Causal inference and the detection of coevolutionary contingencies using dynamic phylogenetic models

E. Ringen, J. S. Martin, S. Claessens, & A. V. Jaeggi

2024-08-23

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1 Introduction

Phylogenetic comparative methods (PCMs) are commonly used to study the coevolution of organismal traits, spanning topics such as anatomy and physiology (Dunn and Ryan 2015; Garland, Bennett, and Rezende 2005; Navalón et al. 2019; O'Connor and Cornwallis 2022; Thayer et al. 2018), life history and behavior (Bielby et al. 2007; Clayton and Cotgreave 1994; MacLean et al. 2012; Salguero-Gómez et al. 2016), and cultural evolution (Mace and Holden 2005; Navarrete et al. 2016; Watts et al. 2016). "Coevolution" in its broadest sense refers to repeatable patterns of trait covariation over time, which can be investigated using a diverse family of statistical techniques depending on the research question and type of data available (Garamszegi 2014; Harvey and Pagel 1991; Nunn 2011). Many PCMs focus on understanding how traits coevolve by estimating evolutionary correlations and rates of change across phylogenetic trees, which can be used to better understand natural history and inform ancestral state reconstructions. However, evolutionary correlation does not imply evolutionary causation between traits (Shipley 2016). Thus, PCMs facilitating inference of why traits coevolve and change together across time are indispensable tools for testing macroevolutionary and ecological theory.

Among others, phylogenetic generalized linear (mixed) models (Grafen 1989; Hadfield and Nakagawa 2010; Symonds and Blomberg 2014), phylogenetic path analysis (Gonzalez-Voyer and Hardenberg 2014; Hardenberg and Gonzalez-Voyer 2013), and Pagel's (1994) discrete method are the most popular approaches for assessing trait coevolution on macroevolutionary timescales. While each of these methods has clear benefits and performs well under specific scenarios, they are each limited in their generality by strong assumptions regarding the direction of causal effects among traits, the process of evolutionary change, and/or the statistical properties of the traits under investigation. As we detail further below, these modeling assumptions are likely to be violated in most datasets used for phylogenetic analysis, increasing the risk of inferential error and ultimately inhibiting our ability to explain the dynamics of phenotypic evolution. We therefore introduce a novel class of PCMs designed to address these challenges in a cohesive and flexible statistical framework, using Bayesian MCMC algorithms in the Stan statistical programming language (Carpenter et al. 2017).

We begin by briefly reviewing the benefits and constraints of current PCMs, particularly with regard to causal inference, exploring how their limitations motivate further methods development. We then formally

and conceptually introduce a novel class of *dynamic phylogenetic models* (DPMs). We provide a worked synthetic example of coevolution between two continuous traits – illustrated as female promiscuity and sperm length (Fitzpatrick et al. 2009) – to assess the accuracy and uncertainty of inferences made with dynamic phylogenetic models across a reasonable range of sample sizes for phylogenetic research. We also provide an accompanying coding tutorial using our `coevolve` R package to aid empiricists in easily applying basic dynamic phylogenetic models to their own datasets. We then demonstrate the generality and flexibility of our method with two empirical applications, which extend the model to complex, high-dimensional scenarios using dynamic latent variables. In particular, we show how our method can generate insights into the causal dynamics of trait coevolution across both genetic and cultural evolutionary timescales, using (1) a comparative dataset on primate brain size, sociality, diet, and life history traits (DeCasien, Williams, and Higham 2017), which reduced to two coevolving latent variables, and (2) two studies on the evolution of social complexity across a global and a regional sample of pre-industrial human societies (Ringen, Martin, and Jaeggi 2021; Sheehan et al. 2023).

2 Current approaches and motivation for a novel method

Fundamental to PCMs is the adjustment of raw trait associations for shared evolutionary history using a phylogenetic tree (or set of trees) and a statistical model. In a basic sense, phylogenetic adjustment is crucial for facilitating causal inference, as shared evolutionary history tends to generate trait correlations among closely related species with similar phenotypes, creating the illusion of convergent coevolution even when traits evolve independently. Adjustment for phylogeny reduces bias and variance due to clustering over the tree, thus reducing the risks of type I (false-positive), type II (false-negative), type M (magnitude), and type S (sign) errors during statistical inference (*CITE*). Nonetheless, adjustment for phylogeny is not a magic fix for all sources of unobserved confounding, nor does it guarantee that resulting estimates are causally interpretable (*CITE*). Therefore, while all PCMs employ statistical techniques that reduce bias, these methods vary widely in the degree to which they isolate the causal pathways by which traits coevolve over time, with most commonly used methods focusing largely on evolutionary correlation (Figure 1a) rather than causation (Figure 1b-e).

2.1 Phylogenetic generalized linear (mixed) models

Phylogenetic generalized linear (mixed) models are linear regression models used to quantify trait covariance between species or populations due to shared evolutionary history (likewise for their predecessor, independent contrasts) (Blomberg et al. 2012; Grafen 1989; Hadfield and Nakagawa 2010; Lynch 1991; Symonds and Blomberg 2014). The most commonly assumed model of trait coevolution in these regression models is Brownian motion (also called a Wiener process, *CITE*), which assumes that the trait variance for a given species or population is proportional to the total branch length of the tree leading from root to tip, and the covariance between any two species is proportional to their amount of shared history. This process is often biologically interpreted as reflecting patterns of evolutionary drift (or neutral evolution) across a phylogenetic tree. Empirically, the actual degree of covariance between related species or populations may be less than expected under a pure Brownian motion model, such as when evolution is rapid or, more mundanely, when measurement error is substantial. In these cases, the "phylogenetic signal" will be weak such that even closely related taxa may have dramatically different trait values (Blomberg, Garland Jr, and Ives 2003; Kamilar and Cooper 2013), motivating various branch length transformations to the basic model to account for this greater independence among taxa (Pagel 1999). In cases where traits do not evolve according to a modified Brownian motion process, other adjustments can also be made to the basic phylogenetic generalized linear model to modify the expected rate change in trait (co)variation. For instance, early burst models of adaptive radiation allow the rate of change to decrease over time (*CITE*), while Ornstein-Uhlenbeck models characterize evolutionary change as the product of both stochastic and deterministic processes (*CITE*).

Despite their differences, attempts to use parameters from any of these models for causal inference share often unstated and undesirable assumptions, namely that there is no selection acting on the traits included

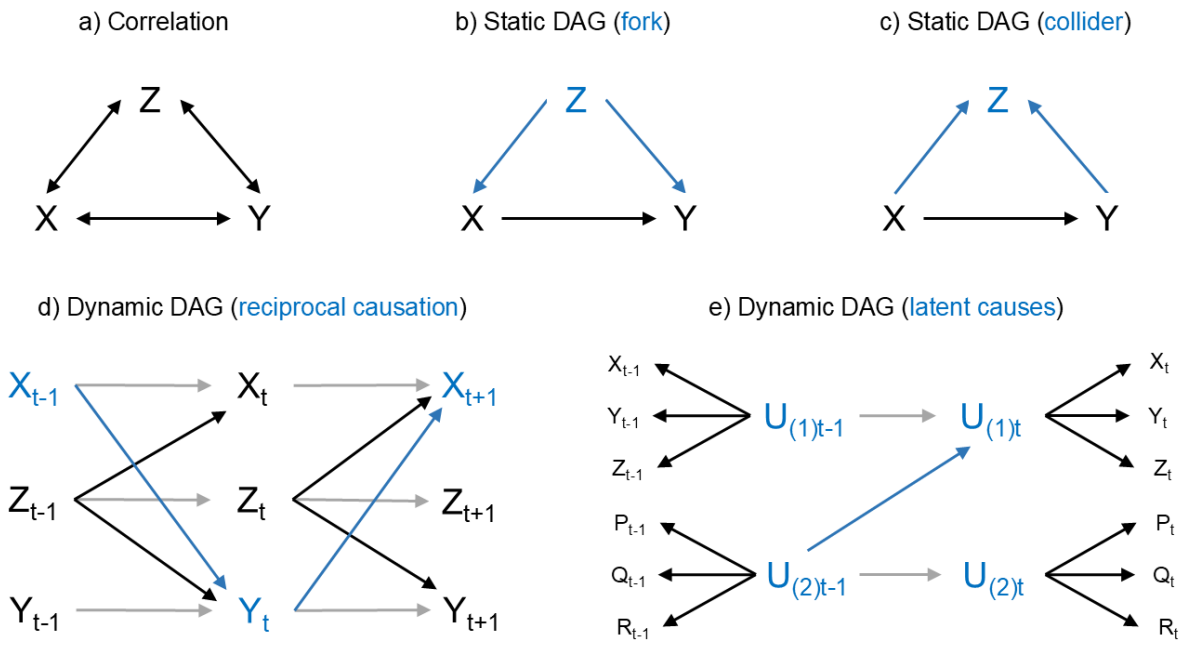


Figure 1: *Explanatory models of trait coevolution.* Examples of distinct formal approaches to describing and explaining patterns of coevolution among traits (bold letters), with important properties represented in each graph highlighted by blue arrows and text. Approaches range in complexity from (a) simple models of phylogenetic correlations (indicated by bidirectional arrows), which are useful for predictive purposes but fail to explain the causes of trait associations; to (b-e) directed acyclic graph models, which can be used to directly test the causal effects (directed arrows) driving trait associations across a phylogeny. Explicit causal models are crucial for deciding which traits should be included or excluded from a multivariate analysis to avoid potential biases due to phenomena such as so-called forks (b) and colliders (c). Causal models can also be further distinguished by whether they model relationships among traits as static (b-c) or dynamic (d-e) effects. Only dynamic models can be used to account for feedback processes (t_{-1} <U+2192> t <U+2192> t_{+1}) generated by reciprocal causation (blue arrows) among traits and autoregressive effects within traits (grey arrows) over time (d). For high dimensional problems, it is important to also consider whether inclusion of latent causes into the dynamic DAG (e), capturing dimensions of evolutionary integration among multiple traits, can provide a more parsimonious and theoretically insightful causal model (see subsection 3.6).

Box 1: Evolutionary regression coefficients underestimate causal effects

An unintuitive aspect of traditional PCMs is that regression coefficients estimated using cross-sections of contemporary species are used to infer the change in some trait Y in response to another trait X over evolutionary time. The slope of a response trait on a predictor trait is known as the "evolutionary regression coefficient" (Pagel 1993). For the evolutionary regression coefficient to be interpreted as a causal effect, we must, as with all statistical models, assume that there are no unmeasured confounders and that our parametric model is correct. But there is an additional caveat: because selection is (generally) a gradual process, the total causal effect of one trait on another – which we define as the change in the optimal trait value of Y as a function of the value of X (see Schölkopf et al. 2013 for a closely related definition of causal effects in systems of ordinary differential equations) – can take a long time to be fully realized. As such, we capture species in the process of adaptation rather than at equilibrium. Consequently, traditional methods underestimate the total causal effect of one trait another. This attenuation bias is a joint product of the strength of selection on the response trait and the rate of change in the predictor trait. Intuitively, if one trait changes too quickly, then the other will always be playing catch-up. See Figure 2 for an illustration of this process. A solution to this problem is to move beyond traditional regression-based PCMs and explicitly model trait change and the adaptive process using dynamic phylogenetic models.

as predictors, that all predictor traits are independent and direct causes of the response trait, and that there is no reciprocal causation between traits over time (see Figure 1). These are strong assumptions that are likely to be violated in most empirical datasets, limiting the applicability of phylogenetic generalized linear (mixed) models for theory testing and development. In addition, the most commonly used implementation of these models, phylogenetic generalized least squares regression, suffers from various other constraints such as being limited to traits with Gaussian errors, overfitting due to a lack of parameter regularization, and a failure to accommodate many common data features such as repeated measures of species trait values, missing data, and measurement error. While multilevel/mixed-effects models can be used to address these statistical concerns (Hadfield and Nakagawa 2010; Ives and Garland Jr 2010; Ives and Helmus 2011; Lynch 1991; Martin et al. 2020; Ringen, Duda, and Jaeggi 2019), they nonetheless share the same basic limitations with regard to causal inference and are arguably not ideal for modelling adaptation (Hansen 2014) (Box 1).

2.2 Phylogenetic path analysis

Fortunately, a major innovation in statistical research in recent decades has been to realize that even conventional regressions with observational data can be used to infer and test for the presence of causal relationships once appropriate assumptions are encoded into the model structure (Shipley 2016). This is because specific causal models, usually represented as directed acyclic graphs (DAGs, see Figure 1, Box 1), imply specific patterns of correlation (or covariance) among variables (Pearl 2009; Rubin 2005). An elaborate toolbox now exists to determine the correct test of causal hypotheses given a DAG. This has become the state-of-the-art in fields like epidemiology and is increasingly taught and used in evolutionary biology and ecology (Deffner, Rohrer, and McElreath 2022; McElreath 2020; Shipley 2016; Warrell and Gerstein 2020).

If one wants to test for a causal effect of X (say diet) on Y (say brain size), one has to carefully consider the relationship of potential third variables Z (say sociality or life history), which, depending on the structure of the DAG, should either be included in the regression analysis or not (see Figure 5). These considerations are what distinguish causal inference from standard predictive applications of multivariate regression. For instance, most researchers know that if Z causes both X and Y (a "fork", Figure 1b), Z should be included to remove spurious associations between X and Y. However, if Z is a collider (both X and Y cause Z; Figure 1c), Z should not be included in the regression model, as doing so introduces a spurious association between X and Y. For example, if diet (X) and brain size (Y) independently caused longevity (Z) but were otherwise uncorrelated, conditioning on longevity would introduce a spurious negative association between diet and brain size. This further implies that the appropriate set of covariates for testing a causal effect of

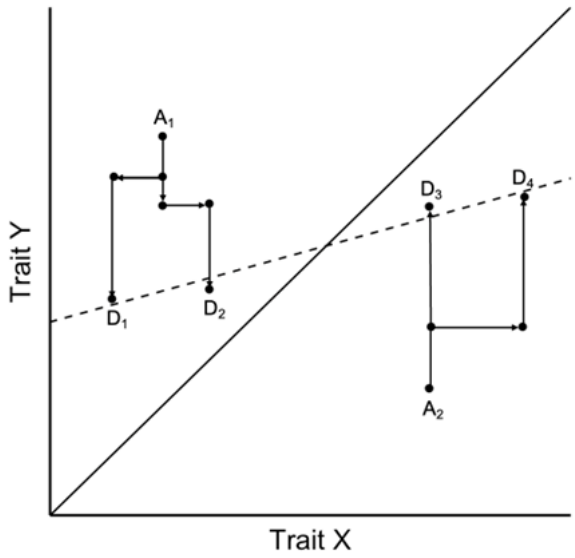


Figure 2: *Gradual adaptation flattens the evolutionary regression coefficient.* Ancestral species ("A") evolve towards their optimal trait Y value, which depends on the value of trait X shown on the solid line. Ancestors speciate and their descendants ("D") inherit their maladapted trait values, resulting in a slope between the two traits (dashed line) that is flatter than the true relationship. Redrawn from Figure 14.4 in Hansen (2014).

X on Y may not be the same as the set for testing the causal effect of Z on Y. Nevertheless, all coefficients in a multivariate regression are commonly interpreted as causal effects, a phenomenon known as the "table 2 fallacy" (Westreich and Greenland 2013). Most comparative analyses have not yet used the causal inference toolbox and typically include several variables in a phylogenetic generalized least squares model without formal causal justification. This implicitly assumes that all variables are independent causes, which is often problematic.

Phylogenetic path analysis represents a particular application of the causal inference toolbox that overcomes some of the issues outlined above for phylogenetic generalized least squares (Gonzalez-Voyer and Hardenberg 2014; Hardenberg and Gonzalez-Voyer 2013). With phylogenetic path analysis, researchers pre-define specific causal models, ideally avoiding colliders and other traps (Figure 1), and compare their fit to the data. Specifically, a DAG implies a set of variables that should be statistically independent, which is then tested using phylogenetic regression, such that several causal hypotheses can be compared. For instance, Navarrete et al. (2016) presented several phylogenetic path analysis models that all included links between life history and social group size with brain size, with variable direct or indirect effects of diet and technical intelligence. Thus, phylogenetic path analysis is well-suited for comparing existing hypotheses, with DAGs generated "top-down" from prior theory or empirical literature (Gonzalez-Voyer and Hardenberg 2014).

However, the applicability of phylogenetic path analysis is limited to static DAGs lacking any reciprocal effects over time, such as Figure 1b and c where X, Y, and Z do not have any temporal structure that could be used to resolve reciprocal causal effects. Such models assume a constant direction of effect between two variables, with bidirectional causal effects only being able to be described as unresolved patterns of correlation (Figure 1a). This is an issue because positive (or negative) evolutionary feedback loops are predicted by many theoretical models in response to life history tradeoffs and various forms of frequency- and density-dependence, suggesting that reciprocal causation is likely to be a common phenomenon explaining patterns of trait coevolution. For example, McNamara (2022) presents a model of the coevolution of paternal care and extrapair paternity, which is driven by a frequency-dependent feedback loop over evolutionary time between males' tendency to seek extrapair copulations and the benefits of their paternal care. The inability to directly test for such dynamic feedback effects is a fundamental limitation of phylogenetic path analysis

and any other phylogenetic method reliant on static causal models, thus motivating the development of dynamic methods.

2.3 Pagel's discrete dynamic method

To date, Pagel's (1994) "discrete" method is the only general application of dynamic causal inference to phylogenetic data. By reconstructing the coevolutionary sequence of two traits along the branches of a phylogeny, researchers can infer causality and potentially reciprocal causality through temporal contingency, referred to as "Granger causality" in economics (Granger 1969). For example, researchers can infer that X evolved first and then made the evolution of Y more likely. This method has resulted in many high-profile publications (Cornwallis et al. 2017; Fitzpatrick et al. 2009; Kappeler and Pozzi 2019; Sheehan et al. 2018; Shultz, Opie, and Atkinson 2011; Watts et al. 2016), all of which are built on the stronger causal inference of temporal contingency.

Unfortunately, despite its innovative approach to causal inference, Pagel's method is fundamentally limited to investigating the coevolution of binary traits, where dynamic evolutionary transitions can be conceptualized in terms of probabilities of switching between two discrete states. Application of this method to evolutionary topics involving continuous phenomena (e.g. morphology, life history, brain size, etc.) thus requires that researchers falsely dichotomize naturally occurring variation in traits (e.g. Cornwallis et al. 2017; Fitzpatrick et al. 2009; Sheehan et al. 2018; Watts et al. 2016), which is well-established as an undesirable statistical practice leading to loss of power, biased effect sizes, and inconsistent results (Dawson and Weiss 2012; Royston, Altman, and Sauerbrei 2006). For instance, Fitzpatrick et al. (2009) dichotomized continuously-measured sperm length and speed by classifying traits below the species mean as "low" and those above the species mean as "high", and they collapsed a four-point scale of female promiscuity, based on behavioral and paternity data, into "low" (levels 1 and 2) and "high" (levels 3 and 4). Using Pagel's method, they then showed that sperm evolved to be faster and longer in response to increases in female promiscuity.

The limitation of Pagel's method to binary traits is further exacerbated when studying the (co)evolution of multiple continuous traits, which often motivates the incorporation of latent variables into a dynamic causal analysis (Figure 1e). Latent variables are widely used in biology to capture theoretically pertinent constructs that are difficult to directly operationalize and quantify using a single measurement, such as size and shape dimensions in morphometrics (*CITE*), environmental quality and climate metrics in ecology (*CITE*), life history variation (Stott et al. 2024), and canonical axes of correlational selection in quantitative genetics (*CITE*). While latent variables are sometimes used pragmatically as placeholders for deeper causal analysis (Shipley 2016), many evolutionary models directly conceptualize and model causal effects occurring among latent variables. For instance, life-history traits are theorized to evolve along a latent fast-slow continuum (Bielby et al. 2007; Healy et al. 2019; Stott et al. 2024; Wright et al. 2019) that can both cause and in turn be caused by the correlated evolution of a niche complexity continuum reflecting differences in brain size and cognitive ability, social group size, and foraging behavior (Dunbar and Shultz 2017; Miller, Barton, and Nunn 2019; Navarrete et al. 2016) (see subsubsection 3.6.1). While prior research has used methods such as phylogenetic principal component analysis (pPCA; *CITE*) to apply Pagel's method to latent variables (e.g. Cornwallis et al. 2017), this approach still requires falsely dichotomizing latent variables and thus suffers from the same statistical biases as with any other continuous measure. Synthesizing continuous and multivariate methods with dynamic causal inference thus remains a major challenge for contemporary PCMs.

3 Dynamic phylogenetic models

The limitations of current PCMs can be overcome by synthesizing their respective strengths into a single generalizable framework for dynamic phylogenetic causal inference, applicable to both discrete and continuous traits. We have attempted to do so through the development of a novel class of PCMs we refer to simply as dynamic phylogenetic models (DPMs) of trait coevolution. Our DPM is driven by a multivariate stochastic differential equation, similar to a multivariate Ornstein-Uhlenbeck (OU) process (see Box 2 for an overview of the basic OU model) with greater model flexibility and complexity than is permitted by standard

Box 2: The Ornstein-Uhlenbeck (OU) model

The OU model is a mean-reverting, stationary Gauss-Markov process. It describes change in a trait due to both Gaussian noise and reversion towards some central value. In some fields, such as economics, the mean reversion component is called "drift" and the stochastic noise is called "diffusion" (i.e., the Vasicek model, *CITE*). In evolutionary biology, the mean-reverting quality is loosely interpreted as "selection" and accordingly the Gaussian noise is labeled "drift" (*CITE*). The basic OU form is

$$dy(t) = \alpha(\theta - y_t) + \sigma dW(t)$$

where α controls the strength of mean-reverting selection, θ is the mean trait value, and σ controls the strength of drift. When $\alpha = 0$, the model is pure drift Brownian motion. The simplest OU models assume a single evolutionary optimum (θ), or estimate an ancestral optimum along with a global optimum. More elaborate OU models imagine that θ changes as a function of other variables, turning it into a coevolutionary process with varying selection regimes (i.e., the Hansen model; *CITE*). These approaches exploit the fact that, if selection regimes are piecewise-constant (the optimum is the same within each segment of the phylogenetic tree but allowed to vary at branching points), the OU process can be discretized, giving a closed form solution for the expectation and covariance matrix of a trait. Butler and King (*CITE*) provided a maximum-likelihood algorithm for this approach, and a Bayesian implementation was developed by Ross et al. (*CITE*). Depending on the research question, the piecewise constant assumption can be quite restrictive and a poor approximation when tree segments are long, due to the assumption of a constant optimum within each discrete tree segment.

implementations. Our approach arose from two complementary goals: first, to extend previous Bayesian implementations of the multivariate OU model such that the optimal trait value θ is updated dynamically (rather than assuming piecewise constant, see Box 2). Second, we wanted a general form for the optimal trait value θ that could be used to assess directionality (X $\rightarrow$ Y vs. Y $\rightarrow$ X) and contingencies (X, then Y) in evolution, akin to Pagel's (1994) method but without restrictions on the type or number of traits.

3.1 Formal model

Rather than estimating a phylogenetic variance-covariance matrix as is standard for PCMs, we instead partition the deterministic and stochastic components of the coevolutionary process, adapting the continuous-time structural equation modeling approach of Driver, Oud, and Voelkle (2017) to the phylogenetic context. The evolutionary history of any species or population is modeled as a time series where the deterministic dynamics of an OU process play out over the length of each tree segment, and the stochastic drift components (which are by definition orthogonal to selection) are added to the end of each segment as independent samples from the standard normal distribution, scaled by the expected covariance for a given segment duration $t(s)$

$$d\eta(t) = (\mathbf{A}\eta(t) + \mathbf{b}) + \mathbf{G}dW(t) \tag{1}$$

Here $\eta(t)$ is a vector of the latent variables at time t, the matrix $\mathbf{A}$ represents "selection" with autoregressive terms on the diagonal equivalent to α in the univariate OU process (Box 2) and off-diagonals representing the effect of each trait on the others (e.g., $\mathbf{A}[2,1]$ represents the effect of η_1 on η_2), and $\mathbf{b}$ is a vector of continuous time intercepts that, along with $\mathbf{A}$, determine the asymptotic values of η. The matrix $\mathbf{G}$ is the Cholesky decomposition of the "drift" covariance matrix $\mathbf{Q}$, such that $\mathbf{Q} = \mathbf{G}\mathbf{G}'$, which scales the stochastic Weiner process. The square root of the diagonals in $\mathbf{Q}$ are equivalent to σ in the OU process (Box 2). Although in other types of time-series analyses it is possible to estimate the off-diagonals of $\mathbf{Q}$ (i.e., the covariance of the stochastic drift terms), it is not possible to simultaneously estimate them in the phylogenetic context while also estimating the off-diagonals of $\mathbf{A}$, so in our model we assume that they are zero. This is equivalent to the assumption made in Pagel's (1994) method that multiple traits cannot transition together instantaneously.

Following Driver, Oud, and Voelkle (2017), the solution to Equation 1 for any time interval $t - t_0$ is

$$\begin{aligned} \eta(t) &= e^{\mathbf{A}(t-t_0)}\eta(t_0)\mathbf{A}^{-1}[e^{\mathbf{A}(t-t_0)}]\mathbf{b} + \int_{t_0}^{t} e^{\mathbf{A}(t-s)}\mathbf{G}dW(s) \\ \text{cov}\left[\int_{t_0}^{t} e^{\mathbf{A}(t-s)}\mathbf{G}dW(s)\right] &= \text{irow}\left(\mathbf{A}_{\#}^{-1}\left[e^{\mathbf{A}_{\#}(t-t_0)} - \mathbf{I}\right]\text{row}(\mathbf{Q})\right) \end{aligned} \tag{2}$$

where $\mathbf{A}_{\#} = (\mathbf{A} \otimes \mathbf{I}) + (\mathbf{I} \otimes \mathbf{A})$, with $\otimes$ denoting the Kronecker product. $\mathbf{I}$ is an identity matrix, row() is an operation that takes elements of a matrix row-wise and puts them in a column vector, and irow() is the inverse of the row operation (Driver, Oud, and Voelkle 2017). In the OU process, $\frac{dy}{dt} = 0$ when $y = \theta$ (Box 2). Equivalently, we can calculate the equilibrium value θ for each trait η_i in the DPM as

$$\theta_{\eta_i} = \frac{-(\sum_{j \neq i} \mathbf{A}[i,j]\eta_j) + \mathbf{b}_i}{\mathbf{A}[i,i]} \tag{3}$$

Note that the vector θ_η does not equal the time asymptotic trait values for the system as a whole ($\mathbf{b}_{\Delta\infty}$; *CITE*), which are determined by

$$\mathbf{b}_{\Delta\infty} = \mathbf{A}^{-1}\mathbf{b} \tag{4}$$

3.2 Implementation

Our strategy for mapping this model onto a phylogenetic tree is described in algorithm 1 and visualized in Figure 3. In short, we iterate over segments of the tree and allow the coevolutionary process to play out over the length of each tree segment.

Algorithm 1: Implementation of the dynamic phylogenetic model

for $n \in N$ **do**
- Divide the evolutionary history of n into S segments, where each segment starts with a parent node and ends with a child node or tip.
- Calculate the length of each segment $s \in S$, i.e., duration of time between the parent and child nodes.
- Initialize the ancestral trait values η_0.
- **for** $s \in S$ **do**
  - Solve for $\eta(s)$:

$$\eta(s) = e^{\mathbf{A}(t(s))}\eta(s-1)\mathbf{A}^{-1}[e^{\mathbf{A}(t(s)-t(s-1))}]\mathbf{b} + \int_{t(s-1)}^{t(s)} e^{\mathbf{A}(t(s-1)-t(s))}\mathbf{G}dW(t(s))$$

- **end for**

end for

We have implemented this algorithm in a Bayesian inferential framework using the Stan programming language (Carpenter et al. 2017) (see Supplementary Materials for example Stan code). Stan employs the computationally efficient Hamiltonian Monte Carlo MCMC algorithm to sample from the posterior distribution. Implementing the model in Stan also allows for a greater degree of model flexibility. Users can fit the model to any number of traits following any response distribution, lifting the limitation of Pagel's (1999) method to binary traits and eschewing the need for false dichotimization. Stan also allows for additional features, such as the incorporation of regularization through prior distributions, the inclusion of measurement error on traits, and handling missing data.

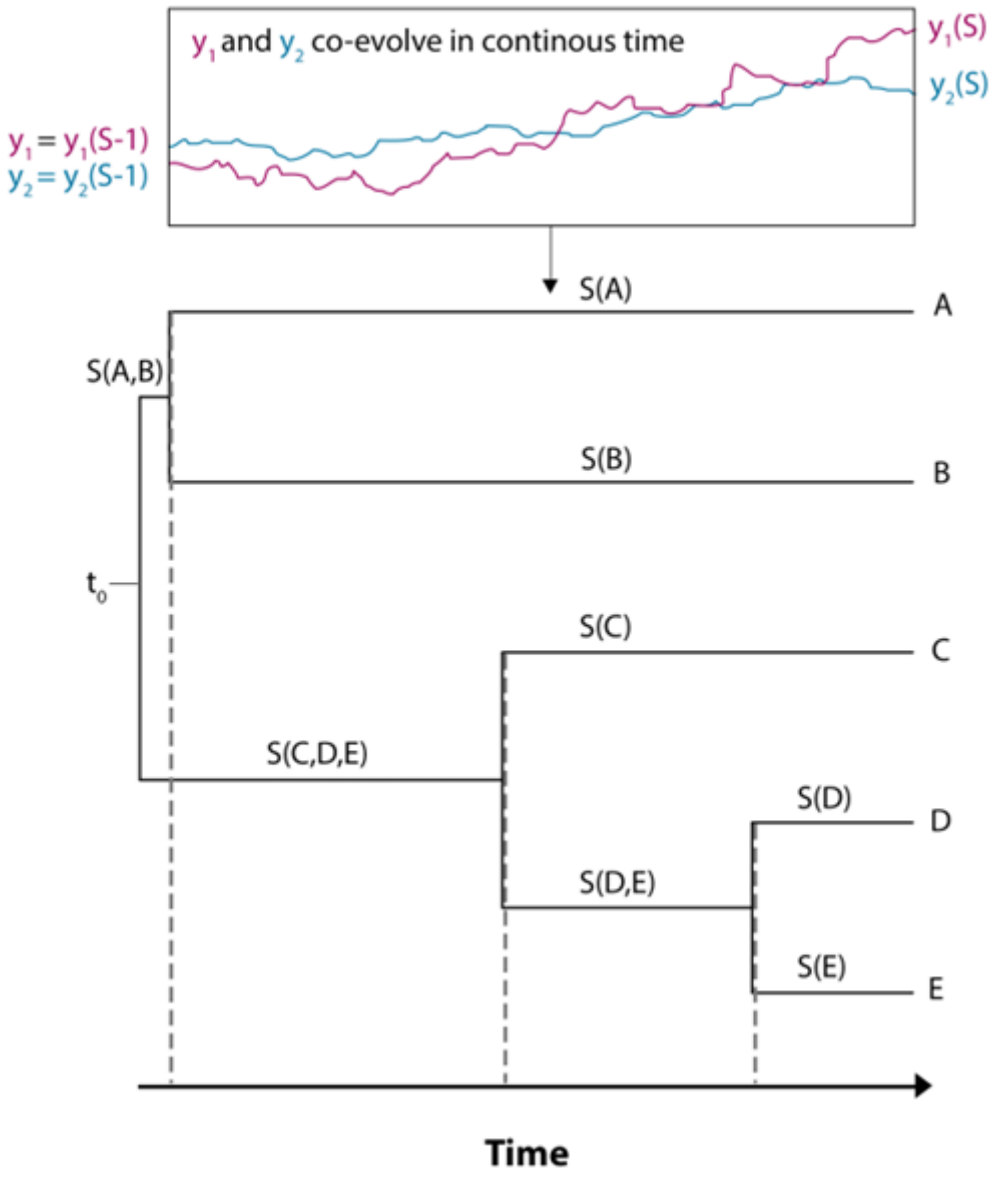


Figure 3: *The DPM algorithm.* Example of how our phylogenetic algorithm works (see algorithm 1 for formal structure). The evolutionary history of each tip (A-E) is described as a time series starting at $t(s)$ and comprised of a set of segments S. The OU process runs over each segment in order, where the initial trait values for any segment s are the terminal values from $s-1$. Each sequence is a combination of shared segments (where tips are evolving together) and unshared segments (where tips are evolving independently). For example, S(A) and S(B) start with the same trait values because they both descend directly from the parent segment S(A,B). But further evolution along S(A) and S(B) happens independently. Dashed lines indicate segment split points.

3.3 The `coevolve` R package

To aid researchers in applying the DPM to their own data, we have developed the `coevolve` R package (https://github.com/ScottClaessens/coevolve). This fully documented package provides a user-friendly interface for applying DPMs, with simple functions for generating DPM Stan code and data lists, fitting DPMs to data using the `cmdstanr` package (Gabry et al. 2024), and post-processing and plotting model results.

The development version of the `coevolve` package can be installed and loaded using the following R code:

```
#install.packages("devtools")
library(devtools)
install_github("ScottClaessens/coevolve")
library(coevolve)
```

Currently, the `coevolve` package supports binary, ordinal, count, continuous, and positive real traits. The package can also handle missing trait data, repeated observations per taxa, and model comparison using approximate leave-one-out cross-validation. We plan to implement further response distributions and latent variables in future versions of the package, but for now users are encouraged manually adapt the Stan code produced by the package to add these features.

3.4 Synthetic example

In order to demonstrate that our formal model is able to accurately recover true coevolutionary dynamics, we apply the `coevolve` package to a simple synthetic example.

Consider an ancestral species with two continuous traits. As an illustration, let's assume that we are studying the highly diverse cichlid fish of Lake Tanganyika, and the two continuous traits are female promiscuity and sperm size (Figure 4a). Our hypothesis is that changes in female promiscuity will select for changes in sperm size, but not vice versa (Fitzpatrick et al. 2009). We simulate the coevolution of these traits over time, allowing traits to have autoregressive effects (i.e., change in a trait influences itself in the future), coevolutionary effects (i.e., change in a trait influences the other trait in the future), and stochastic drift. In this particular simulation, we specify that changes in female promiscuity influence future changes in sperm size, but not vice versa.

The data generating model is as follows:

$$\begin{aligned} \text{spermsize}_t &= (0.90 \times \text{spermsize}_{t-1}) + (0.85 \times \text{promiscuity}_{t-1}) + \epsilon_{\text{spermsize}} \\ \text{promiscuity}_t &= (0.90 \times \text{promiscuity}_{t-1}) + \epsilon_{\text{promiscuity}} \\ \epsilon_{\text{spermsize}}, \epsilon_{\text{promiscuity}} &\sim \mathcal{N}(0, 0.05) \end{aligned} \tag{5}$$

We allow this coevolutionary process to play out over discrete timesteps. In each timestep, there is a small probability ($p = 0.05$) of a speciation event, where the parent species splits into two child species. Coevolutionary dynamics then continue independently on each separate phylogenetic branch. We continue this process of evolution and speciation until we have a sufficient sample size of cichlid species.

We then use the `coev_fit()` function from the `coevolve` package to fit our statistical model using the resulting phylogenetic tree and the "contemporary" values of female promiscuity and sperm size from the final timestep. The code is as follows:

```
fit <-
  coev_fit(
    data = sim$data,
    variables = list(
      Promiscuity = "normal",
```

```
      SpermSize = "normal"
    ),
    id = "species",
    tree = sim$tree
  )
```

This function takes a data frame of trait values, a list of variables to coevolve in the model (along with their associated response distributions), the column in the dataset that links to the phylogeny, and a phylogeny object of class `phylo`. The function generates the code and data list for Stan, compiles the model, and fits the model.

We find that, when fitted to data from 100 cichlid species, the statistical model converges normally and is able to accurately recover the true coevolutionary dynamics. The model summary is as follows:

```
summary(fit)
```

```
## Variables: Promiscuity = normal
##            SpermSize = normal
##      Data: sim$data (Number of observations: 100)
##     Draws: 4 chains, each with iter = 2000; warmup = 2000; thin = 1
##            total post-warmup draws = 8000
##
## Autoregressive selection effects:
##             Estimate Est.Error  2.5% 97.5% Rhat Bulk_ESS Tail_ESS
## Promiscuity    -1.18      0.64 -2.55 -0.10 1.00     3163     2509
## SpermSize      -0.53      0.41 -1.52 -0.02 1.00     4468     3729
##
## Cross selection effects:
##                                Estimate Est.Error  2.5% 97.5% Rhat Bulk_ESS
## Promiscuity <U+27F6> SpermSize     2.02      0.91  0.18  3.75 1.00     2292
## SpermSize <U+27F6> Promiscuity     0.76      0.75 -0.69  2.25 1.00     4056
##                                Tail_ESS
## Promiscuity <U+27F6> SpermSize     3603
## SpermSize <U+27F6> Promiscuity     5704
##
## Drift scale parameters:
##             Estimate Est.Error 2.5% 97.5% Rhat Bulk_ESS Tail_ESS
## Promiscuity     0.88      0.07 0.75  1.03 1.00     8528     6276
## SpermSize       0.69      0.06 0.57  0.83 1.00     4179     5830
##
## Continuous time intercept parameters:
##             Estimate Est.Error  2.5% 97.5% Rhat Bulk_ESS Tail_ESS
## Promiscuity    -0.03      0.55 -1.12  1.06 1.00     4645     5166
## SpermSize      -0.04      0.73 -1.47  1.43 1.00     5206     5459
```

By plugging the fitted model parameters into Equation 3 above, we are able to calculate $\Delta\theta_z$, which is defined as the change in the equilibrium value of each trait resulting from an absolute deviation increase in the other trait. Figure 4b visualizes the resulting posterior distributions of $\Delta\theta_z$ for both directions of selection, which we can retrieve using the function `coev_calculate_delta_theta()`. For example, we might run:

```
coev_calculate_delta_theta(
  object = fit,
  response = "Promiscuity",
```

```
  predictor = "SpermSize"
)
```

We can see from Figure 4b that the posterior distribution of $\Delta\theta_z$ for promiscuity → sperm size is positive (median posterior value = 2.96, 95% credible interval [0.31 50.97]) suggesting that an absolute deviation increase in female promiscuity results in an increase in the equilibrium value of sperm size. However, the posterior distribution of $\Delta\theta_z$ for sperm size → promiscuity includes zero (median posterior value = 0.92, 95% CI [-2.17 6.1]) suggesting that an absolute deviation increase in sperm size does not change the equilibrium value of female promiscuity. This result accurately reflects the data generating process.

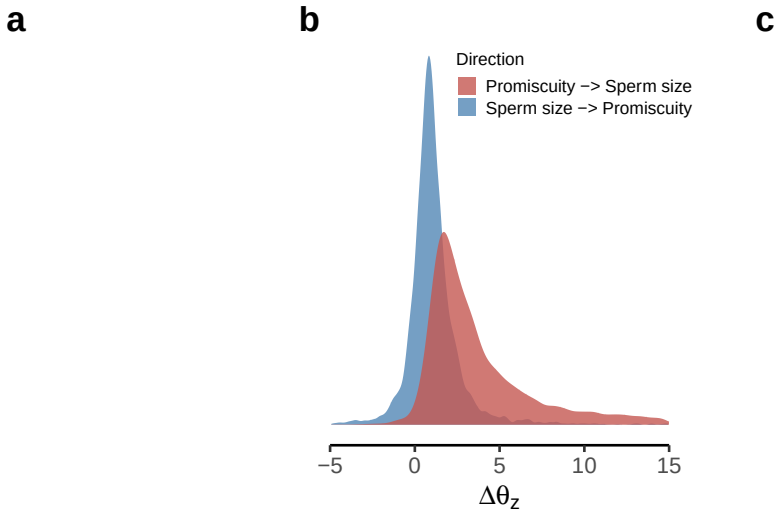


Figure 4: *Synthetic data example.* (a) We use the example of the coevolution of female promiscuity and sperm size in cichlid fish as an illustration. Using Pagel's discrete method, Fitzpatrick et al. (2009) dichotomized their initially continuous measures of female promiscuity and sperm size and swimming speed, finding that changes in promiscuity preceded changes in sperm traits. (b) Our synthetic data analysis shows that the dynamic phylogenetic model correctly identifies a positive effect of female promiscuity on sperm size, but not of sperm size on promiscuity. (c) The flow field shows how continuous trait values of female promiscuity and sperm size are expected to change over evolutionary time as a function of the state of the other variable. Black and white fish represent males and females, respectively.

We can further visualize the influence of female promiscuity on sperm size by plotting a phase plane of the evolutionary dynamics implied by the model. Figure 4c depicts the expected change in female promiscuity and sperm size depending on their current states. Both variables have been standardized as z-scores for easier interpretation. The solid lines in the figure represent nullclines where female promiscuity (blue) and sperm size (red) are at equilibrium, depending on the state of the other. This plot is easily generated using the following code:

```
coev_plot_flowfield(
  object = fit,
  var1 = "Promiscuity",
  var2 = "SpermSize",
  nullclines = TRUE
)
```

The phase plane shows that changes in female promiscuity have large effects on the direction of selection on sperm size. When female promiscuity is low (i.e., mating is predominantly monogamous), investment

in large sperm is selected against, but when female promiscuity is high, large sperm are favored. However, changes in sperm size do not have the same effect on the direction of selection on female promiscuity or the equilibrium value of female promiscuity. That is, changes in sperm size do not affect the evolution of mating systems in our example.

To determine whether this pattern of results generalizes beyond a single simulation run, we iterated the above simulation 100 times for three different sample sizes: 20, 50, and 100 species. Table @ref{tab:power} shows the proportion of models where the posterior $\Delta\theta_z$ values have 95% credible intervals excluding zero. The results show that the model is relatively underpowered to detect the causal effect of promiscuity on sperm size when there are only 20 species, but it is able to detect this causal effect with sufficient power when the sample size is 50 or 100.

Table 1: The proportion of models in which the 95% credible interval for $\Delta\theta_z$ is greater than zero for three different sample sizes. 100 simulations were run for each sample size. NOTE THIS IS NOT FULL SIMULATION

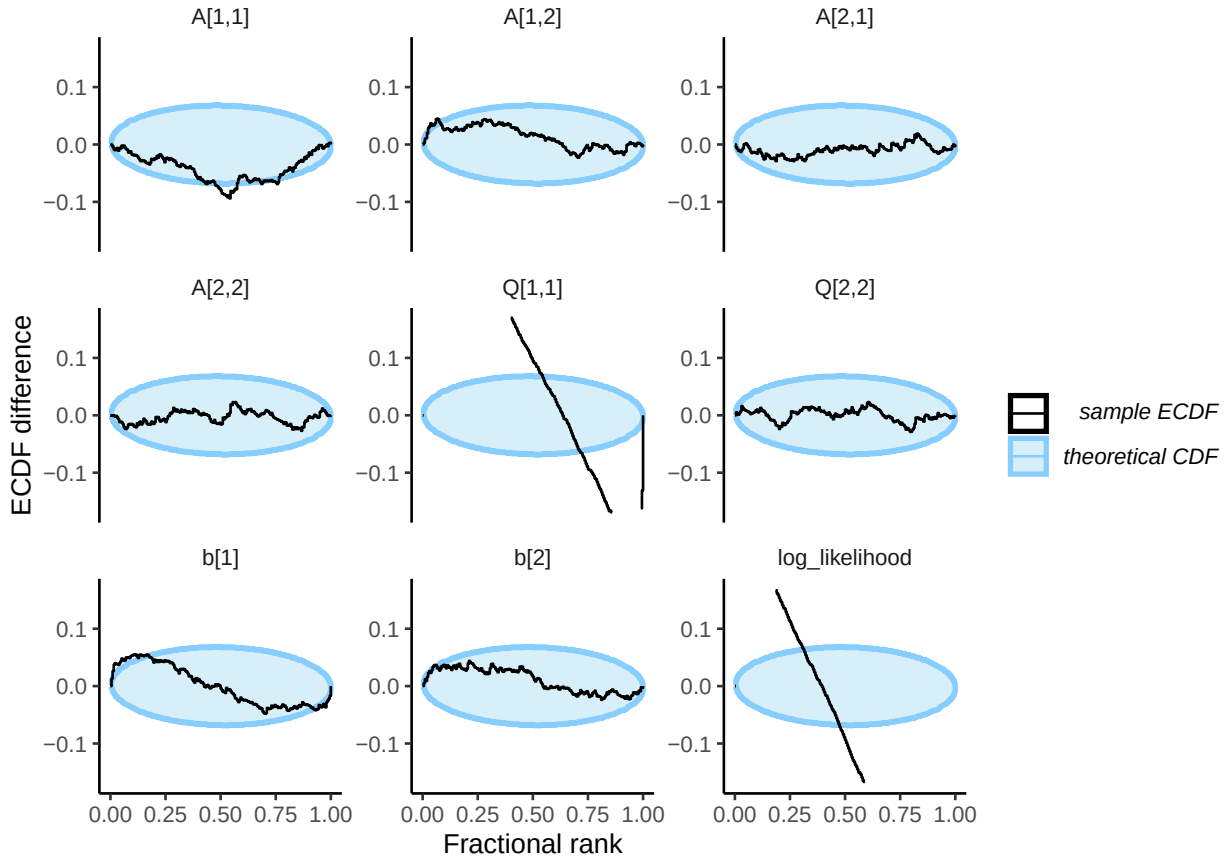
Direction	n = 20	n = 50	n = 100
Promiscuity -> Sperm size	0.05	0.17	0.61
Sperm size -> Promiscuity	0.01	0.03	0.10

3.5 Simulation-based calibration

In order to further validate our Bayesian model, we ran a simulation-based calibration exercise. Simulation-based calibration is a method to assess whether samples of the posterior distribution from a generator (in this case, our Stan model) faithfully capture the "true" posterior distribution (Modrák et al. 2023; Talts et al. 2018). We infer this faithfulness by simulating parameters from the prior distribution, using the simulated parameters to simulate multiple datasets, and then fitting the model to those datasets. In a well-calibrated Bayesian model, the resulting samples of the "data-averaged posterior" should be indistinguishable from the prior distribution. If there is a discrepancy between the two, this suggests that the model is either misspecified or otherwise inaccurately computing the posterior distribution.

We used the `SBC` R package (Modrák et al. 2023) to run our simulation-based calibration exercise. We generated 500 simulated datasets from the prior ($N = 20$, one continuous trait and one binary trait) and fitted the DPM to these simulated datasets. We then determined whether the posterior ranks from the fitted models were uniformly distributed by plotting, for several key model parameters and the log-likelihood, the differences between the empirical cumulative distribution function and a perfectly uniform cumulative distribution function. If posterior ranks are perfectly uniformly distributed as expected, this should result in a flat line, though some variation from flat is to be anticipated. Ideally, we would expect to see the lines falling within the expected deviations at the 95% level (blue ovals). Indeed, this is what we find for our model (subsection 3.5), suggesting that the Stan program is well-calibrated.

\begin{figure}



{

}

\caption{*Simulation-based calibration.* Differences between the empirical cumulative distribution function (ECDF) and a perfectly uniform cumulative distribution function. Blue ovals represent theoretical deviations for these values at the 95% level.} \end{figure}

3.6 Empirical applications

To illustrate the flexibility of our method, we provide empirical applications that involve several variables ultimately best analyzed using latent variables. In subsubsection 3.6.1, we present a novel analysis of existing data on primate brain size evolution, while subsubsection 3.6.2 summarizes two published examples of cultural evolution in humans.

3.6.1 Brain size evolution in primates

Brain size, both in absolute terms and relative to body size, varies tremendously among taxa, with primates as a group having larger brains than other mammals, and humans having the largest brains of all primates (Boddy et al. 2012; Isler et al. 2008; Miller, Barton, and Nunn 2019; Smaers et al. 2021). This has motivated a large number of comparative phylogenetic analyses on correlates of brain size, typically framed as testing social or ecological benefits or constraints for evolving larger brains (DeCasien, Williams, and Higham 2017; Dunbar and Shultz 2017; Isler and Schaik 2009). However, many have recognized that this approach has reached an impasse, as different hypotheses can be supported depending on the specific dataset (Powell, Isler, and Barton 2017; Wartel, Lindenfors, and Lind 2019) due to variable patterns of missingness and high collinearity among predictors (e.g. diet, social group size, and life history), and

reciprocal causation is inherent in most theoretical models (e.g. larger brains allow access to better diet which removes energetic constraints on larger brains) (Isler and Van Schaik 2014). Add the fact that Pagel's discrete method has to our knowledge never been applied to brain size due to its continuous nature, and this field is virtually predisposed to the use of DPMs. The issues and innovations highlighted here should also generalize to many other morphological and socio-ecological traits.

To illustrate the potential of DPMs to analyze brain size evolution, we used the dataset of DeCasien, Williams, and Higham (2017), augmented by life-history variables from Herculano-Houzel (2019), and a consensus phylogeny from 10ktrees (Arnold, Matthews, and Nunn 2010). We contrasted a typical phylogenetic multiple regression (Fig. 5a), in which all predictors are interpreted as independent causes of brain size, with a latent factor model (Fig. 5b), which was derived in a bottom-up way using exploratory graph analysis (see Supplementary Materials). Note that instead of modeling absolute brain size and adjusting for body size as is typical for phylogenetic regression (Fig. 5a), the model in Fig. 5b directly modeled relative brain size as the allometric slope for the brain-body relationship; we also accounted for variable allometric relationships using random intercepts and slopes at the species level, thus allowing species to differ in their brain-body allometry (Smaers et al. 2021). Note also that this DPM used Bayesian imputation rather than removing species with missing data and losing valuable information and statistical power. In contrast to the approach in Fig. 5a, which compares independent causes and might conclude that "primate brain size is predicted by diet but not sociality" (DeCasien, Williams, and Higham 2017), the causal structure in Fig 5b groups diet, social group size, and relative brain size into a single latent factor, termed here "niche complexity", with another latent factor, "life-history pace", with loadings from the variables body size, age at maturity, and longevity. These latent factor groupings are not surprising, given that much work in primate socio-ecology has emphasized connections between diet and group size (Sterck, Watts, and Van Schaik 1997; Wrangham 1980), and both are consistently found to correlate with relative brain size (DeCasien, Williams, and Higham 2017; Dunbar and Shultz 2017); the integration of life history traits and body size is also well established (*CITE*).

A DPM analyzing the coevolution of these two latent factors (see Supplementary Materials) showed that evolutionary changes in life-history pace (LH) cause evolutionary changes in niche complexity (NC), but not vice versa (Fig. 5c). In other words, evolution of larger size and slower life history was inferred to precede increases in relative brain size, group size, and frugivory. This is consistent with the observations that repeated changes in size characterized mammalian evolution as lineages radiated into new niches (Pagel, O'Donovan, and Meade 2022) and changes in body size associated with new niches were arguably primarily responsible for changes in relative brain size (Smaers et al. 2021). Further, debates about the relative importance of diet and sociality in brain size evolution (DeCasien, Williams, and Higham 2017; Dunbar and Shultz 2017) seem to be moot as both are sides of the same socio-ecological coin intimately tied to relative brain size (see also Powell, Isler, and Barton 2017; Wartel, Lindenfors, and Lind 2019). Many extensions of this model are possible, and more theoretically motivated causal models should be explored to further advance our understanding of brain size evolution.

3.6.2 Evolution of social complexity across human societies

Human societies are currently and have historically been tremendously diverse, ranging from nomadic hunter-gatherers to large-scale states or empires. Many anthropologists have used comparative approaches to better understand this diversity (Hooper and Jaeggi 2024; Nunn 2011), including many applications of Pagel's discrete method (Sheehan et al. 2018; Watts et al. 2016). Here we summarize two studies that have used our DPM.

Ringen, Martin, and Jaeggi (2021) examined the rise of social complexity using a global sample of 186 pre-industrial societies that ranged from hunter-gatherers to agrarian empires (Figure 6A). They included nine measures of complexity and three measures of subsistence (Figure 6B), and compared three competing static causal models for the relationships between these variables: an "agricentric" model wherein agricultural intensification causes all other variables, a single factor model in which a latent "complexity" factor explains all observed variables (as in Turchin et al. 2018), and a two-factor model (as proposed by Chick 1997). Model comparison supported the two-factor model (the correlation structure of which is

shown in Figure 5B). The authors labeled one factor "resource-use intensification" (RI) and the other "technological and social differentiation" (TSD). A DPM analyzing the co-evolution of the two latent variables RI and TSD showed that, while these two variables are highly correlated (Figure 6C left), evolutionary changes in RI were inferred to cause changes in TSD but not vice versa (Figure 6C right). In other words, subsistence intensification was a leader, not a follower, in the evolution of complex societies. These results contrast with and complement Turchin et al. (2018), who found a single dimension of complexity (though in a mostly agricultural sample) and Sheehan et al. (2018), who found reciprocal causation between agricultural intensification and political complexity (though in a more regional sample).

Sheehan et al. (2023) studied the coevolution of religious and political authority in 97 Austronesian societies. Religious and political authority were both coded as four-level ordinal variables: absent, sub-local authority, local authority, and supra-local authority. Mapping these variables onto a phylogeny of Austronesian languages revealed that both religious and political authority had high phylogenetic signals, suggesting that a DPM could reasonably be used to assess the coevolution of these variables over cultural evolutionary time. Instead of dichotomizing the two ordinal variables to use Pagel's discrete method, as previous work had done (Sheehan et al. 2018; Watts et al. 2016), the authors explicitly modelled both variables as ordinal in a DPM. The model revealed that both religious and political authority coevolved reciprocally over time. In other words, increases in religious authority led to strong positive selection on political authority and, likewise, increases in political authority led to strong positive selection on religious authority. This is akin to "runaway selection" where both traits enter a positive feedback loop and increase concomitantly. This coevolutionary relationship makes sense in light of Austronesian ethnographies, which describe how both forms of authority are tightly interdependent and have often historically served to legitimize one another (e.g. Goodenough 2002).

4 Conclusion

xxx

References

Arnold, Christian, Luke J Matthews, and Charles L Nunn. 2010. "The 10kTrees Website: A New Online Resource for Primate Phylogeny." *Evolutionary Anthropology: Issues, News, and Reviews* 19 (3): 114–18. https://doi.org/10.1002/evan.20251.

Bielby, Jon, Geogina M Mace, Olaf RP Bininda-Emonds, Marcel Cardillo, John L Gittleman, Katherine E Jones, C David L Orme, and A Purvis. 2007. "The Fast-Slow Continuum in Mammalian Life History: An Empirical Reevaluation." *The American Naturalist* 169 (6): 748–57. https://doi.org/10.1086/516847.

Blomberg, Simon P, Theodore Garland Jr, and Anthony R Ives. 2003. "Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile." *Evolution* 57 (4): 717–45. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x.

Blomberg, Simon P, James G Lefevre, Jessie A Wells, and Mary Waterhouse. 2012. "Independent Contrasts and PGLS Regression Estimators Are Equivalent." *Systematic Biology* 61 (3): 382–91. https://doi.org/10.1093/sysbio/syr118.

Boddy, A M, M R McGowen, C C Sherwood, L I Grossman, M Goodman, and D E Wildman. 2012. "Comparative Analysis of Encephalization in Mammals Reveals Relaxed Constraints on Anthropoid Primate and Cetacean Brain Scaling." *Journal of Evolutionary Biology* 25 (5): 981–94. https://doi.org/10.1111/j.1420-9101.2012.02491.x.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76. https://doi.org/10.18637/jss.v076.i01.

Chick, Garry. 1997. "Cultural Complexity: The Concept and Its Measurement." *Cross-Cultural Research* 31 (4): 275–307. https://doi.org/10.1177/106939719703100401.

Clayton, Dale H, and Peter Cotgreave. 1994. "Relationship of Bill Morphology to Grooming Behaviour in Birds." *Animal Behaviour* 47 (1): 195–201. https://doi.org/10.1006/anbe.1994.1022.

Cornwallis, Charlie K, Carlos A Botero, Dustin R Rubenstein, Philip A Downing, Stuart A West, and Ashleigh S Griffin. 2017. "Cooperation Facilitates the Colonization of Harsh Environments." *Nature Ecology & Evolution* 1 (3): 0057. https://doi.org/10.1038/s41559-016-0057.

Dawson, Neal V, and Robert Weiss. 2012. "Dichotomizing Continuous Variables in Statistical Analysis: A Practice to Avoid." *Medical Decision Making* 32 (2): 225–26. https://doi.org/10.1177/0272989X12437605.

DeCasien, Alex R, Scott A Williams, and James P Higham. 2017. "Primate Brain Size Is Predicted by Diet but Not Sociality." *Nature Ecology & Evolution* 1 (5): 0112. https://doi.org/10.1038/s41559-017-0112.

Deffner, Dominik, Julia M Rohrer, and Richard McElreath. 2022. "A Causal Framework for Cross-Cultural Generalizability." *Advances in Methods and Practices in Psychological Science* 5 (3): 25152459221106366. https://doi.org/10.1177/25152459221106366.

Driver, Charles C, Johan HL Oud, and Manuel C Voelkle. 2017. "Continuous Time Structural Equation Modeling with R Package ctsem." *Journal of Statistical Software* 77: 1–35. https://doi.org/10.18637/jss.v077.i05.

Dunbar, Robin IM, and Susanne Shultz. 2017. "Why Are There so Many Explanations for Primate Brain Evolution?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1727): 20160244. https://doi.org/10.1098/rstb.2016.0244.

Dunn, Casey W, and Joseph F Ryan. 2015. "The Evolution of Animal Genomes." *Current Opinion in Genetics & Development* 35: 25–32. https://doi.org/10.1016/j.gde.2015.08.006.

Fitzpatrick, John L, Robert Montgomerie, Julie K Desjardins, Kelly A Stiver, Niclas Kolm, and Sigal Balshine. 2009. "Female Promiscuity Promotes the Evolution of Faster Sperm in Cichlid Fishes." *Proceedings of the National Academy of Sciences* 106 (4): 1128–32. https://doi.org/10.1073/pnas.080999010.

Gabry, Jonah, Rok Češnovar, Andrew Johnson, and Steve Bronder. 2024. *cmdstanr: R Interface to 'CmdStan'*. https://mc-stan.org/cmdstanr/.

Garamszegi, László Zsolt. 2014. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*. Springer.

Garland, Jr, Theodore, Albert F. Bennett, and Enrico L. Rezende. 2005. "Phylogenetic Approaches in Comparative Physiology." *Journal of Experimental Biology* 208 (16): 3015–35. https://doi.org/10.1242/jeb.01745.

Gonzalez-Voyer, Alejandro, and Achaz von Hardenberg. 2014. "An Introduction to Phylogenetic Path Analysis." In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, edited by László Zsolt Garamszegi, 201–29. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_8.

Goodenough, Ward H. 2002. *Under Heaven's Brow: Pre-Christian Religious Tradition in Chuuk*. American Philosophical Society.

Grafen, Alan. 1989. "The Phylogenetic Regression." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326 (1233): 119–57. https://doi.org/10.1098/rstb.1989.0106.

Granger, Clive WJ. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica: Journal of the Econometric Society*, 424–38. https://doi.org/10.2307/1912791.

Hadfield, JD, and S Nakagawa. 2010. "General Quantitative Genetic Methods for Comparative Biology: Phylogenies, Taxonomies and Multi-Trait Models for Continuous and Categorical Characters." *Journal of Evolutionary Biology* 23 (3): 494–508. https://doi.org/10.1111/j.1420-9101.2009.01915.x.

Hansen, Thomas F. 2014. "Use and Misuse of Comparative Methods in the Study of Adaptation." In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, edited by László Zsolt Garamszegi, 351–79. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_14.

Hardenberg, Achaz von, and Alejandro Gonzalez-Voyer. 2013. "Disentangling Evolutionary Cause-Effect Relationships with Phylogenetic Confirmatory Path Analysis." *Evolution* 67 (2): 378–87. https://doi.org/10.1111/j.1558-5646.2012.01790.x.

Harvey, Paul H, and Mark D Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press.

Healy, Kevin, Thomas HG Ezard, Owen R Jones, Roberto Salguero-Gómez, and Yvonne M Buckley. 2019. "Animal Life History Is Shaped by the Pace of Life and the Distribution of Age-Specific Mortality and Reproduction." *Nature Ecology & Evolution* 3 (8): 1217–24. https://doi.org/10.1038/s41559-019-0938-7.

Herculano-Houzel, Suzana. 2019. "Longevity and Sexual Maturity Vary Across Species with Number of Cortical Neurons, and Humans Are No Exception." *Journal of Comparative Neurology* 527 (10): 1689–1705. https://doi.org/10.1002/cne.24564.

Hooper, Paul L, and Adrian V Jaeggi. 2024. "Political Organization." In *Human Behavioral Ecology*, edited by Jeremy Koster, Brooke Scelza, and Mary K Shenk, 180–202. Cambridge Studies in Biological and Evolutionary Anthropology. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108377911.009.

Isler, Karin, E Christopher Kirk, Joseph M A Miller, Gene A Albrecht, Bruce R Gelvin, and Robert D Martin. 2008. "Endocranial Volumes of Primate Species: Scaling Analyses Using a Comprehensive and Reliable Data Set." *Journal of Human Evolution* 55 (6): 967–78. https://doi.org/10.1016/j.jhevol.2008.08.004.

Isler, Karin, and Carel P van Schaik. 2009. "The Expensive Brain: A Framework for Explaining Evolutionary Changes in Brain Size." *Journal of Human Evolution* 57 (4): 392–400. https://doi.org/10.1016/j.jhevol.2009.04.009.

Isler, Karin, and Carel P Van Schaik. 2014. "How Humans Evolved Large Brains: Comparative Evidence." *Evolutionary Anthropology: Issues, News, and Reviews* 23 (2): 65–75. https://doi.org/10.1002/evan.21403.

Ives, Anthony R, and Theodore Garland Jr. 2010. "Phylogenetic Logistic Regression for Binary Dependent Variables." *Systematic Biology* 59 (1): 9–26. https://doi.org/10.1093/sysbio/syp074.

Ives, Anthony R, and Matthew R Helmus. 2011. "Generalized Linear Mixed Models for Phylogenetic Analyses of Community Structure." *Ecological Monographs* 81 (3): 511–25. https://doi.org/10.1890/10-1264.1.

Kamilar, Jason M, and Natalie Cooper. 2013. "Phylogenetic Signal in Primate Behaviour, Ecology and Life History." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1618): 20120341. https://doi.org/10.1098/rstb.2012.0341.

Kappeler, Peter M, and Luca Pozzi. 2019. "Evolutionary Transitions Toward Pair Living in Nonhuman Primates as Stepping Stones Toward More Complex Societies." *Science Advances* 5 (12): eaay1276. https://doi.org/10.1126/sciadv.aay1276.

Lynch, Michael. 1991. "Methods for the Analysis of Comparative Data in Evolutionary Biology." *Evolution* 45 (5): 1065–80. https://doi.org/10.1111/j.1558-5646.1991.tb04375.x.

Mace, Ruth, and Clare J Holden. 2005. "A Phylogenetic Approach to Cultural Evolution." *Trends in Ecology & Evolution* 20 (3): 116–21.

MacLean, Evan L, Luke J Matthews, Brian A Hare, Charles L Nunn, Rindy C Anderson, Filippo Aureli, Elizabeth M Brannon, et al. 2012. "How Does Cognition Evolve? Phylogenetic Comparative Psychology." *Animal Cognition* 15: 223–38. https://doi.org/10.1007/s10071-011-0448-8.

Martin, Jordan S, Erik J Ringen, Pavel Duda, and Adrian V Jaeggi. 2020. "Harsh Environments Promote Alloparental Care Across Human Societies." *Proceedings of the Royal Society B* 287 (1933): 20200758. https://doi.org/10.1098/rspb.2020.0758.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan, 2nd Edition*. 2nd ed. CRC Press. http://xcelab.net/rm/statistical-rethinking/.

McNamara, John M. 2022. "Game Theory in Biology: Moving Beyond Functional Accounts." *The American Naturalist* 199 (2): 179–93. https://doi.org/10.1086/717429.

Miller, Ian F, Robert A Barton, and Charles L Nunn. 2019. "Quantitative Uniqueness of Human Brain Evolution Revealed Through Phylogenetic Comparative Analysis." *eLife* 8: e41250. https://doi.org/10.7554/eLife.41250.

Modrák, Martin, Angie H Moon, Shinyoung Kim, Paul Bürkner, Niko Huurre, Kateřina Faltejsková, Andrew Gelman, and Aki Vehtari. 2023. "Simulation-Based Calibration Checking for Bayesian Computation: The Choice of Test Quantities Shapes Sensitivity." *Bayesian Analysis* 1 (1): 1–28. https://doi.org/10.1214/23-BA1404.

Navalón, Guillermo, Jen A Bright, Jesús Marugán-Lobón, and Emily J Rayfield. 2019. "The Evolutionary Relationship Among Beak Shape, Mechanical Advantage, and Feeding Ecology in Modern Birds." *Evolution* 73 (3): 422–35. https://doi.org/10.1111/evo.13655.

Navarrete, Ana F, Simon M Reader, Sally E Street, Andrew Whalen, and Kevin N Laland. 2016. "The Coevolution of Innovation and Technical Intelligence in Primates." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1690): 20150186. https://doi.org/10.1098/rstb.2015.0186.

Nunn, Charles L. 2011. *The Comparative Approach in Evolutionary Anthropology and Biology*. University of Chicago Press.

O'Connor, Emily A, and Charlie K Cornwallis. 2022. "Immunity and Lifespan: Answering Long-Standing Questions with Comparative Genomics." *Trends in Genetics* 38 (7): 650–61. https://doi.org/10.1016/j.tig.2022.02.014.

Pagel, Mark. 1993. "Seeking the Evolutionary Regression Coefficient: An Analysis of What Comparative Methods Measure." *Journal of Theoretical Biology* 164 (2): 191–205. https://doi.org/10.1006/jtbi.1993.1148.

———. 1994. "Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255 (1342): 37–45. https://doi.org/10.1098/rspb.1994.0006.

———. 1999. "The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies." *Systematic Biology* 48 (3): 612–22.

Pagel, Mark, Ciara O'Donovan, and Andrew Meade. 2022. "General Statistical Model Shows That Macroevolutionary Patterns and Processes Are Consistent with Darwinian Gradualism." *Nature Communications* 13 (1): 1113. https://doi.org/10.1038/s41467-022-28595-z.

Pearl, Judea. 2009. *Causality*. Cambridge University Press.

Powell, Lauren E, Karin Isler, and Robert A Barton. 2017. "Re-Evaluating the Link Between Brain Size and Behavioural Ecology in Primates." *Proceedings of the Royal Society B: Biological Sciences* 284 (1865): 20171765. https://doi.org/10.1098/rspb.2017.1765.

Ringen, Erik J, Pavel Duda, and Adrian V Jaeggi. 2019. "The Evolution of Daily Food Sharing: A Bayesian Phylogenetic Analysis." *Evolution and Human Behavior* 40 (4): 375–84.

Ringen, Erik J, Jordan S Martin, and Adrian V Jaeggi. 2021. "Novel Phylogenetic Methods Reveal That Resource-Use Intensification Drives the Evolution of "Complex" Societies." *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/wfp95.

Royston, Patrick, Douglas G Altman, and Willi Sauerbrei. 2006. "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine* 25 (1): 127–41. https://doi.org/10.1136/bmj.332.7549.1080.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100 (469): 322–31. https://doi.org/10.1198/016214504000001880.

Salguero-Gómez, Roberto, Owen R Jones, Eelke Jongejans, Simon P Blomberg, David J Hodgson, Cyril Mbeau-Ache, Pieter A Zuidema, Hans De Kroon, and Yvonne M Buckley. 2016. "Fast–Slow Continuum and Reproductive Strategies Structure Plant Life-History Variation Worldwide." *Proceedings of the National Academy of Sciences* 113 (1): 230–35. https://doi.org/10.1073/pnas.1506215112.

Schölkopf, Bernhard, Dominik Janzing, Jonas Peters, Eleni Sgouritsa, Kun Zhang, and Joris Mooij. 2013. "Semi-Supervised Learning in Causal and Anticausal Settings." In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, edited by Bernhard Schölkopf, Zhiyuan Luo, and Vladimir Vovk, 129–41. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_13.

Sheehan, Oliver, Joseph Watts, Russell D Gray, and Quentin D Atkinson. 2018. "Coevolution of Landesque Capital Intensive Agriculture and Sociopolitical Hierarchy." *Proceedings of the National Academy of Sciences* 115 (14): 3628–33. https://doi.org/10.1073/pnas.1714558115.

Sheehan, Oliver, Joseph Watts, Russell D Gray, Joseph Bulbulia, Scott Claessens, Erik J Ringen, and Quentin D Atkinson. 2023. "Coevolution of Religious and Political Authority in Austronesian Societies." *Nature Human Behaviour* 7 (1): 38–45. https://doi.org/10.1038/s41562-022-01471-y.

Shipley, Bill. 2016. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R*. Cambridge university press.

Shultz, Susanne, Christopher Opie, and Quentin D Atkinson. 2011. "Stepwise Evolution of Stable Sociality in Primates." *Nature* 479 (7372): 219–22. https://doi.org/10.1038/nature10601.

Smaers, Jeroen B, Ryan S Rothman, Daphne R Hudson, Amy M Balanoff, Brian Beatty, Dina KN Dechmann, Dorien de Vries, et al. 2021. "The Evolution of Mammalian Brain Size." *Science Advances* 7 (18): eabe2101. https://doi.org/10.1126/sciadv.abe2101.

Sterck, Elisabeth H M, David P Watts, and Carel P Van Schaik. 1997. "The Evolution of Female Social Relationships in Nonhuman Primates." *Behavioral Ecology and Sociobiology* 41: 291–309. https://doi.org/10.1007/s002650050390.

Stott, Iain, Roberto Salguero-Gómez, Owen R Jones, Thomas HG Ezard, Marlène Gamelon, Shelly Lachish, Jean-Dominique Lebreton, Emily G Simmonds, Jean-Michel Gaillard, and Dave J Hodgson. 2024. "Life Histories Are Not Just Fast or Slow." *Trends in Ecology & Evolution*. https://doi.org/10.1016/j.tree.2024.06.001.

Symonds, Matthew R. E., and Simon P. Blomberg. 2014. "A Primer on Phylogenetic Generalised Least Squares." In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, edited by László Zsolt Garamszegi, 105–30. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_5.

Talts, Sean, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. 2018. "Validating Bayesian Inference Algorithms with Simulation-Based Calibration." *arXiv*. https://doi.org/10.48550/arXiv.1804.06788.

Thayer, Zaneta M, Meredith A Wilson, Andrew W Kim, and Adrian V Jaeggi. 2018. "Impact of Prenatal Stress on Offspring Glucocorticoid Levels: A Phylogenetic Meta-Analysis Across 14 Vertebrate Species." *Scientific Reports* 8 (1): 1–9. https://doi.org/10.1038/s41598-018-23169-w.

Turchin, Peter, Thomas E Currie, Harvey Whitehouse, Pieter François, Kevin Feeney, Daniel Mullins, Daniel Hoyer, et al. 2018. "Quantitative Historical Analysis Uncovers a Single Dimension of Complexity That Structures Global Variation in Human Social Organization." *Proceedings of the National Academy of Sciences* 115 (2): E144–E151. https://doi.org/10.1073/pnas.1708800115.

Warrell, Jonathan, and Mark Gerstein. 2020. "Cyclic and Multilevel Causation in Evolutionary Processes." *Biology & Philosophy* 35 (5): 50. https://doi.org/10.1007/s10539-020-09753-3.

Wartel, Andreas, Patrik Lindenfors, and Johan Lind. 2019. "Whatever You Want: Inconsistent Results Are the Rule, Not the Exception, in the Study of Primate Brain Evolution." *PLoS One* 14 (7): e0218655. https://doi.org/10.1371/journal.pone.0218655.

Watts, Joseph, Oliver Sheehan, Quentin D Atkinson, Joseph Bulbulia, and Russell D Gray. 2016. "Ritual Human Sacrifice Promoted and Sustained the Evolution of Stratified Societies." *Nature* 532 (7598): 228–31. https://doi.org/10.1038/nature17159.

Westreich, Daniel, and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177 (4): 292–98. https://doi.org/10.1093/aje/kws412.

Wrangham, Richard W. 1980. "An Ecological Model of Female-Bonded Primate Groups." *Behaviour*, 262–300.

Wright, Jonathan, Geir H Bolstad, Yimen G Araya-Ajoy, and Niels J Dingemanse. 2019. "Life-History Evolution Under Fluctuating Density-Dependent Selection and the Adaptive Alignment of Pace-of-Life Syndromes." *Biological Reviews* 94 (1): 230–47. https://doi.org/10.1111/brv.12451.