offs. *The Journal of Animal Ecology*. https://doi.org/10.1111/1365-2656.14173

Bloome, D., & Schrage, D. (2021). Covariance Regression Models for Studying Treatment Effect Heterogeneity Across One or More Outcomes: Understanding How Treatments Shape Inequality. *Sociological Methods & Research*, *50*(3), 1034–1072.

Bolund, E., Hayward, A., Pettay, J. E., & Lummaa, V. (2015). Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution; International Journal of Organic Evolution*, *69*(3), 747–755.

Bonner, J. T. (2004). Perspective: the size-complexity rule. *Evolution; International Journal of Organic Evolution*, *58*(9), 1883–1890.

725 Brommer, J., Class, B., & Covarrubias-Pazaran, G. (2019). Multivariate Mixed Models in Ecology and
726 Evolutionary biology: Inferences and implementation in R. In *EcoEvoRxiv*.
727 https://doi.org/10.32942/osf.io/hs38a
728 Brooks, R., Hunt, J., Blows, M. W., Smith, M. J., Bussière, L. F., & Jennions, M. D. (2005). Experimental
729 evidence for multivariate stabilizing sexual selection. *Evolution; International Journal of*
730 *Organic Evolution*, *59*(4), 871–880.
731 Bucklaew, A., & Dochtermann, N. A. (2021). The effects of exposure to predators on personality and
732 plasticity. *Ethology: Formerly Zeitschrift Fur Tierpsychologie*, *127*(2), 158–165.
733 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo,
734 J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical*
735 *Software*, *76*. https://doi.org/10.18637/jss.v076.i01
736 Carvalho, M. J. A., & Mirth, C. K. (2015). Coordinating morphology with behavior during development:
737 an integrative approach from a fly perspective. *Frontiers in Ecology and Evolution*, *3*.
738 https://doi.org/10.3389/fevo.2015.00005
739 Chang, C.-C., Moiron, M., Sánchez-Tójar, A., Niemela, P., & Laskowski, K. (2023). What's the meta-
740 analytic evidence for life-history trade-offs at the genetic level? In *Authorea*.
741 https://doi.org/10.22541/au.168795190.00787983/v1
742 Chebib, J., & Guillaume, F. (2017). What affects the predictability of evolutionary constraints using a
743 G-matrix? The relative effects of modular pleiotropy and mutational correlation. *Evolution;*
744 *International Journal of Organic Evolution*, *71*(10), 2298–2312.
745 Clutton-Brock, T. H., Brotherton, P. N., Russell, A. F., O'Riain, M. J., Gaynor, D., Kansky, R., Griffin, A.,
746 Manser, M., Sharpe, L., McIlrath, G. M., Small, T., Moss, A., & Monfort, S. (2001). Cooperation,
747 control, and concession in meerkat groups. *Science (New York, N.Y.)*, *291*(5503), 478–481.
748 Clutton-Brock, T. H., Russell, A. F., & Sharpe, L. L. (2003). Meerkat helpers do not specialize in particular
749 activities. *Animal Behaviour*, *66*(3), 531–540.
750 Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

751 Damián, X., Ochoa-López, S., Gaxiola, A., Fornoni, J., Domínguez, C. A., & Boege, K. (2020). Natural
752 selection acting on integrated phenotypes: covariance among functional leaf traits increases
753 plant fitness. *The New Phytologist*, *225*(1), 546–557.
754 Danchin, É., & Wagner, R. H. (2010). Inclusive heritability: combining genetic and non-genetic
755 information to study animal behavior and culture. *Oikos* , *119*(2), 210–218.
756 de Jong, G. (1989). Phenotypically plastic characters in isolated populations. In A. Fontdevila (Ed.),
757 *Evolutionary Biology of Transient Unstable Populations*. Springer.
758 de Jong, G. (1995). Phenotypic plasticity as a product of selection in a variable environment. *The*
759 *American Naturalist*, *145*(4), 493–512.
760 de la Mata, R., Zas, R., Bustingorri, G., Sampedro, L., Rust, M., Hernandez-Serrano, A., & Sala, A. (2022).
761 Drivers of population differentiation in phenotypic plasticity in a temperate conifer: A 27-year
762 study. *Evolutionary Applications*, *15*(11), 1945–1962.
763 Dean, D. S., & Majumdar, S. N. (2008). Extreme value statistics of eigenvalues of Gaussian random
764 matrices. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *77*(4 Pt 1), 041108.
765 Dingemanse, Niels J., Barber, I., & Dochtermann, N. A. (2020). Non-consumptive effects of predation:
766 does perceived risk strengthen the genetic integration of behaviour and morphology in
767 stickleback? *Ecology Letters*, *23*(1), 107–118.
768 Dingemanse, Niels J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour:
769 mixed-effect modelling approaches. *The Journal of Animal Ecology*, *82*(1), 39–54.
770 Dingemanse, Niels Jeroen, Araya-Ajoy, Y. G., & Westneat, D. F. (2021). Most published selection
771 gradients are underestimated: Why this is and how to fix it. *Evolution; International Journal of*
772 *Organic Evolution*, *75*(4), 806–818.
773 Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *Journal of Statistical*
774 *Computation and Simulation*, *64*(2), 105–123.
775 Elgart, M., Goodman, M. O., Isasi, C., Chen, H., Morrison, A. C., de Vries, P. S., Xu, H., Manichaikul, A.
776 W., Guo, X., Franceschini, N., Psaty, B. M., Rich, S. S., Rotter, J. I., Lloyd-Jones, D. M., Fornage,

M., Correa, A., Heard-Costa, N. L., Vasan, R. S., Hernandez, R., … Sofer, T. (2022). Correlations between complex human phenotypes vary by genetic background, gender, and environment. *Cell Reports. Medicine*, *3*(12), 100844.

Enzmann, B. L., & Nonacs, P. (2021). Age-related division of labor occurs in ants at the earliest stages of colony initiation. *Behavioral Ecology and Sociobiology*, *75*(2), 35.

Estes, S., & Arnold, S. J. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, *169*(2), 227–244.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics (4th edition)*. Pearson, Prentice Hall.

Ferguson-Gow, H., Sumner, S., Bourke, A. F. G., & Jones, K. E. (2014). Colony size predicts division of labour in attine ants. *Proceedings. Biological Sciences / The Royal Society*, *281*(1793). https://doi.org/10.1098/rspb.2014.1411

Fogarty, L., & Wade, M. J. (2022). Niche construction in quantitative traits: heritability and response to selection. *Proceedings. Biological Sciences / The Royal Society*, *289*(1976), 20220401.

Gavrilets, S., & Scheiner, S. M. (1993). The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *Journal of Evolutionary Biology*, *6*(1), 31–48.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis Third edition*. CRC Press.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/2011.01808

Grebe, N. M., Fitzpatrick, C., Sharrock, K., Starling, A., & Drea, C. M. (2019). Organizational and activational androgens, lemur social play, and the ontogeny of female dominance. *Hormones and Behavior*, *115*, 104554.

Haaland, T. R., Wright, J., & Ratikainen, I. I. (2021). Individual reversible plasticity as a genotype-level bet-hedging strategy. *Journal of Evolutionary Biology*, *34*(7), 1022–1033.

803 Haave-Audet, E., Besson, A. A., Nakagawa, S., & Mathot, K. J. (2022). Differences in resource
804 acquisition, not allocation, mediate the relationship between behaviour and fitness: a
805 systematic review and meta-analysis. *Biological Reviews of the Cambridge Philosophical*
806 *Society*, *97*(2), 708–731.
807 Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag.
808 Heino, M., Díaz Pauli, B., & Dieckmann, U. (2015). Fisheries-Induced Evolution. *Annual Review of*
809 *Ecology, Evolution, and Systematics*, *46*(1), 461–480.
810 Henderson, C. R., Jr. (1982). Analysis of covariance in the mixed model: higher-level,
811 nonhomogeneous, and random regressions. *Biometrics*, *38*(3), 623–640.
812 Hendry, A. P. (2016). Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics.
813 *The Journal of Heredity*, *107*(1), 25–41.
814 Henry, G. A., & Stinchcombe, J. R. (2023). G-matrix stability in clinally diverging populations of an
815 annual weed. *Evolution; International Journal of Organic Evolution*, *77*(1), 49–62.
816 Hoffman, M. D., & Gelman, A. (2011). The no-U-Turn Sampler: Adaptively setting path lengths in
817 Hamiltonian Monte Carlo. In *arXiv [stat.CO]*. arXiv.
818 https://jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf
819 Houslay, T. M., Nielsen, J. F., & Clutton-Brock, T. H. (2021). Contributions of genetic and nongenetic
820 sources to variation in cooperative behavior in a cooperative mammal. *Evolution;*
821 *International Journal of Organic Evolution*, *75*(12), 3071–3086.
822 Huang, Z.-Y., & Robinson, G. E. (1996). Regulation of honey bee division of labor by colony age
823 demography. *Behavioral Ecology and Sociobiology*, *39*(3), 147–158.
824 Huber, S. E., Lenz, B., Kornhuber, J., & Müller, C. P. (2017). Prenatal androgen-receptor activity has
825 organizational morphological effects in mice. *PloS One*, *12*(11), e0188752.
826 Hudak, A., & Dybdahl, M. (2023). Phenotypic plasticity under the effects of multiple environmental
827 variables. *Evolution; International Journal of Organic Evolution*, *77*(6), 1370–1381.

828 Hutchings, J. A. (2011). Old wine in new bottles: reaction norms in salmonid fishes. *Heredity*, *106*(3),
829 421–437.
830 Jeanson, R., Fewell, J. H., Gorelick, R., & Bertram, S. M. (2007). Emergence of Increased Division of
831 Labor as a Function of Group Size. *Behavioral Ecology and Sociobiology*, *62*(2), 289–298.
832 Jusufovski, D., & Kuparinen, A. (2020). Exploring individual and population eco-evolutionary feedbacks
833 under the coupled effects of fishing and predation. *Fisheries Research*, *231*, 105713.
834 Karsai, I., & Wenzel, J. W. (1998). Productivity, individual-level and colony-level flexibility, and
835 organization of work as consequences of colony size. *Proceedings of the National Academy of*
836 *Sciences of the United States of America*, *95*(15), 8665–8669.
837 Koch, E. L., Sbilordo, S. H., & Guillaume, F. (2020). Genetic variance in fitness and its cross-sex
838 covariance predict adaptation during experimental evolution. *Evolution; International Journal*
839 *of Organic Evolution*, *74*(12), 2725–2740.
840 Koutsidi, M., Lazaris, A., Peristeraki, P., Tserpes, G., & Tzanatos, E. (2024). Quantification of
841 intraspecific and interspecific competition in fish species of the Aegean Sea. *ICES Journal of*
842 *Marine Science: Journal Du Conseil*, *81*(2), 334–347.
843 Kraft, P. G., Wilson, R. S., Franklin, C. E., & Blows, M. W. (2006). Substantial changes in the genetic
844 basis of tadpole morphology of Rana lessonae in the presence of predators. *Journal of*
845 *Evolutionary Biology*, *19*(6), 1813–1818.
846 Kruuk, L. E. B., & Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity
847 between relatives. *Journal of Evolutionary Biology*, *20*(5), 1890–1903.
848 Kruuk, Loeske E. B. (2004). Estimating genetic parameters in natural populations using the 'animal
849 model.' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*,
850 *359*(1446), 873–890.
851 Kuzawa, C. W., & Bragg, J. M. (2012). Plasticity in Human Life History Strategy: Implications for
852 Contemporary Human Variation and the Evolution of Genus Homo. *Current Anthropology*,
853 *53*(S6), S369–S382.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain body size allometry. *Evolution; International Journal of Organic Evolution*, *33*(1Part2), 402–416.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution; International Journal of Organic Evolution*, *37*(6), 1210–1226.

Lande, Russell. (1976). NATURAL SELECTION AND RANDOM GENETIC DRIFT IN PHENOTYPIC EVOLUTION. *Evolution; International Journal of Organic Evolution*, *30*(2), 314–334.

Lankau, R. A. (2011). Rapid Evolutionary Change and the Coexistence of Species. *Annual Review of Ecology, Evolution, and Systematics*, *42*(Volume 42, 2011), 335–354.

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *55*(2), 139–185.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* , *128*(7), 912–928.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Link, W. A., & Eaton, M. J. (2011). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*(1). https://doi.org/10.1111/j.2041-210X.2011.00131.x

Lofeu, L., Brandt, R., & Kohlsdorf, T. (2017). Phenotypic integration mediated by hormones: associations among digit ratios, body size and testosterone during tadpole development. *BMC Evolutionary Biology*, *17*(1), 175.

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.

Martin, J. (2021). Estimating nonlinear selection on behavioral reaction norms. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/u26tz

Martin, J. S., & Jaeggi, A. V. (2022). Social animal models for quantifying plasticity, assortment, and selection on interacting phenotypes. *Journal of Evolutionary Biology*, *35*(4), 520–538.

Martin, Jordan S., Jaeggi, A. V., & Koski, S. E. (2023). The social evolution of individual differences: Future directions for a comparative science of personality in social behavior. *Neuroscience and Biobehavioral Reviews*, *144*, 104980.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McElreath, R. (2020). *Statistical Rethinking; A Bayesian Course with Examples in R and Stan; Second Edition*. CRC Press.

McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Mahler, D. L., Kolbe, J. J., Losos, J. B., & Brodie, E. D., 3rd. (2018). Adaptive radiation along a deeply conserved genetic line of least resistance in Anolis lizards. *Evolution Letters*, *2*(4), 310–322.

Mitchell, D. J., & Houslay, T. M. (2021). Context-dependent trait covariances: how plasticity shapes behavioral syndromes. *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology*, *32*(1), 25–29.

Montoya, E. R., Terburg, D., Bos, P. A., Will, G.-J., Buskens, V., Raub, W., & van Honk, J. (2013). Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. *Psychoneuroendocrinology*, *38*(8), 1362–1369.

Nishio, M., & Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. *Genetics, Selection, Evolution: GSE*, *51*(1), 73.

Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, *20*(3), 831–844.

Oh, K. P., & Shaw, K. L. (2013). Multivariate sexual selection in a rapidly evolving speciation phenotype. *Proceedings. Biological Sciences / The Royal Society*, *280*(1761), 20130482.

Oomen, R. A., & Hutchings, J. A. (2022). Genomic reaction norms inform predictions of plastic and adaptive responses to climate change. *The Journal of Animal Ecology*, *91*(6), 1073–1087.

905 Paenke, I., Sendhoff, B., & Kawecki, T. J. (2007). Influence of plasticity and learning on evolution under
906 directional selection. *The American Naturalist*, *170*(2), E47-58.
907 Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models
908 in ecology: an introduction with mgcv. *PeerJ*, *7*, e6876.
909 Pfennig, D. W. (2021). *Phenotypic Plasticity & Evolution* (1st Edition). CRC Press.
910 Phillips, P. C., & Arnold, S. J. (1989). Visualizing Multivariate Selection. *Evolution; International Journal*
911 *of Organic Evolution*, *43*(6), 1209–1222.
912 Pigliucci, M. (1996). Modelling phenotypic plasticity. II. Do genetic correlations matter? *Heredity*, *77 (*
913 *Pt 5)*, 453–460.
914 Qi, P.-F., Jiang, Y.-F., Guo, Z.-R., Chen, Q., Ouellet, T., Zong, L.-J., Wei, Z.-Z., Wang, Y., Zhang, Y.-Z., Xu,
915 B.-J., Kong, L., Deng, M., Wang, J.-R., Chen, G.-Y., Jiang, Q.-T., Lan, X.-J., Li, W., Wei, Y.-M., &
916 Zheng, Y.-L. (2019). Transcriptional reference map of hormone responses in wheat spikes.
917 *BMC Genomics*, *20*(1), 390.
918 R Core Team. (2023). *R: A language and environment for statistical computing*.
919 Revell, L. J. (2007). The G matrix under fluctuating correlational mutation and selection. *Evolution;*
920 *International Journal of Organic Evolution*, *61*(8), 1857–1872.
921 Riutort-Mayol, G., Bürkner, P.-C., Andersen, M. R., Solin, A., & Vehtari, A. (2022). Practical Hilbert space
922 approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and*
923 *Computing*, *33*(1), 17.
924 Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution;*
925 *International Journal of Organic Evolution*, *50*(4), 1392–1403.
926 Roff, D. A., & Fairbairn, D. J. (2012). A test of the hypothesis that correlational selection generates
927 genetic correlations. *Evolution; International Journal of Organic Evolution*, *66*(9), 2953–2960.
928 Rolian, C. (2020). Ecomorphological specialization leads to loss of evolvability in primate limbs.
929 *Evolution; International Journal of Organic Evolution*, *74*(4), 702–715.

930 Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of
931 residual variance - estimation of variance components using double hierarchical generalized
932 linear models. *Genetics, Selection, Evolution: GSE*, *42*(1), 8.
933 Royauté, R., Hedrick, A., & Dochtermann, N. A. (2020). Behavioural syndromes shape evolutionary
934 trajectories via conserved genetic architecture. *Proceedings. Biological Sciences / The Royal*
935 *Society*, *287*(1927), 20200183.
936 Schaum, C.-E., Buckling, A., Smirnoff, N., & Yvon-Durocher, G. (2022). Evolution of thermal tolerance
937 and phenotypic plasticity under rapid and slow temperature fluctuations. *Proceedings.*
938 *Biological Sciences / The Royal Society*, *289*(1980), 20220834.
939 Schluter, D. (1996). Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution;*
940 *International Journal of Organic Evolution*, *50*(5), 1766–1774.
941 Service, Philip M., & Rose, M. R. (1985). Genetic Covariation Among Life-History Components: The
942 Effect of Novel Environments. *Evolution; International Journal of Organic Evolution*, *39*(4),
943 943–945.
944 Sharpe, D. M. T., & Hendry, A. P. (2009). Life history change in commercially exploited fish stocks: an
945 analysis of trends across studies. *Evolutionary Applications*, *2*(3), 260–275.
946 Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development
947 predicts genotypic differences in plasticity. *Evolution; International Journal of Organic*
948 *Evolution*, *72*(10), 2167–2180.
949 Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual 2.33*.
950 Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268.
951 Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian Inference
952 Algorithms with Simulation-Based Calibration. In *arXiv [stat.ME]*. arXiv.
953 http://arxiv.org/abs/1804.06788
954 Thomas, M. L., & Elgar, M. A. (2003). Colony size affects division of labour in the ponerine ant
955 Rhytidoponera metallica. *Die Naturwissenschaften*, *90*(2), 88–92.

Thomson, C. E., Winney, I. S., Salles, O. C., & Pujol, B. (2018). A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PloS One*, *13*(10), e0197720.

Trumble, B. C., Jaeggi, A. V., & Gurven, M. (2015). Evolving the neuroendocrine physiology of human and primate cooperation and collective action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1683), 20150014.

Ulrich, Y., Saragosti, J., Tokita, C. K., Tarnita, C. E., & Kronauer, D. J. C. (2018). Fitness benefits and emergent division of labour at the onset of group living. *Nature*, *560*(7720), 635–638.

Uusi-Heikkilä, S. (2020). Implications of size-selective fisheries on sexual selection. *Evolutionary Applications*, *13*(6), 1487–1500.

van Noordwijk, A. J., & de Jong, G. (1986). Acquisition and Allocation of Resources: Their Influence on Variation in Life History Tactics. *The American Naturalist*, *128*(1), 137–142.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Van Tienderen, P. H. (1995). Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, *10*(5), 212–217.

Via, Sara, & Lande, R. (1985). Genotype-Environment Interactions and the Evolution of Phenotypic Plasticity. *Evolution; International Journal of Organic Evolution*, *39*(3), 505–522.

vom Saal, F. S. (1979). Prenatal exposure to androgen influences morphology and aggressive behavior of male and female mice. *Hormones and Behavior*, *12*(1), 1–11.

Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 41–59.

Westneat, D. F., Potts, L. J., Sasser, K. L., & Shaffer, J. D. (2019). Causes and Consequences of Phenotypic Plasticity in Complex Environments. *Trends in Ecology & Evolution*, *34*(6), 555–568.

980 Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., &
981 Nussey, D. H. (2010). An ecologist's guide to the animal model. *The Journal of Animal Ecology*,
982 *79*(1), 13–26.
983 Wittman, T. N., Robinson, C. D., McGlothlin, J. W., & Cox, R. M. (2021). Hormonal pleiotropy structures
984 genetic covariance. *Evolution Letters*, *5*(4), 397–407.
985 Wood, C. W., & Brodie, E. D., 3rd. (2015). Environmental effects on the structure of the G-matrix.
986 *Evolution; International Journal of Organic Evolution*, *69*(11), 2927–2940.
987 Yewers, M. S. C., Jessop, T. S., & Stuart-Fox, D. (2017). Endocrine differences among colour morphs in
988 a lizard with alternative behavioural strategies. *Hormones and Behavior*, *93*, 118–127.