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

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 other researchers who helped to produce 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.

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 extend prior multilevel regression models for quantitative genetic inference (so-called animal models) to develop a novel covariance reaction norm (CRN) model, which can be used to detect how trait (co)variances respond to 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 use simulations to validate its implementation for Bayesian inference in Stan, as well as to compare its performance to standard character state and random regression approaches. Findings demonstrate superior accuracy and power for detecting environmental effects on genetic covariance with modest sample sizes. I then apply the CRN 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, heterogeneity, context-dependent, 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 adaptive evolution when these traits are under positive selection (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; Niels Jeroen Dingemanse et al., 2021; J. S. Martin, Araya-Ajoy, et al., 2024), testing theoretical models of developmental integration and niche specialization (Damián et al., 2020; J. S. Martin et al., 2023; Rolian, 2020), 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 $\boldsymbol{G}$ matrix and phenotypic $\boldsymbol{P}$ matrix can be used to describe these multivariate (co)variances and predict their evolutionary consequences (Lande, 1979; Lande & Arnold, 1983). Various quantities derived from $\boldsymbol{G}$ and $\boldsymbol{P}$ 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). Extensive theoretical work has investigated the evolution of these matrices under varying genetic, demographic, and environmental conditions (Arnold et al., 2008). For instance, epistasis is expected to shape $\boldsymbol{G}$ by promoting the evolution of covarying mutational effects across traits (Jones et al., 2014). Migration can drive the evolution of genetic (co)variances by facilitating gene flow between diverging populations, reshaping $\boldsymbol{G}$ in response to the distinct selection pressures on introgressed alleles (Guillaume & Whitlock, 2007). Ecological conditions inducing correlational selection among multiple traits play a crucial role in the evolution of genetic covariance, as well as the stability of $\boldsymbol{G}$ across time (Jones et al., 2003). Relatedly, environmental fluctuations within populations can cause correlated shifts in selective optima across multiple traits, promoting the evolution of $\boldsymbol{G}$ by increasing pleiotropy and modularity in trait expression (do O & Whitlock, 2023). When the genotype-to-phenotype map is highly nonlinear, rapid and complex evolutionary changes in **G** may also occur that cannot be straightforwardly predicted by patterns of selection (Milocco & Salazar-Ciudad, 2022).

While **G** matrices are expected to rapidly evolve under many scenarios, extensive work has also investigated and provided empirical support for the micro- and macroevolutionary stability of $\boldsymbol{G}$ and $\boldsymbol{P}$ across time (Arnold et al., 2008; Delahaie et al., 2017; Estes & Arnold, 2007; Henry & Stinchcombe, 2023; McGlothlin et al., 2018; Rohner & Berger, 2023), motivating an emphasis on understanding the role of genetic (co)variances in channeling and constraining multivariate evolution (Chebib & Guillaume, 2017; Chevin, 2013; Garcia-Costoya et al., 2023; Phillips & Arnold, 1989; Schluter, 1996; Walsh & Blows, 2009). As a consequence, it is often underappreciated that estimated genetic and phenotypic (co)variances are the products of underlying genotype- and phenotype-by-environment interactions (Mats Björklund & Gustafsson, 2015; de Jong, 1989; Elgart et al., 2022; J. S. Martin et al., 2023; J. S. Martin, Westneat, et al., 2024; Pigliucci, 1996; Service, Philip M. & Rose, 1985; Sara Via & Lande, 1985). When such interactions are relevant for fitness and the benefits of responding

to environmental variation outweigh the costs of producing a response, plasticity can evolve in multivariate trait expression (de Jong, 1995; Draghi & Whitlock, 2012; Gavrilets & Scheiner, 1993; Haaland et al., 2021). Genetic and phenotypic (co)variances may, therefore, 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**).

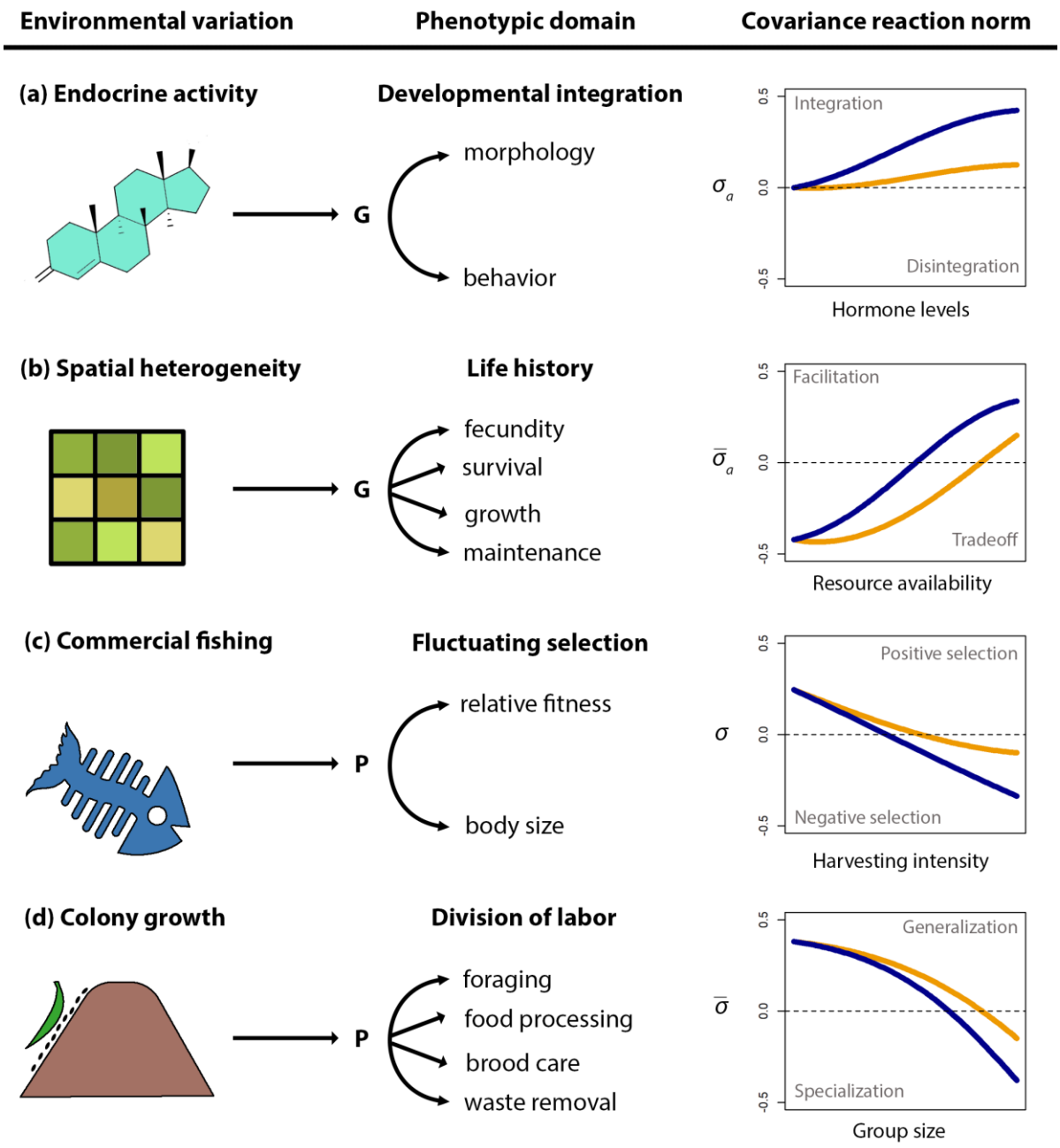
Consider, for example, that previous research across a wide range of taxa has shown that endocrine activity and the resulting hormonal milieu experienced during both prenatal and postnatal development exhibit dose-dependent effects on the integration (positive genetic covariance) of various morphological and behavioral phenotypes in adult organisms (e.g. in lizards, Yewers et al., 2017, Wittman et al., 2021; 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; **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 or facilitate 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 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 due to 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).

Modeling such multivariate environmental interactions is a crucial but easily overlooked step in effectively explaining the ongoing evolution of plastic phenotypes in a rapidly changing world (Westneat et al., 2019; Hudak & Dybdahl, 2023). 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 $\boldsymbol{G}$ and $\boldsymbol{P}$ matrices across contexts. Important empirical efforts have been made to investigate the fluctuations in $\boldsymbol{G}$ and $\boldsymbol{P}$ that result from plasticity in multivariate traits, as well as potentially rapid microevolution, in response to environmental heterogeneity and ongoing change in natural populations (e.g. Björklund et al., 2013; Bolund et al., 2015; Wood & Brodie, 2015). Analytic tools for efficiently inferring these complex patterns have been limited, however, particularly outside of the laboratory or agricultural contexts, where organisms are often exposed to continuous and high-dimensional patterns of spatial and temporal variation in their local microhabitats. Multivariate, multilevel regression models (also known as mixed effects, hierarchical, and random regression models) are well-established in the literature and widely applied for empirically estimating $\boldsymbol{G}$ and $\boldsymbol{P}$ (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

151 robustness for describing heritable (co)variation in wild populations, in comparison to classical
152 methods that rely on the assumptions of balanced breeding experiments or specific kin-class
153 comparisons (Kruuk & Hadfield, 2007). While current implementations of the multivariate animal
154 model can be used to investigate environmental effects on trait (co)variances, they remain limited in
155 their general application to complex environments and for field studies of natural populations.
156 Therefore, the present paper develops flexible extensions of current methods to better predict
157 variation in $\boldsymbol{G}$ and $\boldsymbol{P}$ matrices attributable to continuous, nonlinear, and multivariate environmental
158 effects, such as those discussion in **Fig. 1**, even in the absence of specialized breeding designs or
159 repeated individual measurements.

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.

Motivation for a novel method

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 (Lynch & Walsh, 1998; Sara Via & Lande, 1985). However, as argued above, environmental effects on $\boldsymbol{P}$ and $\boldsymbol{G}$ 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, variation in the rank-order of individuals' genetic values 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 contexts usually indicate heritable variation in plasticity due to GxE (Mitchell & Houslay, 2021). Consequently, while the character state model is extremely useful for systems experiencing a small number of environmental contexts, it will tend to 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 also reduce statistical power and tend to downwardly bias 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 slopes 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.

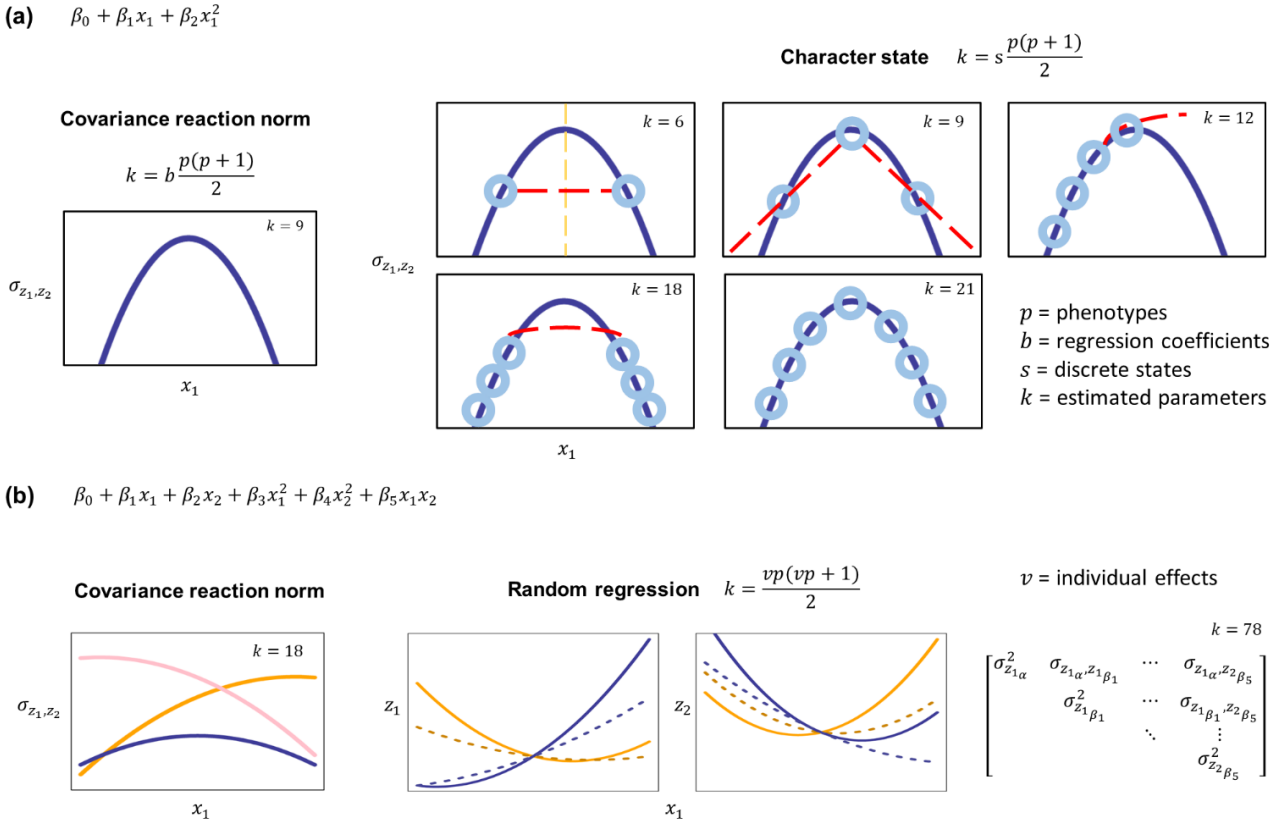
Random regression models can also be applied in the absence of experimental breeding designs when repeated individual-level measurements are available (Nussey et al., 2007). For instance, consider a scenario where the genetic or phenotypic (co)variance between behavior and morphology increases as a function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

A typical solution in this case would be to discretize age and estimate separate age class-specific (co)variance matrices of individuals' intercepts and slopes. This strategy falls prey to the same limitations of discretization discussed above for character state approaches. Discretization can instead be avoided using interaction effects, such as estimating random slopes for the effect of age x resource availability on both behavior and morphology. However, 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 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 for investigating GxE and PxE. This is accomplished by synthesizing character state and random regression approaches within a broader class of multilevel regression models. After formally outlining the CRN model, I subsequently validate this model for empirical application with simulations, 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 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. Link functions are ignored for simplicity. (**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, given an appropriate study design to estimate individual-level intercepts and slopes for both traits across environments. The solid and dashed lines show two hypothetical individual RNs for x_1 across two levels of x_2 (blue and orange). Interpolating the population CRN without prior knowledge requires over 4x as many parameters in comparison to the CRN.

Covariance reaction norms

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

Modeling genetic (co)variances as reaction norms

The $\boldsymbol{G}$ 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 in the **supplementary material**.

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 $\boldsymbol{G}$ matrix predicted from a CRN in the environmental context measured for subject *n*

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

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

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

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* $\boldsymbol{G}$ 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 $\boldsymbol{r_{a(X)_{1,p}}} = [r_{a(X_1)_{1,p}}, \ldots, r_{a(X_n)_{1,p}}]'$, which are predicted by the product of the environmental matrix $\boldsymbol{X}$ and the respective trait-specific CRN parameters (additive fixed effects, including global intercepts) for genetic variances $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_1}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_p}}\right]^{\top}$ and correlations $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{r_{1,2}}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{r_{p-1,p}}}\right]^{\top}$. There are, therefore, as many unique $\boldsymbol{G}_{(X)}$ matrices predicted as the number of unique multivariate environmental contexts, but this is achieved by estimating a much smaller set of CRN parameters. Herein I use the term "environmental context" to refer to any specific and unique combination of values for the given set of variables included in $\boldsymbol{X}$. For example, if one is interested in predicting how average seasonal temperature and precipitation affect the genetic (co)variance between growth and reproductive traits, each combination of temperature and precipitation values for a given season will define a different environmental context, with corresponding context-specific predictions for the expected genetic variances, correlations, and covariances among the traits measured under these conditions.

The log and inverse hyperbolic tangent link functions are respectively used to infer the trait-specific CRN parameters defined on the transformed linear scale of genetic variances and correlations. The link function $\operatorname{atanh}(r) = \frac{1}{2}\operatorname{logit}\left(\frac{r+1}{2}\right)$, also known as Fisher's z-transformation, extends the logit transformation defined for probability scale values to the scale of correlation coefficients. It is approximately linear in the range of $-0.3 \leq r \leq 0.3$, becoming increasingly sigmoidal in shape for larger correlation coefficients. The link function $\log(\sigma^2)$ facilitates linear prediction while ensuring positive values on the original scale of the necessarily non-zero variance terms. Importantly, this function implies exponential change in genetic variance $\boldsymbol{\sigma^2_{a(X)}} = e^{\boldsymbol{X\beta_{\sigma^2}}}$ across environments, in contrast to the quadratic change assumed by standard random regression models where individual slopes are expressed as Gaussian deviations from linear responses. Given that mean-centering is a common and generally well-motivated choice in regression analysis (Schielzeth, 2010; but see Mitchell & Houslay, 2021; Westneat et al., 2020), quadratic change is likely to be an unrealistic assumption for many traits and environments, as it implies symmetric increases in genetic variance across positive and negative values. In contrast, exponential functions allow for asymmetric change, such that, for example, temperature can both rapidly increase and, at the extreme, sharply decrease genetic variance irrespective of centering, consistent with established relationships for many metabolic and growth traits (Schulte, 2015). See **Eq. S8** and **Fig. S2** for an example of individual reaction norms generating exponential change in genetic variance. An alternative inverse softplus link function can also be an appropriate choice for genetic variances in the CRN model, as it produces less convex change on the variance scale in comparison to the more commonly used log link, providing greater flexibility for prediction (see **Fig. S3** and **supplementary materials** for further discussion).

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 alternative 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, i.e. the expected effect of individuals' genotypes on their phenotypes changes as a function of the environment. Given the assumption that environmental effects are independent of genetic effects, this GxE necessarily implies plasticity in the phenotype. Direct interpretation of the CRN fixed effect sizes will generally 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 CRN approach outlined above can also be taken to predict continuous effects on residual or permanent environmental (co)variances.

A regression analysis involving direct prediction of trait variances 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, given that any distributional parameter could be modeled as a function of covariates, giving rise to the possibility of triple, quadruple, etc. hierarchical models. 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.

Phenotypic CRNs

Empirical studies may lack the genetic information necessary to estimate **Eq. 2** 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; Martin et al., 2024). However, if multiple measurements are made on the same subjects across time, 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. 2** can be straightforwardly modified to produce a phenotypic CRN, described by a simplified multivariate normal distribution

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

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

Here $\boldsymbol{W}$ is a *n* x *i* matrix indexing repeated measurements for the random intercepts across *i* individuals and *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 $\boldsymbol{z}$. This matrix can also be introduced to **Eq. 2** when repeated measures are used to infer genetic effects. See **Eq. S9** for further generalization to random regression CRN models. The phenotypic random effects $\left[\boldsymbol{\mu}^{\top}_{(x)_1}, \ldots, \boldsymbol{\mu}^{\top}_{(x)_p}\right]^{\top}$ are 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 $\left[\boldsymbol{\beta}^{\top}_{\sigma^2_1}, \ldots, \boldsymbol{\beta}^{\top}_{\sigma^2_p}\right]^{\top}$ and correlations $\left[\boldsymbol{\beta}^{\top}_{r_{1,2}}, \ldots, \boldsymbol{\beta}^{\top}_{r_{p-1,p}}\right]^{\top}$ estimated on transformed scales,

418 equivalently to **Eq. 2** (not shown here for brevity). Any non-zero fixed effects predicting $\boldsymbol{P}_{(X)}$ provide
419 evidence for phenotype-by-environment (PxE) interactions, i.e. plasticity. See Bliard, Martin et al.
420 (2024) for detailed discussion and applications of bivariate phenotypic CRNs to detect life history
421 tradeoffs under multiple sampling regimes common in population ecology.

422 ***Model extensions***

423 The CRN is a method for facilitating the direct prediction of $\boldsymbol{P}$ and $\boldsymbol{G}$ matrices, rather than a
424 specific model structure per se, and so it is important to emphasize that the models presented in **Eq.**
425 **2-3** are simplified for ease of comprehension, focusing solely on linear prediction using fixed
426 regression coefficients. When applying the CRN for empirical analysis, researchers should always
427 consider whether this basic model structure requires further extension to effectively describe their
428 system. For instance, additional random intercepts and slopes may be useful CRN parameters to
429 capture stochastic change in trait (co)variances across environmental contexts, due to factors such as
430 unmeasured fluctuations in climate or habitat quality across nest sites or years (see the CRN tutorial
431 for example code, **data availability**). The choice of model structure for a specific empirical application
432 of the CRN is in principle no different than for any other regression analysis and so should be informed
433 by the same general considerations, such as whether it is pertinent to explicitly model measurement
434 error in environmental predictors, to account for censoring and non-random missingness of data, or
435 to use non-Gaussian distributions to describe trait (co)variation (see Bolker et al., 2009; Gelman et al.,
436 2020; Gelman & Hill, 2006; McElreath, 2020; Schielzeth, 2010). In the **supplementary material**, further
437 details are given on how the CRN can be extended for specific issues related to the use of repeated
438 individual measures, genetic prediction for unobserved phenotypes, and the estimation of cross-
439 environment correlations. I also present random regression CRN models that can be used to
440 investigate nested patterns of GxE within and across environments by combining RNs of trait means
441 and (co)variances (**Eq. S9**).

442

Statistical implementation

CRNs cannot currently be estimated using standard statistical software packages, due to a lack of in-built functionality for expressing elements of covariance matrices as generalized linear predictors. Fortunately, however, the extremely flexible Stan probabilistic programming language can be used to construct bespoke animal models of desired complexity within a Bayesian inferential framework, facilitating estimation of CRNs models using cutting-edge Markov Chain Monte Carlo (MCMC) methods (Hoffman & Gelman, 2011; Nishio & Arakawa, 2019; Martin & Jaeggi, 2022). This includes CRNs incorporating various kinds of complexity not discussed here, such as spatiotemporal autocorrelation. Interested readers should consult the Stan User's Guide and Reference Manual (https://mc-stan.org/docs/) and growing body of worked Case Studies (https://mc-stan.org/learn-stan/case-studies.html) for further details. Detailed discussion of contemporary Bayesian statistics and general issues in multilevel Bayesian modeling are also 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).

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). Therefore, estimation of the CRN in Stan is best achieved through use of a mathematically equivalent

but more efficient reparameterization of the $\boldsymbol{G}_{(X)}$ and $\boldsymbol{P}_{(X)}$ matrices than is described by the standard parameterization presented in **Eq. 2-3**. These computational solutions are extensively discussed in the supplementary material (**Eq. S1-5**). Fortunately, in many cases, such details can be safely ignored by empiricists applying the CRN model, as R functions and a tutorial are provided (see **data availability**) to straightforwardly facilitate these computational gains, while also generating more intuitive model estimates and predictions with respect to the standard model structure presented above.

Validation for Bayesian inference

I used simulation-based calibration, a gold-standard procedure for validating Bayesian algorithms (Gelman et al., 2020), to validate the proposed CRN model implementation in Stan. Briefly, simulation-based calibration involves simulating datasets across a broad range of effect size distributions, applying the proposed Bayesian model to these datasets, and then formally comparing the distributions of simulated and estimated parameter values (Talts et al., 2018; **Fig. S1a**). Further details on these simulations and discussion of simulation-based calibration as a methodology are provided in the **supplementary material**. As can been seen in supplementary **Fig. S1b**, results indicated that estimated CRN parameter values were congruent with and not systematically biased from the true values used to generate the simulated datasets, demonstrating that the proposed CRN implementation in Stan facilitates valid inference across a broad range of effect sizes, even in the absence of repeated measures or large sample size.

Comparison to alternative methods

The SBC validated the CRN model but did not test its performance in comparison to closely related approaches. Therefore, I performed an additional simulation to more directly compare the accuracy and power of inferences from the CRN model to a standard character state model relying on discretizing the environment, as well as a random regression model relying on estimation of individual intercepts and slopes. The simulation considered the effect of a single environmental variable on the genetic covariance among two Gaussian traits with modest marginal heritability $\left(\sigma_{G(x=0)}/\sigma_{P(x=0)} =\right.$

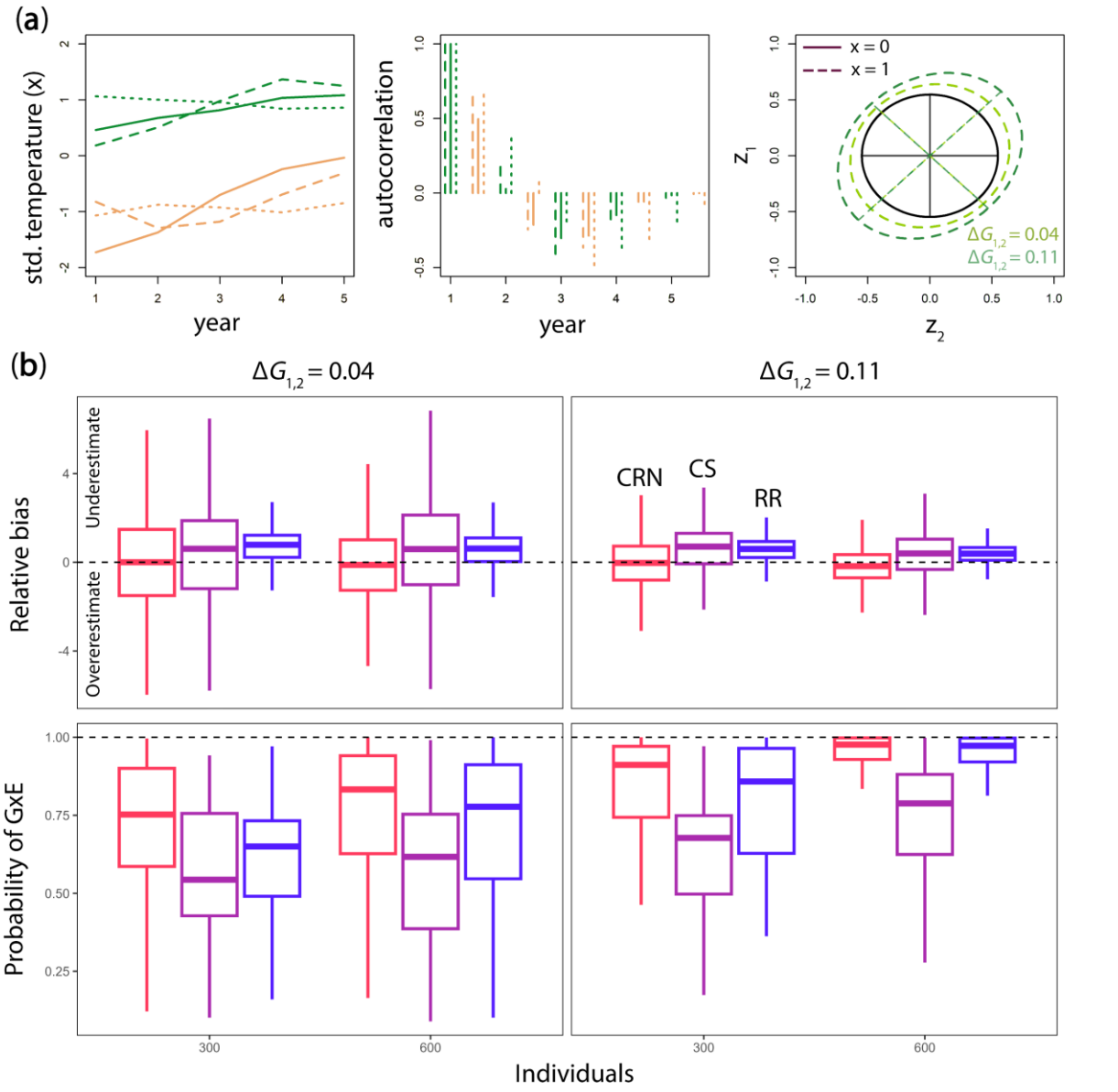
493 0.3), such as a growth and metabolic trait. The environmental variable was conceptualized as a
494 measure of average temperature during the summer and winter months, and individuals were
495 measured across 5 years, resulting in 5 * 2 = 10 environmental contexts. This relatively small number
496 of distinct environmental contexts ensured results relevant for a broader range of field studies lacking
497 deep time series. Stochastic time series of seasonal temperature were generated with varying degrees
498 of autocorrelation and weak directional trend (**Fig. 3a;** see **supplementary material** for details).
499 Individuals were assumed to have an annual life cycle and be measured during both seasons of a
500 randomly selected year, so that each subject had two repeated measurements, allowing for the
501 estimation of individual slopes in the random regression model and between-context genetic
502 correlations in the character state model. Continuous environmental information was retained for the
503 random regression and CRN analyses, while measures were binned into 'summer' and 'winter'
504 environments for the character state model. The simulated scenario isolated the performance of the
505 models under the idealized condition of a single varying environmental factor, minimizing divergence
506 in the number of free parameters, which is otherwise expected to favor the CRN in more complex
507 scenarios with multiple predictors and highly nonlinear effects (**Fig. 2**).

508 Effect sizes in evolutionary ecology tend to be rather small (Kimmel et al., 2023). Therefore,
509 in one simulated condition, I assumed small effect sizes for the impact of temperature on the genetic
510 variances and correlation ($\beta_{\sigma_\alpha^2} = 0.3$, $\beta_{r_\alpha} = 0.1$), resulting in a very small GxE effect ($\Delta G_{1,2} = 0.04$ for
511 $\Delta x = 1$) on the genetic covariance (**Fig. 3a**). The simulation thus assessed each models' performance
512 in estimating a minimally detectable but biologically realistic effect size. I also simulated a condition
513 with larger effect sizes ($\beta_{\sigma_\alpha^2} = 0.6, \beta_{r_\alpha} = 0.2, \Delta G_{1,2} = 0.11$) for comparison. In both conditions, cross-
514 environment correlations were positive but not unity ($r = 0.8$), producing modest among-individual
515 variation in the rank-order of genetic values across environmental contexts. Simulated datasets varied
516 in the number of subjects (300, 600) sampled to assess the influence of sample size on each model's
517 performance. I simulated and applied each model to 50 datasets per sample and effect size condition.

Performance was compared using the predicted change in genetic covariance among traits following a 1 SD increase in temperature from the mean (**Fig. 3a**). Note that this is a conservative comparison favorable to the alternative methods, as the character state model would otherwise show increasing bias for both smaller and larger temperature changes (due to standardizing temperature), and the random regression would show increasing bias for negative temperature changes (due to the assumption of quadratic change). I compared relative bias among models to capture accuracy in recovering the simulated GxE, calculated by $\frac{\Delta G_{1,2} - \Delta \hat{G}_{1,2}}{\Delta G_{1,2}}$ where $\Delta G_{1,2}$ and $\Delta \hat{G}_{1,2}$ are respectively the true and estimated change between $x = 0$ and $x = 1$. I also used the posterior probability $p(\Delta \hat{G}_{1,2} > 0)$ to quantify statistical power for Bayesian inference. This measure captures the degree of posterior uncertainty in the presence and direction of GxE.

Results are shown in **Fig. 3b**. The CRN exhibited accurate and unbiased detection of the true change in genetic covariance across sample and effect sizes, with increasing precision at larger sample and effect sizes. The CRN also exhibited the highest power for detecting GxE, which also increased steadily with sample and effect size. The character state model showed the lowest precision across conditions and tended to underestimate the larger effect size with the smaller sample. As expected, binning of temperature variation also resulted in the character state model having the lowest power for detecting GxE, as well as a very modest increase in power with increasing sample size. The random regression model exhibited high precision but largely because it consistently underestimated the true effect size (relative bias > 0). The power of the random regression was greater than the character state model and comparable but slightly lower than the CRN across conditions. These findings indicate that the CRN model performs well at modest sample and effect sizes. They show that, even for a simple and idealized univariate comparison, the CRN can outperform standard character state and random regression models in detecting and quantifying environmental effects on genetic covariance.

Figure 3. Model performance and comparison at minimal effect size.



Footnote. (**a**) Overview of the simulation. Left plot: each synthetic dataset contained 5 years of summer (green) and winter (orange) temperatures, standardized to unit variance (std.). See **supplementary materials** for details. Each simulation run generated a unique time series (different line types), with varying degrees of stochasticity and directional change. Middle plot: Time series also differed in their degree of temporal autocorrelation. Right plot: trait values were simulated for two Gaussian phenotypes z_1 and z_2 (e.g. a growth and metabolic trait), assuming small ($\Delta G_{1,2} = 0.04$) or moderate effects ($\Delta G_{1,2} = 0.11$) of temperature on their genetic covariance. Under an average temperature ($x = 0$), there was no covariance among traits (solid ellipse), while in relatively warm seasons ($x = 1$), there was positive genetic covariance (dashed ellipses) depending on effect size (color). (**b**) Results from 50 simulated datasets per sample and effect size condition. Performance was compared among models (red = covariance reaction norm [CRN], purple = character state [CS], blue = random regression [RR]) for detecting the true change in genetic covariance with increasing temperature. Top plot: relative bias (values close to 0 indicate accurate recovery of the effect size). Bottom plot: posterior probability (Bayesian power) in support of a positive temperature effect on the genetic covariance. Values closer to 1 indicate stronger support for the presence of GxE. Results are summarized with box plots.

Worked example: social niche specialization in meerkats

To further 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 my 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; J. S. 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 (Hendry, 2016; Hutchings, 2011; Kuzawa & Bragg, 2012; Paenke et al., 2007; Pfennig, 2021; Via et al., 1995), 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.

583 Additional random effects were included for each trait to capture individual-level permanent
584 environmental effects, group identity during observation, breeding season (accounting for temporal
585 trends within and across years from 1997 to 2018), and individual-by-season interactions. The three
586 phenotypes were modeled using binomial (half-days observed babysitting/total days) and Poisson
587 (count of pup feeding and minutes in sentinel activity) distributions. Following **Eq. 2** and using the
588 computational strategy explained in **Eq. S1-5**, the same environmental covariates used to predict
589 phenotypic means were also used to predict potential changes in quantitative genetic (co)variances
590 among cooperative behaviors. Consider that from the perspective of a gene, organismal attributes
591 such as sex, age, and dominance (serving as proxies for various attendant changes in hormonal
592 activity, social experiences, etc.) are just as much aspects of 'the environment' potentially modulating
593 its expression as more exogenous factors like group size (Hendry, 2016; Hutchings, 2011; Kuzawa &
594 Bragg, 2012; Paenke et al., 2007; Pfennig, 2021; S. Via et al., 1995). These covariates also allowed for
595 appropriately testing the independent (age, sex, and dominance adjusted) effect of group size on
596 genetic correlations among cooperative behaviors. A coding tutorial accompanying this worked
597 example is provided on Github (see **data availability**).

598 ***Results***

599 The CRN analysis uncovered continuous changes in the genetic variances and correlations of
600 meerkats' cooperative behaviors in response to the interactive effects of age, dominance status, and
601 sex, as well as the nonlinear effects of group size, providing clear evidence for GxE shaping the $\boldsymbol{G}$
602 matrix across environments. These effects are visualized as CRNs in **Fig. 4b-c** and summarized
603 quantitatively in **Table S1**. Firstly, considering genetic variances, increasing age was strongly
604 associated with greater genetic variance in babysitting behavior (BS), while age had weaker and more
605 uncertain effects on the genetic variance of foraging and pup feeding (FD) and vigilant guarding
606 behavior (GD). This indicates that heritable individual differences in BS are expected to increase across
607 the lifespan, independently of sex and dominance status. Sex did not have a main effect on the genetic

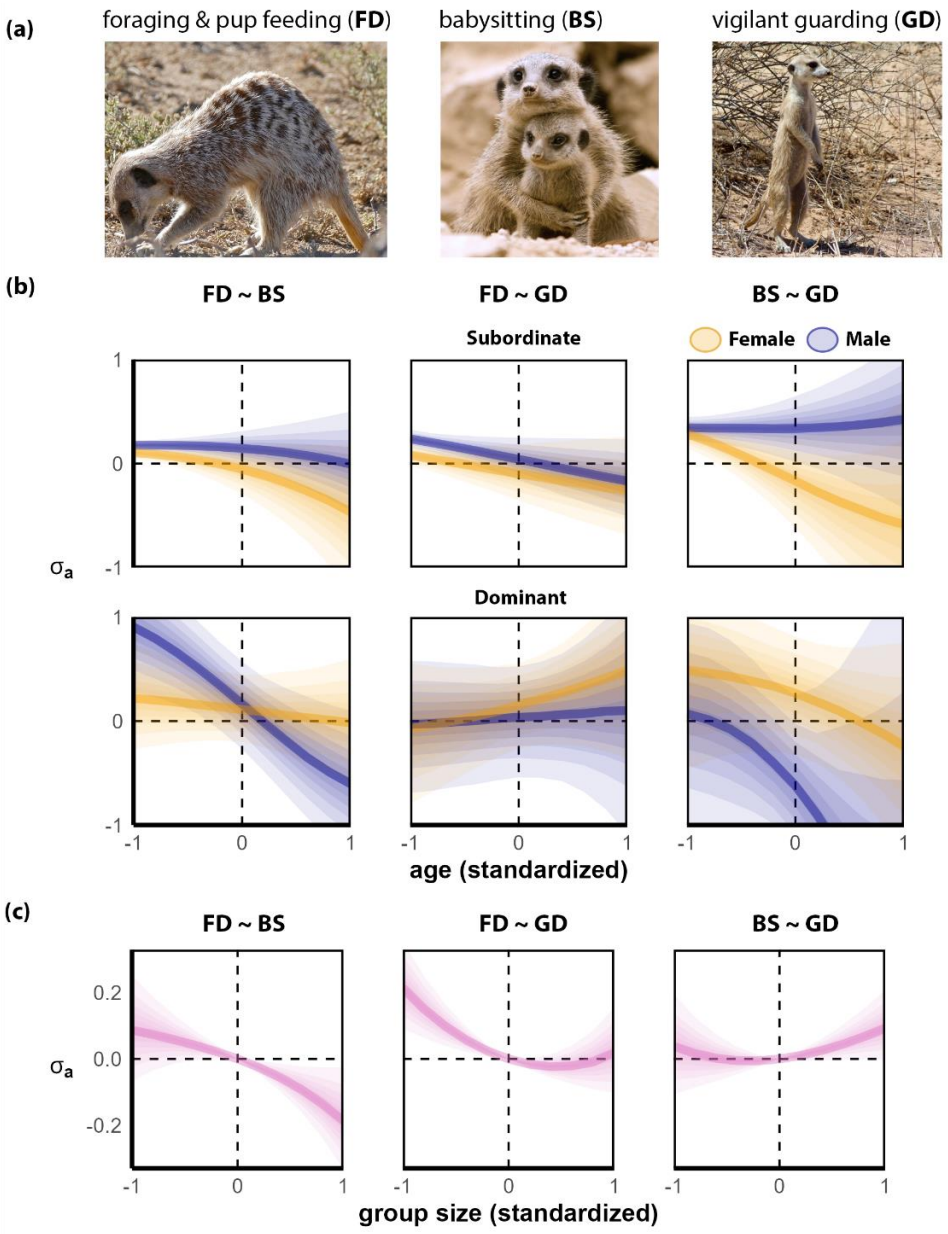
608 variance of any traits, while dominance status had moderate to strong positive effects on the genetic
609 variance of FD and GD. Changes in dominance status were, therefore, a primary driver of changes in
610 the magnitude of heritable individual differences in cooperative behaviors (personality). Dominant
611 individuals showed greater genetic variation than subordinates in their magnitude of FD and GD.
612 Multivariate interactions also occurred between age, sex, and dominance. Genetic variance in BS
613 reduced in response to the interaction of age and sex with dominance, while genetic variance in GD
614 increased as a function of the interaction between age and dominance as well as the three-way
615 interaction among age, sex, and dominance.

616 Environmental variation was also associated with changes in the genetic correlations among
617 cooperative behaviors (**Table S1**). Among subordinates, males exhibited relatively stronger genetic
618 correlations for BS ~ GD than females, which increased with age (**Fig. 4b**). Some evidence was found
619 for reversed sex effects among dominant individuals, but dominance effects exhibited moderate to
620 high statistical uncertainty overall. A clear main effect of age was observed for FD ~ BS, indicating that
621 this genetic correlation tended to decrease across the lifespan, with older individuals being more likely
622 to specialize in FD or BS than younger individuals. Negative age effects were also estimated for FD~BS
623 and BS~GD but with greater statistical uncertainty. Group size decreased both FD~BS and FD~GD,
624 independently of age, sex, and dominance effects, with more uncertainty in the positive effect of
625 group size on BS~GD. Evidence was also found for a positive quadratic effect of group size on FD ~ GD,
626 such that the negative effect was diminished for larger group sizes.

627 Combined effects of the multivariate environment on genetic variances and correlations
628 generate nonlinear CRNs that are visualized in **Fig. 5b-c**. Subordinate males typically show more
629 positive genetic (co)variances across ages than subordinate females, indicating more generalized
630 genetic effects on and heritable individual differences in cooperative behavior. Subordinate females
631 are in turn expected to show more negative genetic covariances among behaviors as they age (**Fig.**
632 **4b**). However, these patterns were complicated among dominant breeders. The direct effects of

633 dominance status on genetic correlations were highly uncertain (**Table S1**) and should be interpreted
634 cautiously, as is reflected by the much larger credible intervals for the predicted age CRNs of dominant
635 individuals (bottom row plots in **Fig. 4b**). Independently of these effects, negative genetic covariance
636 is expected between FD and BS in larger social groups, while a positive genetic covariance is expected
637 between BS and GD in larger social groups (**Fig. 4c**). The genetic covariance between FD and GD is
638 positive in small groups but declines nonlinearly and remains near to zero in average and larger than
639 average group sizes. Taken together, these results provide support for the prediction that fluctuations
640 in group size select for plasticity in the expression of generalized versus specialized cooperative
641 behavior across social groups. Consistent with prior research (Clutton-Brock et al., 2003), social niche
642 specialization is not observed on average across social groups. However, the CRN model reveals that
643 this is because small group sizes promote more positively integrated ($\sigma_a > 0$) genetic effects across
644 cooperative behaviors, while larger group sizes promote negative genetic correlations ($\sigma_a < 0$)
645 indicative of specialized performance of FD versus BS and GD tasks.

646 **Figure 5.** Multivariate CRN of cooperative behavior among meerkats.



647

---

648 **Footnote**. Posterior estimates are shown for multivariate and nonlinear environmental effects on the
649 genetic covariances σ_a among (**a**) meerkats' foraging and pup feeding (FD), babysitting (BS), and
650 vigilant guarding (GD) behavior. Creative commons picture credit: Bernard DUPONT and Jon Pinder
651 (Flickr). (**b**) Posterior CRNs for the interactive effects of sex (orange = female, blue = male), dominance
652 status (top row = subordinate, bottom = dominant), and age (units of months, SD standardized) on σ_a^2
653 (left row = FD~BS, center = FD~GD, right = BS~GD). Shaded bands indicate 10–90% posterior CI from
654 the darkest to lightest bands, respectively, while the dark lines indicate posterior median values. CRN
655 slopes greater or less than zero provide evidence for GxE interactions. (**c**) CRNs for the effect of group
656 size (units of 5, SD standardized) on σ_a, adjusted for the interactive effects of sex, age, and dominance
657 status. Dotted vertical lines indicate the expected covariance at the average group size (0), while
658 dotted horizontal lines indicate $\sigma_a = 0$, so that values above this line provide evidence for task
659 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). 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**). Current character state approaches for analyzing such 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, 3**). 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 in the wild (**Fig. 1**), and thus to study the development and adaptation of plastic phenotypes (**Fig. 5**). The CRN model proposed here provides a validated solution (**Fig. S1**) to this challenge that outperforms alternative approaches (**Fig. 3**), 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 (**Fig. 4a**), building on prior research by Houslay et al. (2021) using standard quantitative genetic methods, 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**), while also uncovering previously undescribed multivariate GxE interactions, such as among sex, age, and dominance status (**Fig. 4b**), which would require many more parameters and larger sample sizes to effectively estimate using standard methods (**Fig. 2**).

In addition to the benefits of the CRN model, researchers should also be mindful of its limitations and key areas for future extension. Statistical power will in many cases be a pertinent issue and potential limitation to consider when applying the CRN to estimate multivariate environmental effects on multiple traits. While 'power' in the classical sense does not straightforwardly translate to fully Bayesian inference, Bayesian power as considered here (the expected probability in support of a hypothesis) is an intuitively analogous quantity of similar importance. The reported simulations suggest that the CRN performs better than alternative methods even for univariate cases and exhibits steadily increasing power with increasing sample size (**Fig. 3b**). However, precisely estimating and detecting the partial effects of multiple variables will inevitably require larger sample sizes than for detecting the total effect of a single variable, as well as greater sampling effort across environmental contexts, to achieve comparable performance. This will particularly be the case for binary measures (e.g. survival or mating success), which generally provide less information per observation in comparison to continuous data (Fay et al., 2022). Even in the univariate case, my simulations suggest that confident detection of small effects will require large sample sizes. Therefore, despite the promising results presented here, investigations of multivariate CRNs will be most fruitfully accomplished with data from experiments and long-term field studies facilitating both large sample sizes and a diverse set of environmental contexts through which the direct influence of distinct factors can be effectively disentangled, such as in the worked example on meerkat social behavior. While general heuristics are useful, power can also vary widely as a function of data structure, model complexity, and effect sizes appropriate for the context under consideration (Johnson et al., 2015). Researchers should, therefore, consider carrying out their own a priori power analyses for the conditions relevant to their intended application of a CRN model.

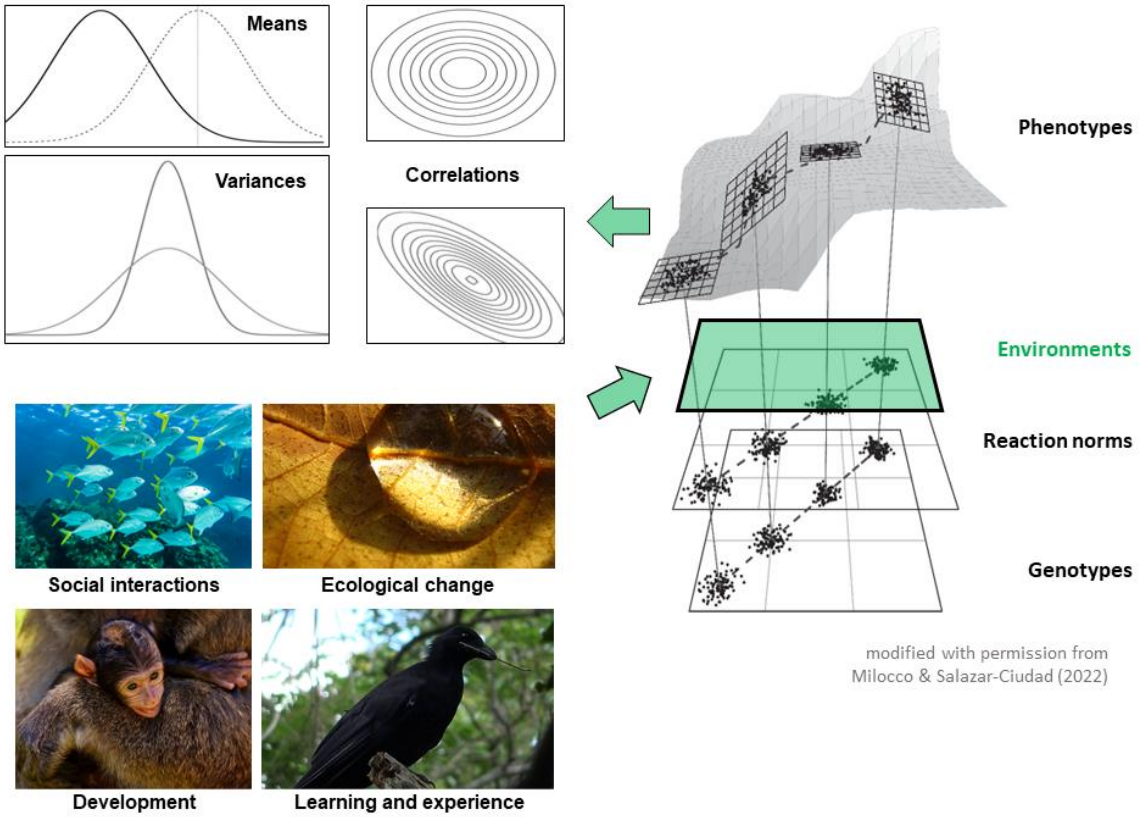
Fortunately, the CRN is expected to provide unbiased inferences even when statistical uncertainty is relatively large (**Fig. S1, 3**). However, the biological validity of these inferences will always be contingent on how well the structure of the CRN approximates the underlying empirical reality being described. As with any regression analysis, researchers should be particularly cautious in

naively interpreting direct effects in the CRN as indicative of causal effects, without thoughtful application of the principles and contemporary techniques required for causal inference (Pearl, Glymour, Jewell, 2016), particularly in observational studies of wild populations. Dynamic feedback between organismal traits and environments may, for instance, induce so-called collider biases (Pearl et al., 2016) that generate spurious associations among causally interdependent environmental factors and trait (co)variances. In this regard, investigating how the CRN performs for environmentally modifying and niche constructing traits (e.g. building on recent quantitative genetic models, Fogarty & Wade, 2022) will be a valuable direction for future research. More generally, future development of the CRN model should aim to better incorporate dynamic relationships to infer directionality and potential reciprocal causality between the environment effects shaping trait expression across multiple spatial and temporal scales. Methods extending ordinary differential equations in Stan to study multivariate trait evolution across phylogenetic trees (Ringen et al., 2021) provide a useful starting point for considering how this might be accomplished, extending the CRN model to study dynamic patterns of GxE using time series of sufficient depth.

Regardless of the specific CRN model being employed, particular attention should always be given to appropriate handling of spatiotemporal variation, as quantitative genetic inference in the wild can be easily biased by unadjusted environmental effects that are correlated among related individuals across space and time (Kruuk & Hadfield, 2007; Munar-Delgado et al., 2023). When interpreting the results of a CRN model, researchers should consider whether apparent evidence of GxE may instead reflect unmeasured or unquantified associations between genes and environmental contexts that were not appropriately adjusted for. For instance, genetic subgroups that are differentially distributed in space (Hadfield et al., 2010) as well as the effects of drift, inbreeding, and selection on a population across time (Sorensen et al., 2001; Sorensen & Kennedy, 1984) can both confound inferences of environmentally induced changes in genetic expression. Statistical tools for handling such issues are already available in the literature and, using the flexible probabilistic programming provided by Stan, can be readily incorporated into a CRN analysis.

737 Further exploration of how the CRN model performs and where in may exhibit bias in
738 recovering patterns of GxE and PxE under more realistic ecological scenarios, where spatiotemporal
739 dynamics are often difficult to effectively measure and formalize, is another important direction for
740 future methodological research, building on the simplified simulations presented here (where
741 temporal autocorrelation was appropriately incorporated into the CRN analysis, **Fig. 3**). In this regard,
742 a key extension of the basic CRN model (**Eq. 2**) will be to use more flexible non-parametric and
743 generalized additive functions, such as splines or Gaussian processes (Pedersen et al., 2019; Riutort-
744 Mayol et al., 2022), that can better capture nonlinear spatiotemporal autocorrelation and
745 environmental effects on trait (co)variances, which are otherwise difficult to estimate with standard
746 polynomials (see the CRN tutorial for example code, **data availability**). Given thoughtful consideration
747 of these limitations and potential extensions, future applications of the CRN model have the potential
748 to greatly enhance our understanding of the evolutionary ecology of multivariate plasticity across a
749 variety of phenotypes in the wild (**Fig. 5**).

Figure 5. Environmental effects on the expression and evolution of multivariate phenotypes.



Footnote. A conceptual figure of GxE and PxE for multivariate traits, modified with permission from Milocco and Salazar-Ciudad (2022). The phenotype-to-genotype map, shown here by lines connecting populations of genotypes (lowest surface) to distributions of phenotypes (highest), is mediated through RNs and the distribution of environments within and across generations. RNs not only structure the expression of trait means, but also the variances, correlations, and (co)variances among traits (i.e. CRNs). Therefore, $\boldsymbol{G}$ and $\boldsymbol{P}$ matrices describing the mapping between genetic and phenotypic variation are often highly sensitive to the environmental contexts in which individuals are measured (GxE and PxE, indicated by green arrows). CRNs may evolve in response to diverse environmental contexts such as the quality and consistency of early parental care, opportunities for and costs of learning, variability and harshness of the climate, habitat degradation, magnitude and predictability of local resources, the density of predators and parasites, the strengths of intra and intersexual competition, social network position and mating system, food web structure, etc. When such environments change (dotted lines) and developmental and/or contextual plasticity has evolved in a population, trait (co)variances may rapidly respond to spatiotemporal heterogeneity within and across generations (top layer planes). Mechanistically and ecologically informed CRN models can be used to better predict how GxE will shape the expression and adaptive evolution of multivariate traits in response to ongoing socio-eco-evolutionary dynamics (Martin et al., 2024). Creative commons picture credit: NickJack and Alexas_Fotos (Pixabay) and Luz Adriana Villa and Corvus moneduloides (Flickr).

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.

Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., & Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution; International Journal of Organic Evolution*, *62*(10), 2451–2461.

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. (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

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.

Brommer, J., Class, B., & Covarrubias-Pazaran, G. (2019). Multivariate Mixed Models in Ecology and Evolutionary biology: Inferences and implementation in R. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/hs38a

797 Brooks, R., Hunt, J., Blows, M. W., Smith, M. J., Bussière, L. F., & Jennions, M. D. (2005). Experimental
798 evidence for multivariate stabilizing sexual selection. *Evolution; International Journal of*
799 *Organic Evolution*, *59*(4), 871–880.
800 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo,
801 J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical*
802 *Software*, *76*. https://doi.org/10.18637/jss.v076.i01
803 Carvalho, M. J. A., & Mirth, C. K. (2015). Coordinating morphology with behavior during development:
804 an integrative approach from a fly perspective. *Frontiers in Ecology and Evolution*, *3*.
805 https://doi.org/10.3389/fevo.2015.00005
806 Chang, C.-C., Moiron, M., Sánchez-Tójar, A., Niemela, P., & Laskowski, K. (2023). What's the meta-
807 analytic evidence for life-history trade-offs at the genetic level? In *Authorea*.
808 https://doi.org/10.22541/au.168795190.00787983/v1
809 Chebib, J., & Guillaume, F. (2017). What affects the predictability of evolutionary constraints using a
810 G-matrix? The relative effects of modular pleiotropy and mutational correlation. *Evolution;*
811 *International Journal of Organic Evolution*, *71*(10), 2298–2312.
812 Chevin, L.-M. (2013). Genetic constraints on adaptation to a changing environment: Genetic
813 constraints in a changing environment. *Evolution; International Journal of Organic Evolution*,
814 *67*(3), 708–721.
815 Clutton-Brock, T. H., Brotherton, P. N., Russell, A. F., O'Riain, M. J., Gaynor, D., Kansky, R., Griffin, A.,
816 Manser, M., Sharpe, L., McIlrath, G. M., Small, T., Moss, A., & Monfort, S. (2001). Cooperation,
817 control, and concession in meerkat groups. *Science (New York, N.Y.)*, *291*(5503), 478–481.
818 Clutton-Brock, T. H., Russell, A. F., & Sharpe, L. L. (2003). Meerkat helpers do not specialize in particular
819 activities. *Animal Behaviour*, *66*(3), 531–540.
820 Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

821 Damián, X., Ochoa-López, S., Gaxiola, A., Fornoni, J., Domínguez, C. A., & Boege, K. (2020). Natural
822 selection acting on integrated phenotypes: covariance among functional leaf traits increases
823 plant fitness. *The New Phytologist*, *225*(1), 546–557.
824 Danchin, É., & Wagner, R. H. (2010). Inclusive heritability: combining genetic and non-genetic
825 information to study animal behavior and culture. *Oikos* , *119*(2), 210–218.
826 de Jong, G. (1989). Phenotypically plastic characters in isolated populations. In A. Fontdevila (Ed.),
827 *Evolutionary Biology of Transient Unstable Populations* (pp. 3–18). Springer.
828 de Jong, G. (1995). Phenotypic plasticity as a product of selection in a variable environment. *The*
829 *American Naturalist*, *145*(4), 493–512.
830 de la Mata, R., Zas, R., Bustingorri, G., Sampedro, L., Rust, M., Hernandez-Serrano, A., & Sala, A. (2022).
831 Drivers of population differentiation in phenotypic plasticity in a temperate conifer: A 27-year
832 study. *Evolutionary Applications*, *15*(11), 1945–1962.
833 Dean, D. S., & Majumdar, S. N. (2008). Extreme value statistics of eigenvalues of Gaussian random
834 matrices. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *77*(4 Pt 1), 041108.
835 Delahaie, B., Charmantier, A., Chantepie, S., Garant, D., Porlier, M., & Teplitsky, C. (2017). Conserved
836 G-matrices of morphological and life-history traits among continental and island blue tit
837 populations. *Heredity*, *119*(2), 76–87.
838 Dingemanse, Niels J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour:
839 mixed-effect modelling approaches. *The Journal of Animal Ecology*, *82*(1), 39–54.
840 Dingemanse, Niels Jeroen, Araya-Ajoy, Y. G., & Westneat, D. F. (2021). Most published selection
841 gradients are underestimated: Why this is and how to fix it. *Evolution; International Journal of*
842 *Organic Evolution*, *75*(4), 806–818.
843 do O, I., & Whitlock, M. C. (2023). The evolution of genetic covariance and modularity as a result of
844 multigenerational environmental fluctuation. *Evolution Letters*, *7*(6), 457–466.
845 Draghi, J. A., & Whitlock, M. C. (2012). Phenotypic plasticity facilitates mutational variance, genetic
846 variance, and evolvability along the major axis of environmental variation: Phenotypic

plasticity facilitates mutational variance, genetic variance, and evolvability. *Evolution; International Journal of Organic Evolution*, *66*(9), 2891–2902.

Elgart, M., Goodman, M. O., Isasi, C., Chen, H., Morrison, A. C., de Vries, P. S., Xu, H., Manichaikul, A. 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.

Fay, R., Authier, M., Hamel, S., Jenouvrier, S., Pol, M., Cam, E., Gaillard, J.-M., Yoccoz, N. G., Acker, P., Allen, A., Aubry, L. M., Bonenfant, C., Caswell, H., Coste, C. F. D., Larue, B., Le Coeur, C., Gamelon, M., Macdonald, K. R., Moiron, M., … Sæther, B.-E. (2022). Quantifying fixed individual heterogeneity in demographic parameters: Performance of correlated random effects for Bernoulli variables. *Methods in Ecology and Evolution*, *13*(1), 91–104.

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.

Garcia-Costoya, G., Williams, C. E., Faske, T. M., Moorman, J. D., & Logan, M. L. (2023). Evolutionary constraints mediate extinction risk under climate change. *Ecology Letters*, *26*(4), 529–539.

872 Gavrilets, S., & Scheiner, S. M. (1993). The genetics of phenotypic plasticity. V. Evolution of reaction
873 norm shape. *Journal of Evolutionary Biology*, *6*(1), 31–48.
874 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data*
875 *Analysis Third edition*. CRC Press.
876 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J.,
877 Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. In *arXiv [stat.ME]*. arXiv.
878 http://arxiv.org/abs/2011.01808
879 Grebe, N. M., Fitzpatrick, C., Sharrock, K., Starling, A., & Drea, C. M. (2019). Organizational and
880 activational androgens, lemur social play, and the ontogeny of female dominance. *Hormones*
881 *and Behavior*, *115*, 104554.
882 Guillaume, F., & Whitlock, M. C. (2007). Effects of migration on the genetic covariance matrix.
883 *Evolution; International Journal of Organic Evolution*, *61*(10), 2398–2409.
884 Haaland, T. R., Wright, J., & Ratikainen, I. I. (2021). Individual reversible plasticity as a genotype-level
885 bet-hedging strategy. *Journal of Evolutionary Biology*, *34*(7), 1022–1033.
886 Haave-Audet, E., Besson, A. A., Nakagawa, S., & Mathot, K. J. (2022). Differences in resource
887 acquisition, not allocation, mediate the relationship between behaviour and fitness: a
888 systematic review and meta-analysis. *Biological Reviews of the Cambridge Philosophical*
889 *Society*, *97*(2), 708–731.
890 Heino, M., Díaz Pauli, B., & Dieckmann, U. (2015). Fisheries-Induced Evolution. *Annual Review of*
891 *Ecology, Evolution, and Systematics*, *46*(1), 461–480.
892 Henderson, C. R., Jr. (1982). Analysis of covariance in the mixed model: higher-level,
893 nonhomogeneous, and random regressions. *Biometrics*, *38*(3), 623–640.
894 Hendry, A. P. (2016). Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics.
895 *The Journal of Heredity*, *107*(1), 25–41.
896 Henry, G. A., & Stinchcombe, J. R. (2023). G-matrix stability in clinally diverging populations of an
897 annual weed. *Evolution; International Journal of Organic Evolution*, *77*(1), 49–62.

Hoffman, M. D., & Gelman, A. (2011). The no-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. In *arXiv [stat.CO]*. arXiv. https://jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf

Houslay, T. M., Nielsen, J. F., & Clutton-Brock, T. H. (2021). Contributions of genetic and nongenetic sources to variation in cooperative behavior in a cooperative mammal. *Evolution; International Journal of Organic Evolution*, *75*(12), 3071–3086.

Huang, Z.-Y., & Robinson, G. E. (1996). Regulation of honey bee division of labor by colony age demography. *Behavioral Ecology and Sociobiology*, *39*(3), 147–158.

Huber, S. E., Lenz, B., Kornhuber, J., & Müller, C. P. (2017). Prenatal androgen-receptor activity has organizational morphological effects in mice. *PloS One*, *12*(11), e0188752.

Hudak, A., & Dybdahl, M. (2023). Phenotypic plasticity under the effects of multiple environmental variables. *Evolution; International Journal of Organic Evolution*, *77*(6), 1370–1381.

Hutchings, J. A. (2011). Old wine in new bottles: reaction norms in salmonid fishes. *Heredity*, *106*(3), 421–437.

Jeanson, R., Fewell, J. H., Gorelick, R., & Bertram, S. M. (2007). Emergence of Increased Division of Labor as a Function of Group Size. *Behavioral Ecology and Sociobiology*, *62*(2), 289–298.

Johnson, P. C. D., Barry, S. J. E., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, *6*(2), 133–142.

Jones, A. G., Arnold, S. J., & Bürger, R. (2003). Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution; International Journal of Organic Evolution*, *57*(8), 1747–1760.

Jones, A. G., Bürger, R., & Arnold, S. J. (2014). Epistasis and natural selection shape the mutational architecture of complex traits. *Nature Communications*, *5*(1), 3709.

Jusufovski, D., & Kuparinen, A. (2020). Exploring individual and population eco-evolutionary feedbacks under the coupled effects of fishing and predation. *Fisheries Research*, *231*, 105713.

924 Karsai, I., & Wenzel, J. W. (1998). Productivity, individual-level and colony-level flexibility, and
925 organization of work as consequences of colony size. *Proceedings of the National Academy of*
926 *Sciences of the United States of America*, *95*(15), 8665–8669.
927 Koch, E. L., Sbilordo, S. H., & Guillaume, F. (2020). Genetic variance in fitness and its cross-sex
928 covariance predict adaptation during experimental evolution. *Evolution; International Journal*
929 *of Organic Evolution*, *74*(12), 2725–2740.
930 Koutsidi, M., Lazaris, A., Peristeraki, P., Tserpes, G., & Tzanatos, E. (2024). Quantification of
931 intraspecific and interspecific competition in fish species of the Aegean Sea. *ICES Journal of*
932 *Marine Science: Journal Du Conseil*, *81*(2), 334–347.
933 Kruuk, L. E. B., & Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity
934 between relatives. *Journal of Evolutionary Biology*, *20*(5), 1890–1903.
935 Kruuk, Loeske E. B. (2004). Estimating genetic parameters in natural populations using the 'animal
936 model.' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*,
937 *359*(1446), 873–890.
938 Kuzawa, C. W., & Bragg, J. M. (2012). Plasticity in Human Life History Strategy: Implications for
939 Contemporary Human Variation and the Evolution of Genus Homo. *Current Anthropology*,
940 *53*(S6), S369–S382.
941 Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain body size
942 allometry. *Evolution; International Journal of Organic Evolution*, *33*(1Part2), 402–416.
943 Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution;*
944 *International Journal of Organic Evolution*, *37*(6), 1210–1226.
945 Lande, Russell. (1976). NATURAL SELECTION AND RANDOM GENETIC DRIFT IN PHENOTYPIC
946 EVOLUTION. *Evolution; International Journal of Organic Evolution*, *30*(2), 314–334.
947 Lankau, R. A. (2011). Rapid Evolutionary Change and the Coexistence of Species. *Annual Review of*
948 *Ecology, Evolution, and Systematics*, *42*(Volume 42, 2011), 335–354.

949 Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Journal*
950 *of the Royal Statistical Society. Series C, Applied Statistics*, *55*(2), 139–185.
951 Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly
952 informative priors in Bayesian analyses. *Oikos* , *128*(7), 912–928.
953 Link, W. A., & Eaton, M. J. (2011). On thinning of chains in MCMC. *Methods in Ecology and Evolution*,
954 *3*(1). https://doi.org/10.1111/j.2041-210X.2011.00131.x
955 Lofeu, L., Brandt, R., & Kohlsdorf, T. (2017). Phenotypic integration mediated by hormones:
956 associations among digit ratios, body size and testosterone during tadpole development. *BMC*
957 *Evolutionary Biology*, *17*(1), 175.
958 Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.
959 MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization
960 of quantitative variables. *Psychological Methods*, *7*(1), 19–40.
961 Martin, J. (2021). Estimating nonlinear selection on behavioral reaction norms. In *EcoEvoRxiv*.
962 https://doi.org/10.32942/osf.io/u26tz
963 Martin, J. S., Araya-Ajoy, Y., Dingemanse, N., Wilson, A., & Westneat, D. (2024). Estimating nonlinear
964 selection on reaction norms: A general framework for labile traits. In *EcoEvoRxiv*.
965 https://ecoevorxiv.org/repository/view/4124/
966 Martin, J. S., & Jaeggi, A. V. (2022). Social animal models for quantifying plasticity, assortment, and
967 selection on interacting phenotypes. *Journal of Evolutionary Biology*, *35*(4), 520–538.
968 Martin, J. S., Jaeggi, A. V., & Koski, S. E. (2023). The social evolution of individual differences: Future
969 directions for a comparative science of personality in social behavior. *Neuroscience and*
970 *Biobehavioral Reviews*, *144*, 104980.
971 Martin, J. S., Westneat, D., Wilson, A. J., Dingemanse, N. J., & Araya-Ajoy, Y. G. (2024). *Frequency-*
972 *dependence favors social plasticity and facilitates socio-eco-evolutionary feedback in*
973 *fluctuating environments*. https://ecoevorxiv.org/repository/view/8058/

974 Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power
975 in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
976 McElreath, R. (2020). *Statistical Rethinking; A Bayesian Course with Examples in R and Stan; Second*
977 *Edition*. CRC Press.
978 McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Mahler, D. L., Kolbe, J. J., Losos, J. B., & Brodie, E. D.,
979 3rd. (2018). Adaptive radiation along a deeply conserved genetic line of least resistance in
980 Anolis lizards. *Evolution Letters*, *2*(4), 310–322.
981 Milocco, L., & Salazar-Ciudad, I. (2022). Evolution of the G Matrix under Nonlinear Genotype-
982 Phenotype Maps. *The American Naturalist*, *199*(3), 420–435.
983 Mitchell, D. J., & Houslay, T. M. (2021). Context-dependent trait covariances: how plasticity shapes
984 behavioral syndromes. *Behavioral Ecology: Official Journal of the International Society for*
985 *Behavioral Ecology*, *32*(1), 25–29.
986 Montoya, E. R., Terburg, D., Bos, P. A., Will, G.-J., Buskens, V., Raub, W., & van Honk, J. (2013).
987 Testosterone administration modulates moral judgments depending on second-to-fourth
988 digit ratio. *Psychoneuroendocrinology*, *38*(8), 1362–1369.
989 Nishio, M., & Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler
990 for estimating genetic parameters and breeding values. *Genetics, Selection, Evolution: GSE*,
991 *51*(1), 73.
992 Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic
993 plasticity in wild populations. *Journal of Evolutionary Biology*, *20*(3), 831–844.
994 Oh, K. P., & Shaw, K. L. (2013). Multivariate sexual selection in a rapidly evolving speciation phenotype.
995 *Proceedings. Biological Sciences / The Royal Society*, *280*(1761), 20130482.
996 Oomen, R. A., & Hutchings, J. A. (2022). Genomic reaction norms inform predictions of plastic and
997 adaptive responses to climate change. *The Journal of Animal Ecology*, *91*(6), 1073–1087.
998 Paenke, I., Sendhoff, B., & Kawecki, T. J. (2007). Influence of plasticity and learning on evolution under
999 directional selection. *The American Naturalist*, *170*(2), E47-58.

Pearl, J. Glymour, M. Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, *7*, e6876.

Pfennig, D. W. (2021). *Phenotypic Plasticity & Evolution* (1st Edition). CRC Press.

Phillips, P. C., & Arnold, S. J. (1989). Visualizing Multivariate Selection. *Evolution; International Journal of Organic Evolution*, *43*(6), 1209–1222.

Pigliucci, M. (1996). Modelling phenotypic plasticity. II. Do genetic correlations matter? *Heredity*, *77 (Pt 5)*, 453–460.

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, B.-J., Kong, L., Deng, M., Wang, J.-R., Chen, G.-Y., Jiang, Q.-T., Lan, X.-J., Li, W., Wei, Y.-M., & Zheng, Y.-L. (2019). Transcriptional reference map of hormone responses in wheat spikes. *BMC Genomics*, *20*(1), 390.

R Core Team. (2023). *R: A language and environment for statistical computing*.

Revell, L. J. (2007). The G matrix under fluctuating correlational mutation and selection. *Evolution; International Journal of Organic Evolution*, *61*(8), 1857–1872.

Ringen, E., Martin, J. S., & Jaeggi, A. (2021). *Novel phylogenetic methods reveal that resource-use intensification drives the evolution of "complex" societies*. https://ecoevorxiv.org/repository/object/4119/download/8181/

Riutort-Mayol, G., Bürkner, P.-C., Andersen, M. R., Solin, A., & Vehtari, A. (2022). Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and Computing*, *33*(1), 17.

Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution; International Journal of Organic Evolution*, *50*(4), 1392–1403.

Roff, D. A., & Fairbairn, D. J. (2012). A test of the hypothesis that correlational selection generates genetic correlations. *Evolution; International Journal of Organic Evolution*, *66*(9), 2953–2960.

Rohner, P. T., & Berger, D. (2023). Developmental bias predicts 60 million years of wing shape evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(19), e2211210120.

Rolian, C. (2020). Ecomorphological specialization leads to loss of evolvability in primate limbs. *Evolution; International Journal of Organic Evolution*, *74*(4), 702–715.

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genetics, Selection, Evolution: GSE*, *42*(1), 8.

Royauté, R., Hedrick, A., & Dochtermann, N. A. (2020). Behavioural syndromes shape evolutionary trajectories via conserved genetic architecture. *Proceedings. Biological Sciences / The Royal Society*, *287*(1927), 20200183.

Schaum, C.-E., Buckling, A., Smirnoff, N., & Yvon-Durocher, G. (2022). Evolution of thermal tolerance and phenotypic plasticity under rapid and slow temperature fluctuations. *Proceedings. Biological Sciences / The Royal Society*, *289*(1980), 20220834.

Schluter, D. (1996). Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution; International Journal of Organic Evolution*, *50*(5), 1766–1774.

Service, Philip M., & Rose, M. R. (1985). Genetic Covariation Among Life-History Components: The Effect of Novel Environments. *Evolution; International Journal of Organic Evolution*, *39*(4), 943–945.

Sharpe, D. M. T., & Hendry, A. P. (2009). Life history change in commercially exploited fish stocks: an analysis of trends across studies. *Evolutionary Applications*, *2*(3), 260–275.

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268.

Thomas, M. L., & Elgar, M. A. (2003). Colony size affects division of labour in the ponerine ant Rhytidoponera metallica. *Die Naturwissenschaften*, *90*(2), 88–92.

1049 Thomson, C. E., Winney, I. S., Salles, O. C., & Pujol, B. (2018). A guide to using a multiple-matrix animal
1050 model to disentangle genetic and nongenetic causes of phenotypic variance. *PloS One*, *13*(10),
1051 e0197720.
1052 Trumble, B. C., Jaeggi, A. V., & Gurven, M. (2015). Evolving the neuroendocrine physiology of human
1053 and primate cooperation and collective action. *Philosophical Transactions of the Royal Society*
1054 *of London. Series B, Biological Sciences*, *370*(1683), 20150014.
1055 Ulrich, Y., Saragosti, J., Tokita, C. K., Tarnita, C. E., & Kronauer, D. J. C. (2018). Fitness benefits and
1056 emergent division of labour at the onset of group living. *Nature*, *560*(7720), 635–638.
1057 Uusi-Heikkilä, S. (2020). Implications of size-selective fisheries on sexual selection. *Evolutionary*
1058 *Applications*, *13*(6), 1487–1500.
1059 van Noordwijk, A. J., & de Jong, G. (1986). Acquisition and Allocation of Resources: Their Influence on
1060 Variation in Life History Tactics. *The American Naturalist*, *128*(1), 137–142.
1061 Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Van Tienderen, P. H. (1995).
1062 Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*,
1063 *10*(5), 212–217.
1064 Via, Sara, & Lande, R. (1985). Genotype-environment interactions and the evolution of phenotypic
1065 plasticity. *Evolution; International Journal of Organic Evolution*, *39*(3), 505–522.
1066 vom Saal, F. S. (1979). Prenatal exposure to androgen influences morphology and aggressive behavior
1067 of male and female mice. *Hormones and Behavior*, *12*(1), 1–11.
1068 Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate
1069 Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution,*
1070 *and Systematics*, *40*(1), 41–59.
1071 Westneat, D. F., Potts, L. J., Sasser, K. L., & Shaffer, J. D. (2019). Causes and Consequences of
1072 Phenotypic Plasticity in Complex Environments. *Trends in Ecology & Evolution*, *34*(6), 555–568.

1073 Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., &
1074 Nussey, D. H. (2010). An ecologist's guide to the animal model. *The Journal of Animal Ecology*,
1075 *79*(1), 13–26.
1076 Wittman, T. N., Robinson, C. D., McGlothlin, J. W., & Cox, R. M. (2021). Hormonal pleiotropy structures
1077 genetic covariance. *Evolution Letters*, *5*(4), 397–407.
1078 Wood, C. W., & Brodie, E. D., 3rd. (2015). Environmental effects on the structure of the G-matrix.
1079 *Evolution; International Journal of Organic Evolution*, *69*(11), 2927–2940.

Supplementary Material

This supplement provides more extensive details on the simulations presented in the main text, as well as additional discussion on key extensions of the basic CRN model presented in Eq. 2 of the main text.

Contents

Computational efficiency

As noted in the main text, direct prediction of large positive-definite covariance matrices is computationally challenging. Therefore, the CRN can be most efficiently estimated in Stan by using reparameterizations of the $\boldsymbol{G}_{(X)}$ and $\boldsymbol{P}_{(X)}$ matrices that are mathematically equivalent to the more intuitive parameterizations presented in Eq. 2-4. Firstly, the *p* x *p* correlation matrix $\boldsymbol{R}_{\boldsymbol{a}}$ containing all genetic (or phenotypic) correlations for *p* phenotypes can be decomposed using a Cholesky factorization such that

$$\boldsymbol{R}_{\boldsymbol{a}} = \boldsymbol{L}_{\boldsymbol{R}}\boldsymbol{L}_{\boldsymbol{R}}^{\top} \tag{S1}$$

where $\boldsymbol{L}_{\boldsymbol{R}}$ is a lower-triangular matrix with unit length rows and positive diagonal elements. These assumptions reduce the number of free parameters necessary for calculating $\boldsymbol{R}_{\boldsymbol{a}}$, as the diagonal elements of $\boldsymbol{L}_{\boldsymbol{R}}$ are determined by the off-diagonal elements of each row. Therefore, estimating $\boldsymbol{L}_{\boldsymbol{R}}$ and subsequently deriving $\boldsymbol{R}_{\boldsymbol{a}}$ using 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 $\boldsymbol{L}_{\boldsymbol{R}}$ 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 $\boldsymbol{L}_{\boldsymbol{R}}$, 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

$$\omega_u = \begin{cases} L_{R[r,c]}, & \text{if } c = 1 < r \\ L_{R[r,c]} \,/\, \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \tag{S2}$$

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

$$L_{R[r,c]} = \begin{cases} \omega_u, & \text{if } c = 1 < r \\ \omega_u * \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \quad \text{(S3)}$$

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 in the main text (Eq. 2), environmental effects can then be specified and estimated more efficiently as predictors of the transformed canonical partial correlations

$$\begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})_1}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})\frac{p(p-1)}{2}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\omega}_1} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\omega}_{\frac{p(p-1)}{2}}} \end{bmatrix} \quad \text{(S4)}$$

Applying the inverse link function $\tanh()$ and using Eq. S3 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 emphasize that the proposed implementation in Stan (Eq. S1-4) 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. 2), the context-specific (co)variance matrices $\boldsymbol{G}_{(\boldsymbol{X})}$ derived from correlation matrices $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ (Eq. 1.3) will necessarily be positive definite.

Covarying environmental predictors can also reduce the efficiency and accuracy of CRN parameter estimation. To reduce the effects of collinearity, the CRN fixed effects $\boldsymbol{\beta}_{\sigma^2}$ 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 $\boldsymbol{X}$ matrix (Harville, 1997). This involves decomposing the predictor matrix $\boldsymbol{X} = \boldsymbol{Q}^*\boldsymbol{R}^*$ into an orthogonal matrix $\boldsymbol{Q}^* = \boldsymbol{Q}\sqrt{n-1}$ and upper-triangle matrix $\boldsymbol{R}^* = \frac{\boldsymbol{R}}{\sqrt{n-1}}$, 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 $\boldsymbol{L_{R(X)}}$ can be further 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 $\boldsymbol{Z_G}$ of standardized individual-level additive genetic deviations (i.e. z-scores of breeding values), a lower-triangular Cholesky decomposition $\boldsymbol{L_A}$ of the relatedness matrix, and a diagonal matrix $\boldsymbol{S_{a(X_n)}} = \text{diag}\left(\left[\sigma_{a(X_n)_1}, \ldots, \sigma_{a(X_n)_p}\right]\right)$ of context-specific genetic standard deviations, an n x p matrix of context-specific genetic values for each phenotype can be predicted by

$$\left[\boldsymbol{a}_{(X_n)_1}, \ldots, \boldsymbol{a}_{(X_n)_p}\right] = \mathbf{L_A Z_G}\left(\mathbf{S_{a(X_n)} L_{R(X_n)}}\right)^{\top} \sim \text{Matrix Normal}\left(\mathbf{0}_{\text{nxp}}, \boldsymbol{A}, \boldsymbol{G}_{(X_n)}\right) \quad \text{(S5)}$$

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

R functions are provided to facilitate these computational gains while also generating more intuitive model estimates and predictions with respect to the standard parameterizations presented in the main text. See the `CRN functions.R` file in the corresponding Github repository for further details.

Simulation-based calibration (model validation)

To provide a general validation of the proposed model, I conducted a simulation-based calibration (SBC) procedure to assess whether the CRN (**Eq. 2**) is an unbiased Bayesian estimator. SBC is a gold-standard procedure for assessing the performance of a Bayesian algorithm across a broad range of effect sizes, using synthetic datasets generated from simulated distributions of parameter values (Talts et al., 2018). This approach removes the need for arbitrarily picking a limited, discrete range of effect sizes for assessing the validity of inferences and, in so doing, reduces the risk of missing unexpected sources of bias in unexplored regions of parameter space. Instead, during SBC, generative distributions of parameter values are simulated, covering both the range of small to moderate effect sizes typically considered in standard biological applications, as well as extremely small or large values that are likely to be rare but in principle possible to observe in practice. The proposed model is then applied to the synthetic datasets generated from these simulated distributions of parameter values (priors), in turn producing distributions of estimated parameter values (posteriors). The formal correspondence between the simulated distributions of expected parameter values and the inferred distributions of estimated parameter values can then be assessed. In particular, if the proposed model implementation facilitates valid and unbiased Bayesian inference, such that the generative values are not systematically over or underestimated compared to the estimated values, the ranks of prior versus posterior values will be uniformly distributed (Talts et al., 2018). A null hypothesis of uniformity can be assessed by visualizing the difference in empirical cumulative density functions (ECDFs) with respect to the distribution of fractional ranks among the generative versus estimated values (Säilynoja, Bürkner, & Vehtari, 2022). If the difference falls within the ECDF difference 95% probability interval of the null hypothesis of uniformity, results support a valid model that generates unbiased inference across the range of generated effect sizes. See **Fig. S1a** for a visualization of the SBC procedure.

200 datasets were simulated for SBC under modest sampling conditions of 300 individuals with a single measurement for each of 3 traits. Measurements were taken in 10 environmental contexts capturing the main effects and interaction between two continuous covariates (e.g. monthly temperatures, ages,

plot densities). This design was used to validate inference from the CRN in cases with minimal information available for estimation of a complex function (no repeated measures, relatively low sample size, very few environmental contexts), which is common in field applications of the animal model. Parameter values were generated using distributions appropriate for typical effect sizes in biological applications (Lemoine, 2019; McElreath, 2020), such that $\boldsymbol{\beta} \sim N(0,1)$ for RN fixed effects determining phenotypic means and genetic (co)variances, and $\boldsymbol{R}_{\epsilon} \sim \mathrm{LKJ}(10)$ and $\boldsymbol{\sigma}_{\epsilon} \sim \mathrm{exponential}(2)$ for residual correlation matrices and standard deviations respectively. Following previous work (Thomson et al., 2018), 10-generation pedigrees were simulated using variable degrees of extra-pair paternity (15-25%) and successful offspring recruitment into the breeding pool (40-70%), generating relatively sparse $\boldsymbol{A}$ matrices comparable to those typically observed in natural populations. Posterior distributions for each dataset were estimated using 2000 MCMC samples across 4 chains with 500 MCMC samples each for warmup. See the 'SBC_CRN.R' file in Github repository for full details on the simulation and model structure.

Results from the SBC analysis are visualized in **Fig. S1b**. The formal analysis using fractional ranks shows that with a 0.95+ probability, posterior inferences were not systematically upwardly or downwardly biased from the true values used to generate the data, indicating that the proposed Bayesian estimator facilitates valid inference of CRNs even under conditions of modest sampling effort and a broad range of effect sizes.

Model comparison simulation

For interpretability, the environmental variable was conceptualized as a measure of average temperature during the summer and winter months individuals were measured across 5 years, resulting in 5 * 2 = 10 environmental contexts. This relatively small number of distinct environmental contexts ensured results relevant for typical field studies lacking deep time series. Seasonal temperatures of arbitrary scale were generated for year t from a stationary autoregressive moving-average function of the form

$$x_t^* = \mu + 0.9\Delta x_{t-1} - 0.5\Delta x_{t-2} + 0.9\Delta\epsilon_{t-1} \quad \text{(S6. 1)}$$

$$\boldsymbol{\epsilon} \sim N(0,0.1)$$

$$\mu = \begin{cases} 1, & summer \\ 0, & winter \end{cases}$$

A small cumulative value was then added to simulate weak directional change (i.e. climate warming).

$$x_t = x_t^* + 0.1t \quad \text{(S6. 2)}$$

For each simulated dataset, this produced an autocorrelated time series of seasonal temperatures with a small upward directional trend (Fig. 3a), reflecting a combination of stochastic temperature fluctuations and weak climate warming across years. The temperature variable $\boldsymbol{x}$ was subsequently standardized to unit variance to charitably compare the performance of the CRN and character state models with binned environmental states (summer = $x_t > 0$, winter = $x_t < 0$), given that $\bar{x}_{\text{summer}} - \bar{x}_{\text{winter}} = 1$. This allowed for more direct comparison of the expected change in genetic (co)variance under a 1-unit change between the continuous temperature variable and binned seasonal state.

Genetic covariances were simulated from a CRN model for two Gaussian phenotypes such that

$$\begin{bmatrix} \boldsymbol{z_1} \\ \boldsymbol{z_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} + \boldsymbol{\alpha}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \boldsymbol{0} + \boldsymbol{\alpha}_{(X)_2} + \boldsymbol{\epsilon_2} \end{bmatrix} \quad \text{(S7)}$$

$$\begin{bmatrix} \boldsymbol{\alpha}_{(X)_1} \\ \boldsymbol{\alpha}_{(X)_2} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(X)} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(X_n)}: \begin{bmatrix} \sigma^2_{a(X_n)_1} & r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p} \\ & \sigma^2_{a(X_n)_2} \end{bmatrix}$$

small effect size: $\begin{bmatrix}\log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{X})_1}\right)\\ \log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{X})_2}\right)\end{bmatrix} = \begin{bmatrix}[\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.3\end{bmatrix}\\ [\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.3\end{bmatrix}\end{bmatrix}; \quad \begin{bmatrix}\operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{X})_{1,2}}\right)\\ \operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{X})_{p-1,p}}\right)\end{bmatrix} = \begin{bmatrix}[\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}0\\ 0.1\end{bmatrix}\\ [\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}0\\ 0.1\end{bmatrix}\end{bmatrix}$

medium effect size: $\begin{bmatrix}\log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{X})_1}\right)\\ \log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{X})_2}\right)\end{bmatrix} = \begin{bmatrix}[\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.6\end{bmatrix}\\ [\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.6\end{bmatrix}\end{bmatrix}; \quad \begin{bmatrix}\operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{X})_{1,2}}\right)\\ \operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{X})_{p-1,p}}\right)\end{bmatrix} = \begin{bmatrix}[\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}0\\ 0.2\end{bmatrix}\\ [\mathbf{1} \quad \boldsymbol{x}]\begin{bmatrix}0\\ 0.2\end{bmatrix}\end{bmatrix}$

$$\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma});\ \boldsymbol{\Sigma} = \boldsymbol{SRS}$$

$$\boldsymbol{R} \sim LKJ(10)$$

$$\mathbf{S} = \operatorname{diag}\left(\sqrt{0.7}, \sqrt{0.7}\right)$$

Mean trait changes with temperature were ignored for simplicity given the purposes of the simulation, though this is of course biologically unrealistic. Note that $\exp(-1.2) = \boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{x}=\mathbf{0})} \approx 0.3$ and $\tanh(0) = \boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{x}=\mathbf{0})} = 0$, so that for the average temperature (*x* = 0), traits exhibited modest heritability $h^2 = \frac{0.3}{0.3+0.7}$ and were uncorrelated. Cross-environment correlations were fixed to *r* = 0.8 by simulating 10 correlated standardized genetic values across *i* individuals for each of the 2 traits, constructing 10 $\boldsymbol{Z}_{\boldsymbol{G}}$ matrices (*i* x 2) from these correlated standard normal values, and subsequently scaling them using context-specific $\boldsymbol{G}_{(\boldsymbol{X})}$ matrices following Eq. S5. Relatedness matrices $\boldsymbol{A}$ were simulated following the same procedure used for the SBC as described above. See the 'methods comparison.R' script in the accompanying Github repository for further details.

It may be argued that this simulation unfairly privileges the CRN model over the random regression, given that the pattern of genetic change is not simulated directly from an individual-level model with corresponding intercepts and slopes. However, there are two key points to consider. Firstly, I have only assessed performance with respect to positive temperature change, i.e. $\Delta G_{1,2} = 0.04 = G_{1,2(x=1)} - G_{1,2(x=0)}$ (Fig. 3a), thus effectively ignoring the larger bias expected for negative values, where the random regression model will incorrectly infer symmetric change due to predicting quadratic change in (co)variances. Secondly, given the effect sizes simulated, linear slopes can

effectively approximate the pattern of individual trait change that would be expected, over the simulated range of the temperature variable, to generate the observed pattern of genetic change simulated from the CRN. In particular, focusing on a single trait z for the small effect size condition and ignoring non-genetic effects, assume the data were simulated by a random regression model of the form

$$z_{1i} = 1 + \boldsymbol{b}e^{0.15x} \quad (\mathbf{S8.1})$$

$$\boldsymbol{b} \sim N(0, 1\boldsymbol{A})$$

where $\boldsymbol{b}$ are additive genetic random slopes. The expected variance of the trait at a given value of *x* is

$$\sigma^2_{z_1} = \sigma^2_{\boldsymbol{b}}\exp(0.3\mathrm{x}) = \exp(0.3x) \quad (\mathbf{S8.1})$$

Consistent with the simulated variance CRN (**Eq. S7**). Plotting this function for different values of $\boldsymbol{b}$ over the typical range of standardized temperature values $[-1,1]$, as shown in **Figure S2** below**,** demonstrates that this function can be well approximated by a standard random regression model with genetically varying intercepts and linear slopes. Of course, when two models make distinct empirical assumptions about the functional form of trait change, it is difficult to use synthetic data to make impartial comparisons between them. A valuable target for future empirical research will, therefore, be to more directly quantify the curvature of GxE for trait (co)variances across many traits, environments, and clades. This will provide general heuristics for the conditions under which specific functional forms are likely to provide better approximations to the underlying reality, which can then be incorporated into specific CRN models.

Estimating cross-environment correlations

Cross-environment correlations are determined by the genetic variance in and covariance among the intercepts and slopes of individuals' mean reaction norms (Brommer, 2013; Mitchell & Houslay, 2021), providing crucial information for predicting evolutionary change when organisms exhibit heritable variation in their plasticity toward the environment (Martin et al., 2024). The basic CRN model (**Eq. 2**) does not directly estimate cross-environment correlations. However, it can be readily extended to do so with an appropriate breeding design or when repeated measures are available on the same individuals across environments.

Note that when repeated measures are available, such that the number of subjects $i < n$, the matrix of standardized individual-level additive genetic deviations $\boldsymbol{Z_G}$ can instead be modeled as an $i\ x\ p$ matrix that is constant across environmental contexts. This implies constant rank-order among subjects (cross-environmental genetic correlations $r = 1$) and thus the absence of individual-level variation in plasticity (Brommer, 2013) with respect to the traits being measured. This approach will generally be most applicable when individuals are measured repeatedly within rather than across environmental contexts, such that only population-level GxE can be estimated. Alternatively, when individuals are measured repeatedly across rather than within environmental contexts, the most flexible and straightforward approach to allow for $r < 1$ is to keep $\boldsymbol{Z_G}$ as an $n\ x\ p$ matrix just as in the CRN model without repeated measurements. This freely estimates subjects' standardized values and rank-order in each unique context. Cross-environment correlations can then be manually computed from the posterior distributions of these context-specific values, allowing for flexible estimation of any within- or among-individual correlations of interest (e.g. genetic assortment coefficients, Martin & Jaeggi, 2022), even in cases where only a subset of individuals are repeated measured across contexts. Cross-environment correlations can also be directly estimated in the model by expanding $\boldsymbol{Z_G}$ to an $n\ x\ p * c$ matrix for *c* contexts, as in a standard character state model, while still allowing for the CRN to predict genetic (co)variances among traits.

Genetic prediction

If one is interested in predicting genetic values for all individuals across all contexts, a more general form is to specify $\boldsymbol{Z_G}$ as a C-dimensional array of $i \; x \; p$ matrices, estimating individuals' expected values for all contexts, even those they weren't observed in. In other words, one can simply extend the basic model to include standardized (unscaled) random effects for all individuals in all possible contexts. This can be useful for various purposes, including estimating cross-environment correlations when some subjects are only observed in a subset of environmental contexts. However, this strategy can become cumbersome very rapidly for models with many environmental contexts and will result in much slower sampling. This motivates use of random regression techniques for high-dimensional and continuously varying environments.

Random regression CRN

Random regression and CRN models can be synthesized to determine how the cross-environment correlations among individuals' genetic values as well as the genetic (co)variances among traits are shaped by nested patterns of environmental variation. Random individual-level slopes can be introduced to the CRN 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 experimental 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 contexts. 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. The random regression CRN is then given by

$$\begin{bmatrix} g_{z_1}(\boldsymbol{z_1}) \\ \vdots \\ g_{z_p}(\boldsymbol{z_p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{W u}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{W u}_{(X)_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (S9)$$

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

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 $\boldsymbol{\beta}_{\boldsymbol{1}(X)_p}, \ldots, \boldsymbol{\beta}_{\boldsymbol{b}(X)_p}$ 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. A phenotypic version of the random regression CRN can also be implemented following Eq. 3.

By allowing genetic variation in response to the environment to also change across environments, the random regression CRN can simultaneously quantify both population- and individual-level parameters shaping environmental effects on the ***G*** matrix, providing more accurate predictions about the dual consequences of developmental and contextual plasticity across multiple scales. Pragmatically, this model will be particularly useful when individuals are only measured across a subset of relevant environmental contexts. For instance, subjects may be repeatedly measured under varying microclimatic and resource conditions within their local patch, providing information to estimate the average within-patch (co)variation of intercepts and slopes, without experiencing among-patch variation in community composition and habitat quality, which may further shape genetic (co)variation of these parameters and thus the magnitude of GxE across environments.

Link functions for trait variances

Using the $\log(\sigma^2)$ function for trait variances facilitates unbiased inference on the latent scale of the CRN model (Fig. S1), as well as the observed scale when it is characterized by exponential change (over the range of prediction) in response to the environment (Fig. 3). The log link is also a commonly used and familiar function that facilitates intuitive interpretation. However, this function may not always be the best choice when estimating a CRN. Below I consider two alternatives, one which is generally not recommended (square root) and another which is likely to be much more broadly applicable (inverse softplus).

Square root link

The Gaussian linear random regression model assumes that trait variances change exponentially with respect to the environment, which in turn implies linear change in genetic standard deviations for positive values (with negative values leading in the extreme to predictions of improper negative standard deviations; see Fig. S3 for a comparison). As discussed in the main text, this is unlikely to be a realistic assumption for many traits, given its implication of symmetric change in genetic variance across positive and negative values of the environment. However, when quadratic change is the true functional form for the variance, one can instead use a square root link function directly on the genetic variance, such that $\sqrt{\boldsymbol{\sigma^2}} = |\boldsymbol{X}|\boldsymbol{\beta_\sigma}$ where predictors are constrained to be positive $\boldsymbol{\beta_\sigma} > \mathbf{0}$ in the model likelihood to ensure identifiability. This implies that $(|\boldsymbol{X}|\boldsymbol{\beta_\sigma})^2 = \sigma^2$, so that the variance changes quadratically as a function of the positive linear environmental effects on the standard deviation. The key issue here is ensuring that $\boldsymbol{X\beta_\sigma} > 0$, so this must be accomplished by scaling both the linear predictors and the environmental values (by taking the absolute value) to be strictly positive, unless one is modeling data where there is no risk of predicting impossible negative values. To the degree that one is certain this model is appropriate, it will generally be much more straightforward and interpretable to fit a standard or CRN random regression model (Eq. S9), and thus I cannot recommend

it for general use outside of very specific cases (e.g. when repeated measures are unavailable but prior work informs a strong expectation of quadratic change).

Inverse softplus link

Another less commonly employed link function may be more useful for modelling genetic variance in the CRN model. An important limitation of the log link is that, for larger effect sizes, it can be prone to upwardly bias estimates on the original data scale, due to the asymmetric influence of estimation error. This follows from Jensen's inequality, where for some convex function f and random variable x

$$f\big(\mathrm{E}(x)\big) \leq \mathrm{E}\big(f(x)\big) \tag{S10.1}$$

which implies that

$$f(\sigma^2) < \mathrm{E}\left(\exp\big(\widehat{\sigma^2}\big)\right) \tag{S10.2}$$

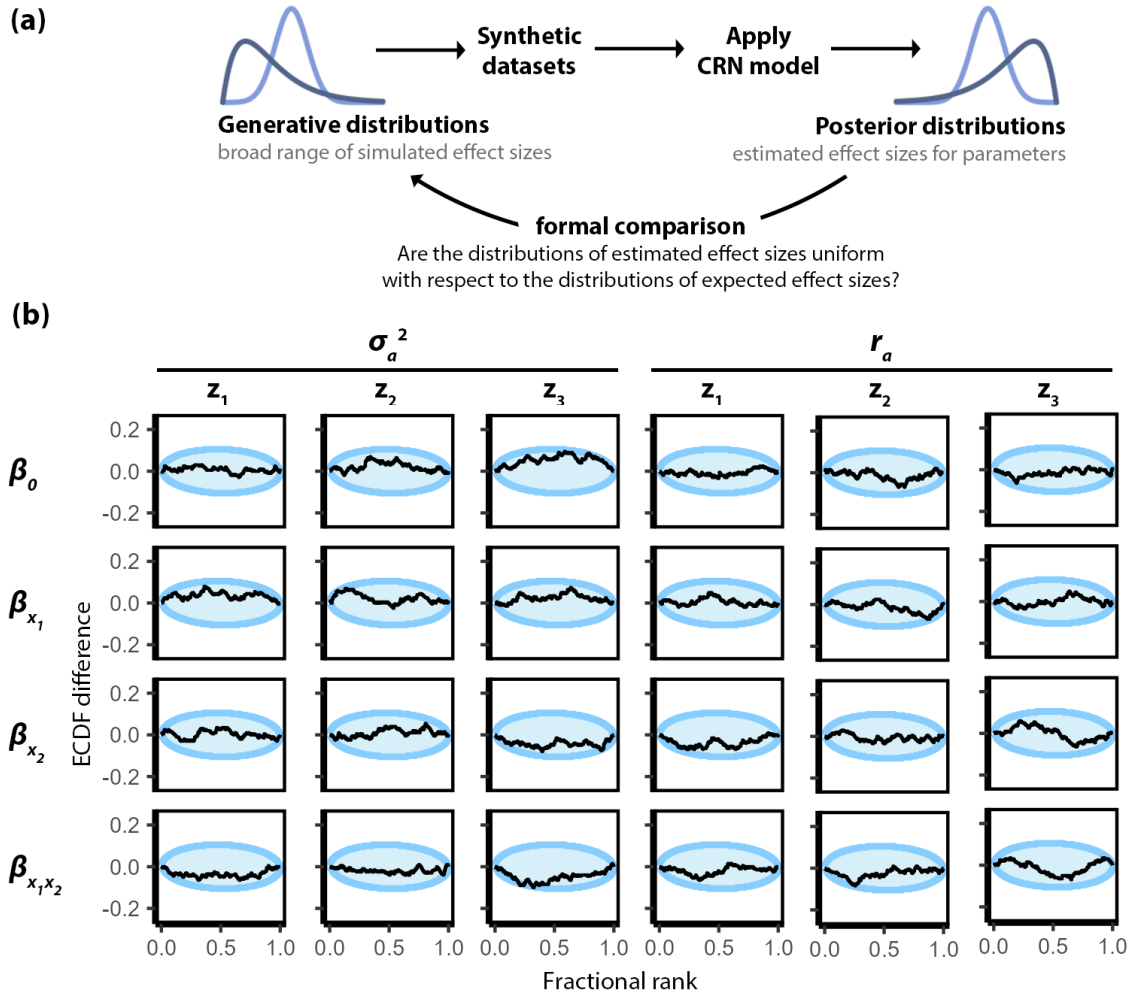
$$\widehat{\sigma^2} \sim N(\sigma^2, \delta)$$

for the true trait variance σ^2 estimated by $\widehat{\sigma^2}$ with random Gaussian error of magnitude δ. In other words, even if error is truly random and thus the expected estimate is unbiased on the log scale, the expected variance estimated on the original scale $\exp\big(\widehat{\sigma^2}\big)$ will still tend to be upwardly biased from the true value, due to application of the convex exponential inverse link function. The importance of this upward bias will be contingent on factors such as the sample size, which will generally decrease error in estimates, as well as the degree to which one is focused on unbiasedly estimating CRN model coefficients on the link scale or predicting exact magnitudes of change in genetic variance on the original data scale. The former will generally be of greater importance for basic research, where theory is most often tested by qualitative predictions (e.g. covariances will become more positive/negative) rather than exact quantitative predictions. The log link should, therefore, be a fine choice for most purposes. One can always hedge their bets by expecting that there may be a small upward bias in original scale predictions in the presence of high statistical uncertainty, while being confident that

latent scale predictions and thus the quantities most often used for hypothesis testing (e.g. CRN parameters) are expected to be unbiasedly estimated.

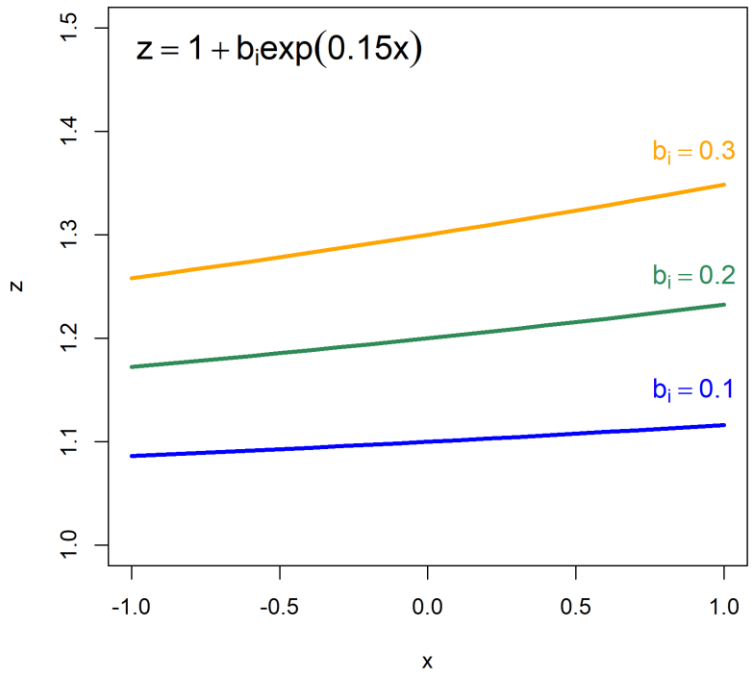
In some cases, however, it may be desirable to use the inverse softplus link function, where $\log(exp(\boldsymbol{\sigma^2}) - 1) = \boldsymbol{X\beta_{\sigma^2}}$ predicts the latent scale values and the softplus function $\boldsymbol{\sigma^2} = \log(1 + exp(\boldsymbol{X\beta_{\sigma^2}}))$ returns the variance on the original scale. Latent predictions from the softplus tend to scale much less convexly with respect to the original data scale (**Fig. S3**). The clear benefit of this link function is, therefore, that stochastic estimation error on the link scale tends to generate less upward bias with respect to the original variance and standard deviation scales, in comparison to the log link. Moreover, because the softplus approximates the identity function as it moves further away from 0 variance, it provides a natural solution for flexibly testing distinct functional forms for the variance components. For instance, one can specify a 2nd-order polynomial on the link scale to more directly test for approximately quadratic change. Note that the softplus can be implemented in Stan using the 'log1p_exp' function.

Figure S1. Simulation-based calibration of the CRN model.



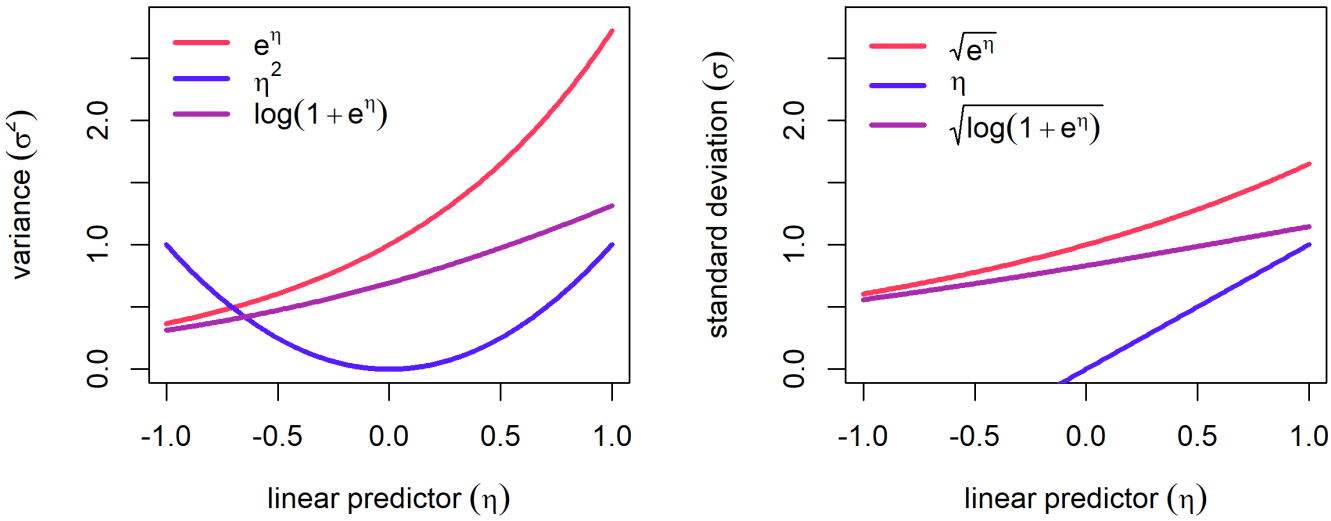
Footnote. Results are shown for SBC analysis of 200 simulated datasets of 3 traits under minimal sampling conditions (N = 300 / 10 environmental contexts) generated from prior distributions defined over the parameters of the quantitative genetic CRN model (**Eq. 2**). (a) Conceptual overview of the SBC procedure. (b) 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.

Figure S2. Individual reaction norms generating exponential change.



Footnote. Examples of individual-level responses (Eq. S6) that generate the exponential change in genetic variance assumed by the CRN model used to simulate datasets for the model comparison (Eq. S7). The true function (black text) incorporates Gaussian random slopes $\boldsymbol{b}$ determining how individuals' trait expression (z) changes in response to temperature (x). Three lines of varying color are shown for individuals (i) with differing random slope values. The visualization demonstrates that over the typical range of temperature values used in the simulation, individual reaction norms from this exponential function could be well approximated by a standard linear random regression model.

Figure S3. Comparison of link functions for the genetic variance.



Footnote. The relationship is plotted between the value of the linear predictor $\eta = \boldsymbol{X_n\beta}$ (x-axis) and the original scale variance (left plot) and standard deviation (right plot) as a function of assuming a log link (exponential for original scale; red), square root link (quadratic change; blue), or inverse softplus link (purple). As can be seen, the inverse softplus link exhibits much less convexity than the log link on the original scale, facilitating approximately sublinear prediction of the variance and standard deviation for small values (the function becomes increasingly linear with larger values).

Table S1. Summary of meerkat 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

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

References

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.

Brommer, J., Class, B., & Covarrubias-Pazaran, G. (2019). Multivariate Mixed Models in Ecology and Evolutionary biology: Inferences and implementation in R. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/hs38a

Bucklaew, A., & Dochtermann, N. A. (2021). The effects of exposure to predators on personality and plasticity. *Ethology*, *127*(2), 158–165.

Dingemanse, Niels J., Barber, I., & Dochtermann, N. A. (2020). Non-consumptive effects of predation: does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? *Ecology Letters*, *23*(1), 107–118.

Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, *64*(2), 105–123.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag.

Kraft, P. G., Wilson, R. S., Franklin, C. E., & Blows, M. W. (2006). Substantial changes in the genetic basis of tadpole morphology of Rana lessonae in the presence of predators. *Journal of Evolutionary Biology*, *19*(6), 1813–1818.

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.

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.

Martin, J. S., Westneat, D., Wilson, A. J., Dingemanse, N. J., & Araya-Ajoy, Y. (2024). Frequency-dependence favors social plasticity and facilitates socio-eco-evolutionary feedback in fluctuating environments. *EcoEvoRxiv*. https://doi.org/10.32942/X2KW5R

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

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.

Säilynoja, T., Bürkner, P.-C., & Vehtari, A. (2022). Graphical test for discrete uniformity and its applications in goodness-of-fit evaluation and multiple sample comparison. *Statistics and Computing*, *32*(2). https://doi.org/10.1007/s11222-022-10090-6

Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development predicts genotypic differences in plasticity. *Evolution; International Journal of Organic Evolution*, *72*(10), 2167–2180.

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual 2.33*.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/1804.06788

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.