The use of multi-response models to improve inferences about natural selection

Authors:

Sarah Dobson[1], Joel Pick[1], Elizabeth Mittell[1], Loeske Kruuk[1] and Craig Walling[1].

[1]Institute of Ecology and Evolution, School of Biological Sciences, The University of Edinburgh, United Kingdom

Corresponding author: Sarah Dobson

Corresponding author's email address: sarah.1.dobson@outlook.com

1. Abstract

Natural selection, the relationship between trait and fitness, is a key determinant of evolutionary change and population adaptation. Therefore, accurate estimation of natural selection is important. In 1983, Lande and Arnold proposed a simple regression-based approach which allows the measurement of selection on a range of traits whilst accounting for confounding variables. However, issues remain with its application in wild populations which can bias estimates, including assumptions around causality and whether selection on a trait is hard or soft. We highlight how, when fitness and traits are measured repeatedly across individuals and/or common environments, we can identify these issues by comparing directional selection gradients decomposed across different hierarchical levels. We outline the theory behind this and show how multi-response models provide a readily available statistical tool to implement this approach. We then use an empirical example to illustrate how to implement our method and interpret the results. Our approach builds upon previous works to allow greater inference to be drawn from existing datasets, particularly when no genetic information is available. This should facilitate improved interpretation of estimates of selection in wild populations, and ultimately, our understanding of the selection process.

Keywords: directional selection, bias, repeated measures, soft selection, hard selection, quantitative genetics

2. Introduction

Natural selection is the difference in survival and/or reproduction of individuals as a result of differences in their trait phenotypes, which is broadly measured as the relationship between fitness and a trait Darwin (1859); (Falconer & Mackay, 1996). The strength of natural selection is an important determinant of the rate of adaptation of a population (Charmantier et al., 2014; Falconer & Mackay, 1996; Hoekstra et al., 2001). Therefore, understanding variation in the strength and direction of natural selection can provide crucial insights into potential causes of differences in fitness between individuals and how populations might evolve in response to changing environmental pressures (Caruso et al., 2017; Siepielski et al., 2009). As a result, accurately estimating natural selection is a key goal in evolutionary biology that has implications across a range of disciplines. Lande and Arnold (1983) proposed a simple linear regression-based approach (the 'Lande-Arnold method') that can account for confounding variables. This has allowed evolutionary biologists and ecologists to measure selection on a range of traits, using the regression of relative fitness w on a trait z:

$$w_i = \mu_0 + \beta_z z_i + e_i. \quad [1]$$

Where w_i is the fitness of an individual relative to the population mean at observation i, μ_0 is the global intercept, z is the trait value and e is the residual. The linear effect of the trait on fitness, β_z, is the directional selection gradient. This Lande-Arnold selection gradient is easy to measure, interpret and can be expanded to measure quadratic and correlational selection

(Lande & Arnold, 1983; Stinchcombe et al., 2008; Svensson, 2023). The Lande-Arnold method of measuring selection led to a rapid expansion in the number of published estimates of selection, to the extent that, most meta-analyses on the strength of selection have more estimates of Lande-Arnold derived selection gradients than estimates from other methods (Caruso et al., 2017; Charmantier et al., 2024; Hoekstra et al., 2001; Kingsolver et al., 2001; Morrissey & Hadfield, 2012; Siepielski et al., 2009; Svensson, 2023).

Despite the benefits of the Lande-Arnold approach to estimating natural selection, and its widespread use in studies of wild populations, issues remain with its application; here we focus on two.

The first, well-documented, issue with the Lande-Arnold approach to estimating directional selection on a trait is that it inherently assumes a causal relationship, i.e., that fitness changes as a direct result of a change in the trait being analysed (Hadfield, 2008; Mitchell-Olds & Shaw, 1987; Morrissey et al., 2010; Rausher, 1992; Svensson, 2023). This can be particularly problematic in wild populations where many confounding environmental factors may result in a trait-fitness relationship and thus a non-zero estimate of selection, when in reality there is no causal relationship between the trait and fitness (Kingsolver et al., 2012; Rausher, 1992). Confounding factors can be dealt with by including them as covariates using the Lande-Arnold method which allows measures of selection to be corrected when confounding variables are both known and measured (Morrissey & Henshaw, 2022). However, in wild observational studies there are theoretically an unlimited number of unknown confounding variables, which poses a huge challenge when relying on Lande-Arnold based approaches to estimate selection and infer causality (Svensson, 2023).

The second, less well-known issue is that whether the Lande-Arnold approach provides an accurate estimate of natural selection depends on whether selection on a trait is hard or soft. Here, we define hard selection as cases where the fitness of individuals depends solely on an individual's absolute trait value and is independent of the trait values of other conspecifics in the local environment (i.e. selection is density and frequency independent). We define soft selection as cases where the fitness of individuals depends on their trait value relative to the trait value of other conspecifics in the local environment with which they interact (i.e. selection is density and frequency dependant) (Bell et al., 2021; Wallace, 1975; Wallace & Lewontin, 1968). For more detailed examples and a discussion of soft and hard selection see Bell et al. (2021). When fitness is relativised by the global mean, the Lande-Arnold method estimates a global linear relationship between trait measurements and relative fitness. If selection is soft, selection gradients estimated this way may be biased towards zero. Together, these issues may contribute to the mismatch between the observed and predicted response to selection that is common across wild populations (Brookfield, 2016; Hansen & Houle, 2004; Merilä et al., 2001; Pemberton, 2010; Pujol et al., 2018).

In wild populations that are monitored across time and/or space, fitness and the trait(s) of interest are often both measured repeatedly on the same individuals across their lifetimes. Fitness and trait(s) will also be measured repeatedly across environments that individuals share with each other, such as the timepoint or location over which selection is measured. This means that any environmental conditions, such as weather, food availability or population density effects that occur in a specific location or timepoint and affect an individual's trait and/or fitness measurements are shared by individuals within that environment. Here, we will introduce how, when fitness and traits are measured repeatedly across individuals and/or common environments, we can use this information to decompose the association between trait(s) and fitness into its causal components. We will show how this decomposition of selection into hierarchical levels can be used to assess whether selection is consistent with causality and to test and correct selection estimates for soft and hard selection. We then show how, without additional data, multi-response models (sometimes referred to as multivariate models) provide a readily available statistical tool to implement this approach. Whilst the statistical models we promote have been previously described and widely used, their application to trait-fitness relationships is currently limited to separating associations into genetic and non-genetic hierarchical levels and thus directly estimating the predicted evolutionary change in a trait without any explicit model of selection (Morrissey et al., 2012; Rausher, 1992; Thomson et al., 2017). This article highlights how, even without genetic information, multi-response models can be applied more broadly to improve our understanding of trait-fitness relationships and thus the process of natural selection.

3. Theory

3.1 Decomposing selection gradients to assess consistency with causality

To understand how we can infer if our estimates of selection are consistent with causality, and test and correct for soft vs hard selection, we reconsider the Lande-Arnold equation in its simplest form (equation [1]), but where a trait and fitness have been measured repeatedly across individuals and common environments:

$$w_{ijk} = \mu_0 + \beta_z z_{ijk} + o_{ijk}. \qquad [2]$$

For simplicity, o is everything that is not explained by z: $o_{ijk} = u_j + c_k + e_{ijk}$ where u_j and c_k are the random effects of individual identity j and common environment. As both the trait and fitness are measured repeatedly across individuals and common environments, we can also decompose the phenotypic value of the trait:

$$z_{ijk} = u_{z,j} + c_{z,k} + e_{z,i}. \qquad [3]$$

Where u, c and e are individual, common environment and residual effects on the trait z, respectively. We combine equations [2] and [3], to give:

$$w_{ijk} = \mu_0 + \beta_z\left(u_{z,j} + c_{z,k} + e_{z,i}\right) + o_{ijk}, \qquad [4]$$

which, when expanded, gives us:

$$w_{ijk} = \mu_0 + \beta_z u_{z,j} + \beta_z c_{z,k} + \beta_z e_{z,i} + o_{ijk}. \qquad [5]$$

Equation [7] shows that if there is a causal effect of z on w, in other words, an increase in any component of the trait phenotype leads to the same increase in fitness, the gradients estimated for each hierarchical level should be the same. This is because the Lande-Arnold method is estimating a single selection gradient β_z, which inherently assumes that all decomposed gradients are of equivalent magnitude and direction. If this is not true, this suggests the relationship between a trait and fitness is not causal and β_z would give a biased estimate of selection. These decomposed gradients represent associations between trait and fitness at different hierarchical levels as follows:

$$w_{ijk} = \mu_0 + \beta_u u_{z,j} + \beta_c c_{z,k} + \beta_e e_{z,i} + o_{ijk}. \qquad [6]$$

Where the β's represent the linear relationship between different hierarchical levels of the trait and relative fitness: β_u among individuals, β_c among common environments, and β_e within individuals, respectively. The selection gradient estimated in a Lande-Arnold regression is composed of these decomposed gradients as shown below:

$$\beta_z = \beta_u \frac{\sigma^2_{u_z}}{\sigma^2_z} + \beta_c \frac{\sigma^2_{c_z}}{\sigma^2_z} + \beta_e \frac{\sigma^2_{e_z}}{\sigma^2_z}. \qquad [7]$$

Where σ^2's represents the total (σ^2_z), among individuals ($\sigma^2_{u_z}$), among common environments ($\sigma^2_{c_z}$), and within individual ($\sigma^2_{e_z}$) variances of the trait. The Lande-Arnold derived selection gradient (β_z) is therefore composed of all these gradients, weighted by the proportion of the total phenotypic variance of the trait explained by variation at these different hierarchical levels. The derivation of equation [7] from equation [6] can be found in supplementary material S1.

As shown above, under a simple model of causality, where the trait directly affects fitness, we expect the selection gradients at each hierarchical level to be equal in both direction and magnitude to each other and the overall β_z (Equation [6]; Figure 1A; (Bonnet & Postma, 2018; Gauzere et al., 2021; Hadfield et al., 2013; Pick et al., 2016). Thus, comparing the magnitude and direction of these decomposed selection gradients, can be used to infer whether selection on a trait is consistent with causality. An exception to this would be when selection is measured on a trait with considerable measurement error, which may result in β_e estimates that are biased downwards (Dingemanse et al., 2021). Inconsistency with causality would be indicated if one or more of the decomposed selection gradients differed from each other. For example, if there is a non-causal relationship between relative fitness and a trait driven by an environmental variable that changes among common environments, this would result in an increase in β_c which would upwardly bias β_z (Eq, 8; Figure 1b), and thus impact our interpretation of selection estimated using the Lande-Arnold method (Rausher, 1992; Stinchcombe et al., 2002).

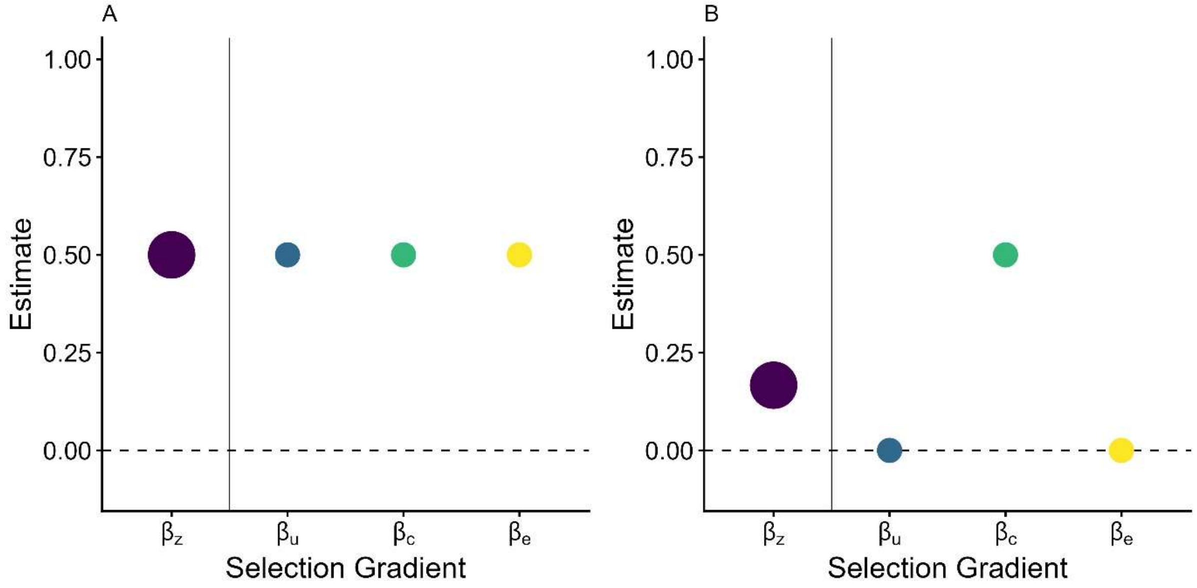

Figure 1: Examples of Lande-Arnold derived selection gradients and their decomposition to selection gradients at different hierarchical levels when selection is: (A) consistent with causality; versus (B) inconsistent with causality and instead confounded by an environmental variable that changes across common environments. β_z (purple) is the Lande-Arnold derived selection gradient, β_u (blue) is the among individual selection gradient, β_c (green) is the common environment selection gradient and β_e (yellow) is the within individual selection gradient.

3.2 Decomposing selection gradients to assess and correct for soft and hard selection

As described above, the Lande-Arnold method of estimating selection gradients, where relative fitness is regressed on the trait(s) of interest, estimates a global linear relationship between trait measurements and relative fitness. However, this approach can bias estimates of selection downwards if selection is soft. Although selection may actually be on a continuum between hard and soft selection, we discuss these as discrete modes of selection for clarity.

To illustrate this, Figures 2A and 2B show a theoretical example where the selection gradients for trait *z* were estimated across two years of measurements under scenarios of hard and soft selection, respectively. Here, all individuals live within the same local environment, and the mean value of the trait changes between years, being larger in year 2. The strength of selection is the same in both scenarios. When selection is hard the fitness of an individual is dependent on its trait value independent of the population mean (Figure 2A). Therefore, as the population trait mean changes across years, so too does mean absolute fitness. Under this scenario, we would obtain the correct selection gradient using the Lande-Arnold method when selection is measured across common environments (Figure 2A). However, when selection is soft, the absolute fitness of individuals for the second year falls into the same range as the year before because an individual's fitness is dependent on the trait value relative to the composition of the population at that time. This leads to an underestimation of the selection gradient when measured from data pooled across both years, while the selection gradient is the same in both years when measured separately for each year (Figure 2B).

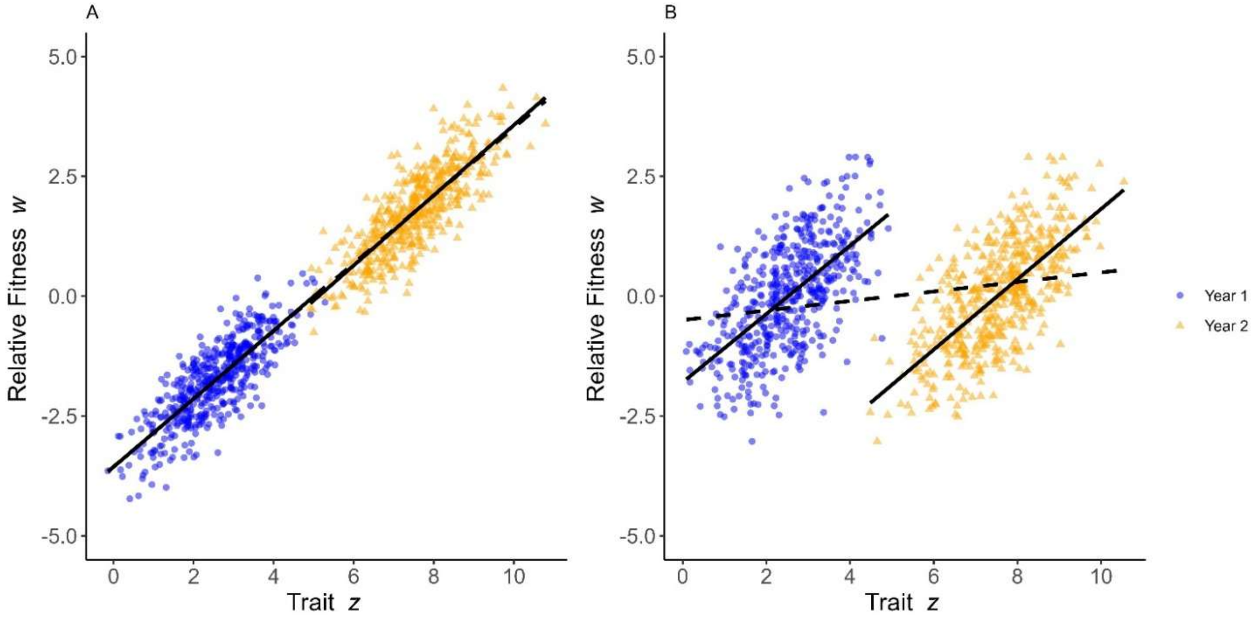

Figure 2: The linear relationship between relative fitness *w* and trait *z* (i.e. the selection gradient) when (A) hard selection and (B) soft selection is occurring, where fitness has been

standardised by the global mean. Blue and orange denote w and z values measured in year 1 (blue) and year 2 (orange), respectively. Solid lines show the selection gradients measured separately for each year, while dashed lines show the selection gradients measured when data from both years are pooled together. In (A) the dashed line cannot be seen as it is identical to the solid line. Under hard selection (A) the selection gradients estimated separately for each year and pooled across years agree, whereas under soft selection (B) the pooled estimate is downwardly biased. Note that the slopes of the true selection gradients do not change between years, which would occur if the strength of selection was varying across years.

Before we describe how selection gradient estimates can be corrected when selection is soft, we must first discuss how the pattern of decomposed selection gradients that is consistent with causality changes depending on whether selection is hard or soft. When hard selection occurs the fitness of an individual is independent of the trait mean of the population. If the trait mean of the population changes across common environments, the associated relative fitness also changes from year to year (Figure 2A). Therefore, we expect the among common environment selection gradient (β_c) to be different from zero and equivalent in both direction and magnitude to all other selection gradients when selection is both causal and hard (Figure 3A). However, if selection is soft, there will be no difference in fitness, and thus relative fitness of individuals among common environments (Figure 2B). Therefore, we do not expect there to be a relationship between mean trait value and mean relative fitness between common environments if the trait mean varies across those common environments. Evidence for soft selection, is therefore provided by a common environment selection gradient (β_c) of zero, but other selection gradients (e.g., β_u and β_e) different from zero and equivalent to each other. If selection is soft and consistent with causality, we should still see the relationship between trait and relative fitness present both within and among individuals (Figure 3B). Soft selection will result in a downward bias in β_z if the contribution of the negligible among common environment selection gradient is not corrected (Figures 2B & 3B), as estimates of β_z will include β_c as per equation [7].

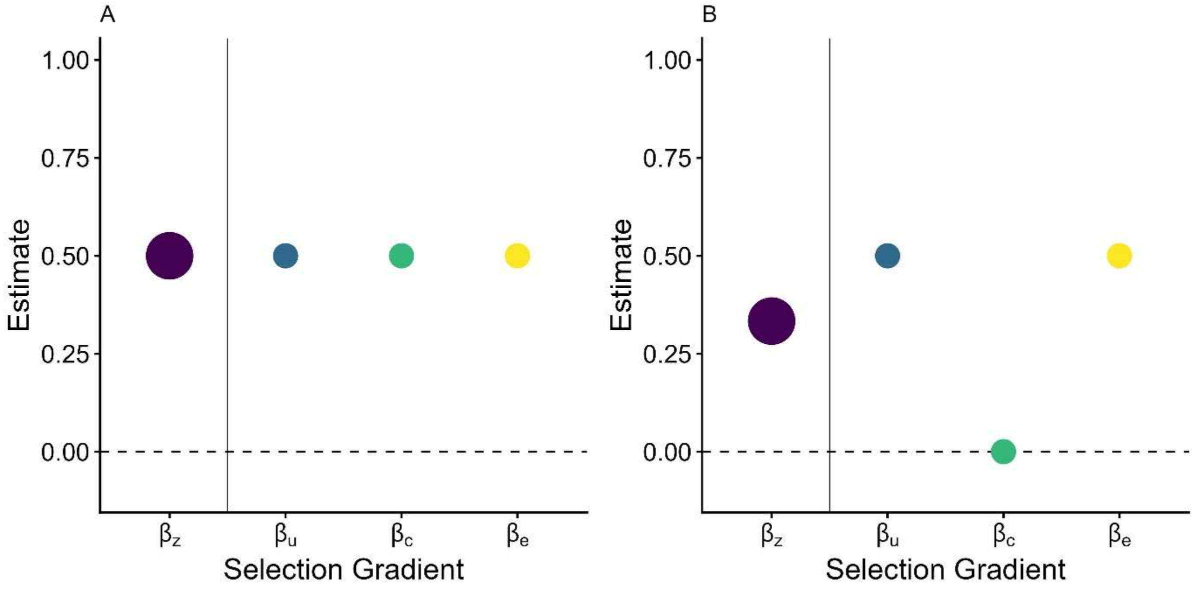

Figure 3: Examples of Lande-Arnold derived selection gradients and their decomposition to selection gradients at different hierarchical levels when selection is: (A) consistent with hard selection; versus (B) consistent with soft selection. β_z (purple) is the Lande-Arnold derived selection gradient, β_u (blue) is the among individual selection gradient, β_c (green) is the common environment selection gradient and β_e (yellow) is the within individual selection gradient.

Corrected estimates of the selection gradient (β_z') can be obtained by removing any variation in relative fitness or the trait across common environments as well as their relationship to each other from equation 9. This includes removal from the total variance σ_z^2. Thus, in the above example the corrected estimate of the selection gradient (β_z') is:

$$\beta_z' = \beta_u \frac{\sigma_{u_z}^2}{\sigma_{u_z}^2 + \sigma_{e_z}^2} + \beta_e \frac{\sigma_{e_z}^2}{\sigma_{u_z}^2 + \sigma_{e_z}^2}. \qquad [8]$$

It is also important to note that other methods have been developed to correct and test for soft selection. More in-depth discussion and visualisation these different approaches affect selection gradient estimates and testing for hard vs soft selection, is in the supplementary material S2.

3.3 Using genetic information to decompose selection gradients

Previous studies have decomposed the covariance between fitness and trait into genetic versus non-genetic hierarchical levels (Morrissey et al., 2012; Rausher, 1992; Stinchcombe et al., 2002). This can be achieved when data on the relatedness among individuals is available (Kruuk, 2004). However, to date these approaches have not discussed what can be learnt from partitioning the non-genetic association further into any of its components, instead focusing on the genetic association. However, in isolation this genetic association does not provide an explicit model of selection, with a genetic association between a trait and fitness potentially resulting from a genetic correlation with another trait that is causally related to fitness (i.e. under selection; (Morrissey et al., 2012; Rausher, 1992; Thomson et al., 2017). Below we highlight how the model we present above can be extended to obtain further information on whether selection on a trait is likely to be causal by partitioning β_u into additive genetic and individual permanent environmental hierarchical levels as shown below:

$$\beta_u u_{z,j} = \beta_a a_{z,j} + \beta_{pe} pe_{z,j}. \qquad [9]$$

Where the additive genetic selection gradient (β_a) is the linear relationship between individual breeding values for a trait and individual breeding values for fitness. The permanent environment selection gradient (β_{pe}) is the linear relationship between mean trait value across individuals' permanent environments and mean relative fitness across individuals' permanent environments. Thus, β_u is weighted by both β_a and β_{pe} as follows:

$$\beta_u = \beta_a \frac{\sigma^2_{a_z}}{\sigma^2_{a_z} + \sigma^2_{pe_z}} + \beta_{pe} \frac{\sigma^2_{pe_z}}{\sigma^2_{a_z} + \sigma^2_{pe_z}}. \qquad [10]$$

Therefore, equation [8] becomes:

$$w_{ijk} = \mu_0 + \beta_a a_{z,j} + \beta_{pe} pe_{z,j} + \beta_c c_{z,k} + \beta_e e_{z,i} + e_{w,i}. \qquad [11]$$

Comparison of the genetic selection gradient with other selection gradients measured in the same model provides more information about whether a causal relationship between the trait and relative fitness is likely (Figure 4). When a trait has a causal effect on fitness, we expect the selection gradient between trait and relative fitness between individuals' breeding values and permanent environments to be equivalent to each other. Therefore, assuming hard selection, selection is consistent with causality when all selection gradients are equivalent (Figure 4A). Decomposing β_u into permanent environment and genetic selection gradients can also facilitate the interpretation of the relationship between a trait and fitness in more complex scenarios. For example, if β_z is non-zero, but both within individual (β_e) and among common environment (β_c) selection gradients are zero (suggesting the association is not causal), decomposition of β_u into β_a and β_{pe} can help us separate cases where β_z is confounded by an environmental variable that is consistent among individuals (Figure 4B) from cases where β_z is being confounded by genetic correlations with another trait that is under selection (Figure 4C). Additionally, decomposing β_u into permanent environment and genetic selection gradients can help us identify cases where selection on the trait is consistent with causality, but the trait is genetically correlated with another trait which is being selected against. In this scenario we expect all selection gradients to be equal in magnitude and direction to each other except the genetic selection gradient β_a, which may be biased downwards by any magnitude depending on the strength of the genetic correlation with the other trait and the strength of selection acting on the other trait (Figure 4D).

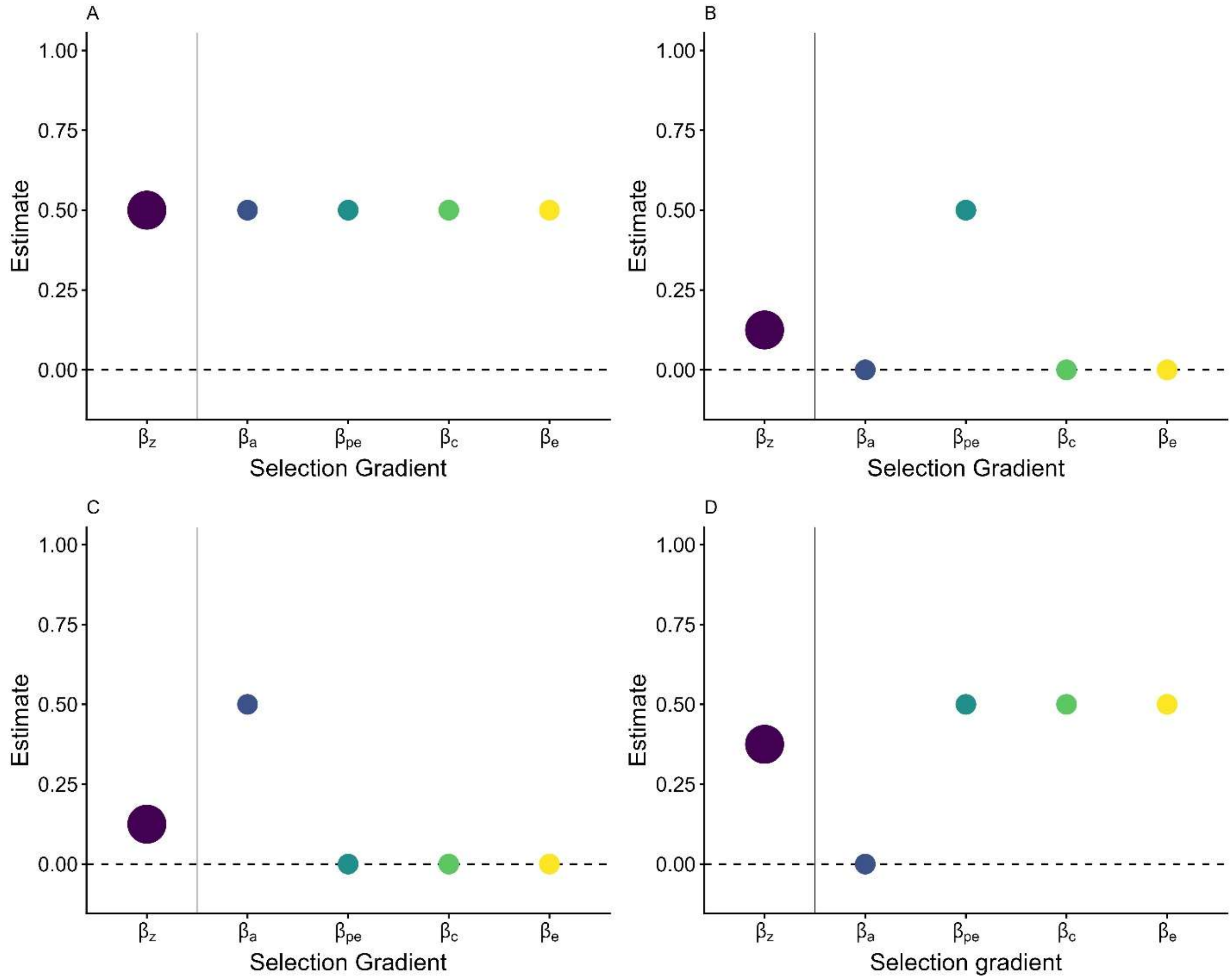

Figure 4. Examples of Lande-Arnold derived selection gradients and decomposed selection gradients where genetic information is available that are (A) consistent with causality, (B) where selection is inconsistent with causality and is instead confounded by a variable that is consistent among individuals, (C) where selection is inconsistent with causality and is instead confounded by a genetic correlation with a trait that is under selection and (D) where selection is consistent with causality but the trait is genetically correlated with a trait that is being selected against. Where β_z (purple) is the Lande-Arnold derived selection gradient, β_a (blue) is the genetic selection gradient, β_{pe} (teal) is the permanent environment selection gradient, β_c (green) is the among common environment selection gradient and β_e (yellow) is the within individual selection gradient. Hard selection is assumed to be occurring in all scenarios. In D, the genetic selection gradient depicted is zero; but it can take any magnitude depending on the strength of the genetic correlation with the other trait and the strength of selection acting on the other trait.

4. Estimating decomposed selection gradients

4.1 Multi-response models

The decomposition of directional selection gradients can be readily achieved with multi-response models. Where multiple response variables are simultaneously modelled, the variances for each response variable and covariances between them at different hierarchical levels can be estimated. These models are already frequently used in quantitative genetics and behavioural ecology to estimate the genetic covariance between a trait and fitness (Kruuk et al., 2002; Morrissey, 2014; Morrissey et al., 2010; Reid, 2012), and the phenotypic and genetic covariances between traits (Clements et al., 2011; Dingemanse & Dochtermann, 2013; Dobson et al., 2023; Houslay & Wilson, 2017; Reid & Wolak, 2018). Thus, there are many packages available, both frequentist and Bayesian, that can run these models, and a number of papers and tutorials outlining how these models are implemented (de Villemereuil et al., 2018; Houslay & Wilson, 2017). While multi-response models are capable of measuring selection gradients (Morrissey et al., 2010), very few studies have used or advocated their use in this context (but see Dingemanse et al. (2021); Mittell et al. (2024)). In wild populations with repeated measures across individuals and common environments, we can fit a multi-response model with relative fitness *w* and the trait *z* as two responses, with random intercepts for individual identity and common environments as specified below:

$$w_{ijk} \sim \mu_{w_0} + \beta_{w_d} d_{w_{ijk}} + u_{w_j} + c_{w_k} + e_{w_{ijk}}. \qquad [12a]$$

$$z_{ijk} \sim \mu_{z_0} + u_{z_j} + c_{z_k} + e_{z_{ijk}}. \qquad [12b]$$

Where global intercepts and random intercepts at each hierarchical level are measured separately for trait *z* and relative fitness *w*, *d* is any variable that may affect relative fitness independent of the trait and β_{w_d} is the linear effect of variable *d* on *w*. When measuring β_z using multiple regression, the random intercepts *u* and *c*, and any unexplained effects *e*, are only modelled and estimated for *w*, and *z* is assumed to be free of error. However, multi-response models measure the variances in and covariances between *z* and *w* across hierarchical levels *u* and *c*, and any residual variance and covariance *e,* therefore allowing error in *z* to be explicitly estimated. The random intercepts for *u* and *c*, and residual errors *e*, are distributed assuming a multivariate normal distribution with variance-covariance structures as follows:

$$[u_{j_w}, u_{j_z}] \sim MVN(0, \Sigma_u), \Sigma_u = \begin{bmatrix} \sigma^2_{u_w} & \sigma_{u_{w,z}} \\ \sigma_{u_{w,z}} & \sigma^2_{u_z} \end{bmatrix}. \qquad [12c]$$

$$[c_{k_w}, c_{k_z}] \sim MVN(0, \Sigma_c), \Sigma_c = \begin{bmatrix} \sigma^2_{c_w} & \sigma_{c_{w,z}} \\ \sigma_{c_{w,z}} & \sigma^2_{c_z} \end{bmatrix}. \qquad [12d]$$

$$[e_{ijk_w}, e_{ijk_z}] \sim MVN(0, \Sigma_e), \Sigma_e = \begin{bmatrix} \sigma^2_{e_w} & \sigma_{e_{w,z}} \\ \sigma_{e_{w,z}} & \sigma^2_{e_z} \end{bmatrix}. \qquad [12e]$$

Where MVN denotes a multivariate normal distribution and Σ are covariance matrices that link common environment *c* and among individual *u* hierarchical levels, and residual effects *e*, to observations. From these matrices we can estimate variances and covariances where $\sigma^2_{u_w}$, $\sigma^2_{c_w}$and $\sigma^2_{e_w}$ are the variances in relative fitness due to differences among individuals, common environments and any residual variance, respectively. $\sigma^2_{u_z}$, $\sigma^2_{c_z}$ and $\sigma^2_{e_z}$ are the variances in the trait due to differences between individuals, between common environments and any variance in the trait left unexplained by the model, respectively. $\sigma_{u_{w,z}}$, $\sigma_{c_{w,z}}$ and $\sigma_{e_{w,z}}$ are the covariances between trait and relative fitness among individuals, common environments and any residual covariance, respectively.

It is important to note here that a regression coefficient is the covariance between the two variables of interest, standardised by the variance of the explanatory variable. As described by (Phillimore et al., 2010), using the estimates of variances and covariances above, we can therefore calculate a regression coefficient (i.e. a directional selection gradient) akin to a Lande-Arnold model as specified below:

$$\beta_z = \frac{\sigma_{w,z}}{\sigma^2_z}. \qquad [13]$$

Total phenotypic covariance is calculated by summing all possible covariances estimated by the model together:

$$\sigma_{w,z} = \sigma_{u_{w,z}} + \sigma_{c_{w,z}} + \sigma_{e_{w,z}}. \qquad [14]$$

Total phenotypic variance in the trait is calculated by summing all traits variances estimated by the model together:

$$\sigma^2_z = \sigma^2_{u_z} + \sigma^2_{c_z} + \sigma^2_{e_z}. \qquad [15]$$

We can estimate decomposed selection gradients using the variances and covariances estimated at each level as specified below:

$$\beta_u = \frac{\sigma_{u_{w,z}}}{\sigma^2_{u_z}}. \qquad [16]$$

$$\beta_c = \frac{\sigma_{c_{w,z}}}{\sigma^2_{c_z}}. \qquad [17]$$

$$\beta_e = \frac{\sigma_{e_{w,z}}}{\sigma^2_{e_z}}. \qquad [18]$$

When selection is soft, we can correct selection estimates β'_z by not including the covariance between relative fitness and the trait among common environments, nor any variation in trait among common environments, when calculating the directional selection gradient as follows (analogous to Equation 8):

$$\beta'_z = \frac{\sigma_{u_{w,z}} + \sigma_{e_{w,z}}}{\sigma^2_{u_z} + \sigma^2_{e_z}}. \qquad [19]$$

4.2 Multi-response animal models

If we have information on relatedness among individuals, we can use a multi-response animal model which involves an additional identity hierarchical level where phenotypic effects are split into genetic and non-genetic effects. This uses relatedness information to estimate breeding values, additive genetic variances and covariances. In this model, the previous hierarchical level associated with among individual effects *u* separates to give us additive genetic *a* and permanent environment *pe* hierarchical levels as below:

$$w_{ijk} \sim \mu_{0_w} + \beta_{w_d} d_{w_i} + a_{j_w} + pe_{j_w} + c_{k_w} + e_{ijk_w}. \qquad [20a]$$

$$z_{ijk} \sim \mu_{0_z} + a_{j_z} + pe_{j_z} + c_{k_z} + e_{ijk_z}. \qquad [20b]$$

From which we can we obtain the following additional variances and covariances

$$\Sigma_a = \begin{bmatrix} \sigma^2_{a_w} & \sigma_{a_{w,z}} \\ \sigma_{a_{w,z}} & \sigma^2_{a_z} \end{bmatrix}. \qquad [20c]$$

$$\Sigma_{pe} = \begin{bmatrix} \sigma^2_{pe_w} & \sigma_{pe_{w,z}} \\ \sigma_{pe_{w,z}} & \sigma^2_{pe_z} \end{bmatrix}. \qquad [20d]$$

β_u can be calculated from this model by summing the additive genetic and permanent environment covariances and variances as follows:

$$\beta_u = \frac{\sigma_{a_{w,z}} + \sigma_{pe_{w,z}}}{\sigma^2_{a_w} + \sigma^2_{pe_z}}. \qquad [21]$$

4.3 Measuring direct selection on multiple traits

One of the major advantages of the Lande-Arnold method is the ability to model selection on multiple traits simultaneously and so estimate direct selection gradients for correlated traits. The multi-response model can be expanded to measure decomposed direct selection gradients on correlated traits as follows:

$$w_{ijk} \sim \mu_{0_w} + \beta_{w_d} d_{w_i} + a_{j_w} + pe_{j_w} + c_{k_w} + e_{ijk_w}. \qquad [22a]$$

$$z_{1_{ijk}} \sim \mu_{0_{z_1}} + a_{j_{z_1}} + pe_{j_{z_1}} + c_{k_{z_1}} + e_{ijk_{z_1}}. \qquad [22b]$$

$$z_{2_{ijk}} \sim \mu_{0_{z_2}} + a_{j_{z_2}} + pe_{j_{z_2}} + c_{k_{z_2}} + e_{ijk_{z_2}}. \qquad [22c]$$

Where $z_1$and z_2 are two different traits measured on the same individual, and *d* is any variable that may affect relative fitness independent of either trait. From these we get the following matrices at different hierarchical levels as follows:

$$\Sigma_u = \begin{bmatrix} \sigma^2_{u_w} & \sigma_{u_{w,z_1}} & \sigma_{u_{w,z_2}} \\ \sigma_{u_{w,z_1}} & \sigma^2_{u_{z_1}} & \sigma_{u_{z_1,z_2}} \\ \sigma_{u_{w,z_2}} & \sigma_{u_{z_1,z_2}} & \sigma^2_{u_{z_2}} \end{bmatrix}. \qquad [23a]$$

$$\Sigma_c = \begin{bmatrix} \sigma^2_{c_w} & \sigma_{c_{w,z_1}} & \sigma_{c_{w,z_2}} \\ \sigma_{c_{w,z_1}} & \sigma^2_{c_{z_1}} & \sigma_{c_{z_1,z_2}} \\ \sigma_{c_{w,z_2}} & \sigma_{c_{z_1,z_2}} & \sigma^2_{c_{z_2}} \end{bmatrix}. \qquad [23b]$$

$$\Sigma_e = \begin{bmatrix} \sigma^2_{e_w} & \sigma_{e_{w,z_1}} & \sigma_{e_{w,z_2}} \\ \sigma_{e_{w,z_1}} & \sigma^2_{e_{z_1}} & \sigma_{e_{z_1,z_2}} \\ \sigma_{e_{w,z_2}} & \sigma_{e_{z_1,z_2}} & \sigma^2_{e_{z_2}} \end{bmatrix}. \qquad [23c]$$

Using methods outlined by (Bernstein, 2009; Froy et al., 2019), we can disentangle direct and indirect selection on traits, and calculate decomposed direct selection gradients by rearranging these variances and covariances among trait z_1, and z_2 , using the among individual direct selection gradients as examples, as follows.

For each hierarchical level, we first rearrange each matrix into two. The first matrix, A, contains the variance and covariance between traits:

$$\Sigma_{u_A} = \begin{bmatrix} \sigma^2_{u_{z_1}} & \sigma_{u_{z_1,z_2}} \\ \sigma_{u_{z_1,z_2}} & \sigma^2_{u_{z_2}} \end{bmatrix}. \qquad [24]$$

The second matrix, B, will contain the covariances between each trait and relative fitness as follows.

$$\Sigma_{u_B} = \begin{bmatrix} \sigma_{\boldsymbol{u}_{w,z_1}} & \sigma_{\boldsymbol{u}_{w,z_2}} \end{bmatrix}. \qquad [25]$$

We then can then estimate decomposed direct selection gradients on z_1 and z_2 by multiplying Σ_{u_B} and the inverse of Σ_{u_A} ($\Sigma_{u_B}\ \Sigma_{u_A}{}^{-1}$), this can then be repeated for each level of the multi-response model specified.

5. Empirical example

5.1 Study Population

To illustrate our approach, we provide an empirical example using individual-based data collected between 1974-2020 from the long-term study of a wild population of red deer (*Cervus elaphus)* living on the Isle of Rum, Scotland to measure selection on male antler weight as proxy for antler size (Kruuk et al., 2002). For more a detailed description on the study population see the supplementary material S3. Briefly, males grow and shed antlers each year and these are used as weapons in male-male contests to control groups of females known as harems. Males with larger antlers are more likely to win fights and mate with more females (Clutton-Brock et al., 1979; Kruuk et al., 2002; Mittell et al., 2024), suggesting a causal relationship between antler size and fitness. Previous analysis suggests strong selection on antler size using the Lande-Arnold method and that antler size had a large heritable component (Kruuk et al., 2002). Selection on antler size on the Rum study population may be soft as males compete over a limited resource (females) and would only need larger antlers than their competitors to outcompete them. Males also regrow antlers each year (Kruuk et al., 2002, Clutton-Brock et al., 1979), and there is evidence that the phenotypic composition of antler mass in the population may change between years (Moyes et al., 2011). Previous studies also found there was no genetic correlation between antler size and breeding success, and that the phenotypic association was being driven mainly by a permanent environment effect (Kruuk, 2004; Kruuk et al., 2014; Kruuk et al., 2002; Mittell et al., 2024). However, these decomposed selection gradients were not directly compared to each other and therefore not explicitly tested for consistency/inconsistency with causality as laid out above and the results were not discussed in the context of soft or hard selection.

For our analyses, we used annual breeding success as our fitness measure, which is the number of calves sired by a male each year as determined by paternity assignments. We then relativised annual breeding success by the global mean, that is, the mean across all years, to obtain our relative fitness measurements. We estimated antler mass by weighing cast antlers on site using an electronic top pan balance accurate to 0.1 g (Kruuk et al., 2002). We then standardised antler weight across years (i.e. the global mean) to a variance of 1 and mean of 0 for analysis. See supplementary material S3 for further details on how our data was collected.

5.2 Statistical analyses

5.2.1 Measuring selection on antler size using the Lande- Arnold method

We ran a linear mixed model as specified in equation [S7], where relative annual breeding success was the response. We fitted standardised antler mass, mean centred age, mean centred age squared and the year annual breeding success was measured as fixed effects. Age and its quadratic term account for variation in annual breeding success due to the ages of males (Kruuk et al., 2014; Kruuk et al., 2002) while year accounts for an increase in the proportion of calves being assigned paternities over the course of the study potentially artificially inflating annual breeding success measures through time (Walling et al., 2010). We included male identity and year as random effects to account for unexplained variation in relative annual breeding success among individuals and years (Kruuk et al., 2014; Kruuk et al., 2002; Mittell et al., 2024). This model, and all models presented here, were implemented using the MCMCglmm v2.36 package (Hadfield, 2010) in R v4.5.0 (R Core Team, 2025).

5.2.2 Measuring selection on antler size using multi-response models

We ran two multi-response models, one where we assumed no pedigree information was available, a common scenario for many wild study populations, and one where we used genetic relatedness calculated from the pedigree to allow for the among individual selection gradient to be further decomposed into a genetic and permanent environment gradients as specified in equations [S8]-[S9]. We included both relative annual breeding success and mean standardised antler weight as responses and fitted the same fixed and random effects as in the previous model. We also fitted male identity as a random effect either once (in models assuming no pedigree information) or twice for animal models, with one linked to a genetic relatedness matrix to estimate both β_a and β_{pe} in models that used relatedness in the form of a pedigree (Kruuk, 2004) (see the supplementary S4 for details). As males grew antlers and participated in the mating season in the same year, we included year as a random effect and allowed a covariance between relative annual breeding and antler mass to be measured to estimate β_c. We then compared the gradients estimated in the same model and determined whether they were statistically different from each other by subtracting their posteriors and calculating the posterior mean difference and their 95% credible intervals (95% CIs).

5.3 Results

We do not discuss the estimates of the fixed effects or random effects variances and covariances in detail here as they are not the focus of the study, but they are reported in Tables S1-S2. As expected, the Lande-Arnold derived estimate of β_z for antler size in the Rum study population was positive (β_z post mean: 0.51, 95% CIs: (0.33 – 0.67), Figure 4, Table S3) with credible intervals that did not overlap zero. Decomposing this selection gradient into among individual, among common environments (years) and within individual selection gradients using a multi-response model with no genetic information suggests that selection on antler size in the Rum study population was consistent with causality and soft. Causality is suggested because both the among and within individual selection gradient were similar in both magnitude and direction to each other, and to the Lande-Arnold derived selection gradients, and were statistically different from zero (β_u: 0.64 (0.39 – 0.94), β_e: 0.46 (0.23 – 0.7), Figure 5, Table S3). However, soft selection is suggested because the among common environment selection gradient was not statistically different from zero and was visually removed from the other gradients (β_c: 0.20 (-0.14 – 0.51), Figure 4, Table S3), implying that there was no relationship between the size of antlers grown in a particular year and the mean relative fitness that year. However, the differences between the posteriors for each gradient do not definitively support this – if selection was soft and consistent with causality, we would expect β_u and β_e not to be statistically different from each other, but for both to be statistically different from β_c. While we found β_u to be statistically different from β_c and not statistically different from β_e, β_c and β_e were not statistically different from each other (Table 1). It is possible that the within individual selection gradient was being biased towards zero by error in the trait measurement (Dingemanse et al., 2021), but we believe this effect to be minimal as antler weight was measured after casting on digital scales accurate to 0.1g, suggesting less scope for error compared to other traits.

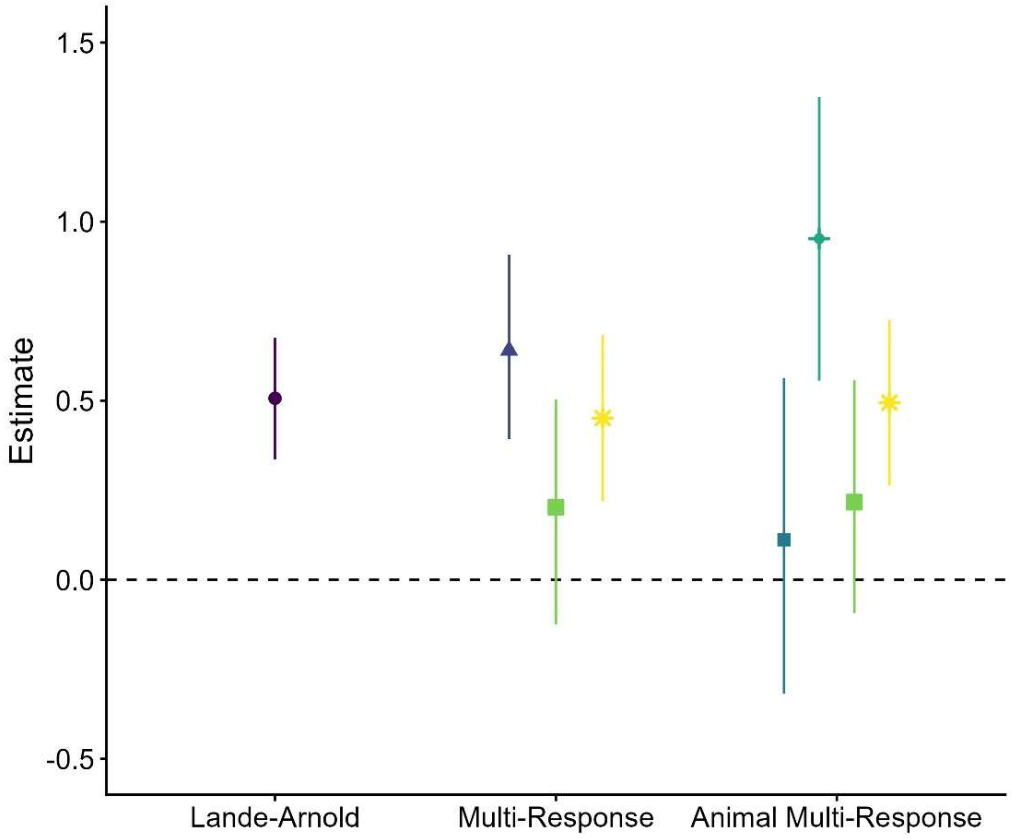

Figure 5: Selection gradient estimates for antler size in male red deer from the Rum study population derived using the Lande-Arnold method, a multi-response model and a multi-response animal model. Where the purple circle is the Lande-Arnold derived selection gradient β_z, the dark blue triangle is the among individual selection gradient β_u, the blue small square is the genetic selection gradient β_a, the teal horizontal dash is the permanent environment selection gradient β_{pe}, the green, large square is the among common

environment (in this case year of measurement) selection gradient β_c and the yellow star is the within individual selection gradient β_e. Points and lines denote the posterior mean estimates and their 95% credible intervals respectively. 95% credible intervals can be used as an indication of whether the gradients are significantly different from 0.

Table 1: The posterior mean differences and associated 95% credible intervals (CIs) between gradients estimated within the multi-response model and within the animal multi-response model. Where β_u is the among individual selection gradient, β_a is the genetic selection gradient, β_{pe} is the permanent selection gradient, β_c is the among common environment selection gradient and β_e is the within individual selection gradient. Gradients that are significantly different from each other (95% CIs do not overlap 0) are highlighted in bold.

Gradients	Difference between posteriors (Mean + 95% CIs)	
	Multi-response Model	Animal Multi-response Model
$\beta_a - \beta_{pe}$	--	**-0.84 (-1.5 – -0.22)**
$\beta_a - \beta_c$	--	-0.11 (-0.7–0.39)
$\beta_a - \beta_e$	--	-0.438(-0.93 – 0.05)
$\beta_{pe} - \beta_c$	--	**0.73 (0.27 –1.19)**
$\beta_{pe} - \beta_e$	--	**0.46 (0.05 –0.92)**
β_u - β_c	**0.44 (0.09 – 0.87)**	--
β_u - β_e	0.19 (-0.15 – 0.51)	--
β_c - β_e	-0.25 (-0.61 – 0.13)	-0.28 (-0.62 – 0.11)

Decomposing the among individual selection gradient into the genetic and permanent environment selection gradients using a multi-response animal model provided further insight. Generally, components that were estimated in both models were similar (Tables S1-S2). Credible intervals were often wider in the animal multi-response model compared to the multi-response model, presumably due to the increase in power required to estimate the genetic components. The genetic selection gradient was not statistically different from zero (β_a: 0.11 (-0.35 – 0.57), Table S3, Figure 5), while the permanent environment selection gradient (β_{pe}: 0.94 (0.57 –1.30), Table S3, Figure 5) was larger than the common environment and residual selection gradients (β_c: 0.22 (-0.16 – 0.52), β_e: 0.50 (0.24 – 0.72), Table S3, Figure 5). Formal comparison of the posterior estimates shows that the permanent environment selection gradient was significantly greater than every other decomposed selection gradient (Table 1), while the genetic selection gradient was not statistically different from the among common environment (Table 1) nor the within individual selection gradient (Table 1).

This pattern suggests two possible scenarios:

1. Selection on antler size was soft because the among common environment selection gradient was not different from 0. It was also consistent with causality because both the residual and the permanent environment gradients were different from 0. However, antler size was genetically correlated with other trait(s) that are selected against, resulting in a genetic selection gradient not different from zero. Additionally, there was a confounding environmental variable that was consistent among individuals and so inflating the permanent environment gradient. In this scenario the Lande-Arnold selection gradient could be correct, due to simultaneously being biased downwards by the genetic selection gradient and upwards by the permanent

environment selection gradient (as found). However, these biases may not be equal and so the Lande-Arnold selection gradient could also be incorrect.

2. Selection on antler size was inconsistent with causality, and the Lande-Arnold selection gradient was being confounded by variables that are consistent between and within individuals, as both the permanent environment selection gradient and the within individual selection gradient respectively were greater than zero and so were biasing the Lande-Arnold selection gradient upwards.

Determining which scenario is more likely would involve careful consideration of the biology of the system and, if possible, attempting to account for potential confounding factors that are consistent among individuals. Parameter estimates could then be compared among models to assess whether the inclusion of potentially confounding variables has reduced the permanent environment selection gradient to be equal in magnitude to the within individual selection gradient or to zero. Considering the Rum red deer population analysed here, anecdotal evidence suggests that any permanent environment confounding could be driven by differences in where males feed outside the study area through winter and spring prior to the mating season; males that rut in our study area are often opportunistically spotted feeding in the same areas in winter/spring each year. However, as these observations are opportunistic and outside the study area, we do not have the necessary data to explicitly test this hypothesis.

If scenario 1 above is correct, and selection on antler size in the Rum study population is consistent with causality and under soft selection, selection estimates from the Lande-Arnold approach would be underestimated, but this can be corrected as detailed above using Eq [21]. Applying this correction to the animal multi-response model, we find that our corrected selection gradient estimate did not differ visually from the uncorrected selection gradient derived from the Lande-Arnold method- as both credible intervals clearly overlap each other (β_z: 0.51 (0.33 – 0.67), β'_z: 0.56 (0.35 –0.75), Figure 6). To confirm this, we calculated the difference between the posteriors of the selection gradient corrected for soft selection, and the uncorrected selection gradient. We found a non-significant difference ($\beta_z - \beta'_z$: -0.11 (-0.23 – 0.006). This is likely because the effect of the among common environment selection gradient is weighted by the among common environment variance in the trait (Eq [7]), and antler mass did not vary much across years ($\sigma^2_{c_z}$,Table S2).

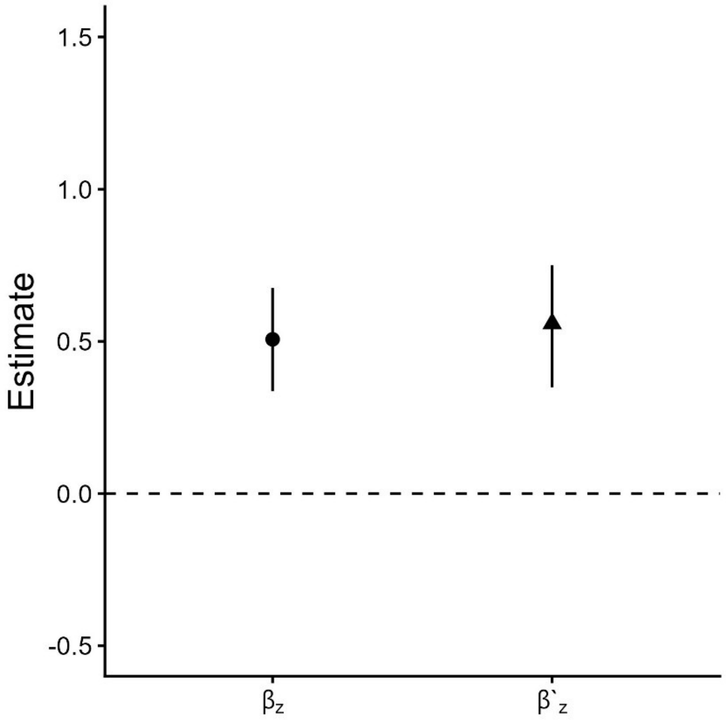

Figure 6: Estimated selection gradients for antler size in male red deer from the Rum study using the Lande-Arnold method which has not been correct for soft selection, (left, β_z),which has been corrected for soft selection, right, β'_z, following equation [10] and using data from the animal multi-response model. Points and lines represent posterior mean estimates and their 95% credible intervals respectively. Note that the credible intervals shown here denote

whether these estimates are statistically different from zero, and do not indicate whether they are statistically different from each other.

6. Discussion

In this article we have described how the Lande-Arnold method of estimating selection based on the phenotypic regression of fitness on a trait assumes that selection is causal (see also Rausher (1992); Stinchcombe et al. (2002); Svensson (2023)), and, when standardising fitness by the global mean, that selection is hard. Importantly we also show that, when these assumptions are not met, selection gradients based on the Lande-Arnold method can be biased. Importantly, we also show that, when these assumptions are not met, selection gradients based on the Lande-Arnold method can be biased. We then outline how, when fitness and a trait are measured repeatedly across individuals and common environments, decomposing selection gradients into their component parts can help infer whether patterns are consistent with causality and hard or soft selection, and to correct selection gradient estimates when selection is soft. We describe how to extract these estimates using multi-response models, and how these models can be expanded to include genetic information and selection on multiple traits. Unlike other methods for determining consistency with causality (Morrissey et al., 2012; Rausher, 1992), when selection is confounded, our method can help identify the source of confounding between trait and fitness.

In our empirical example we used a measure of reproductive success as our fitness measure, but other fitness components such as survival, or metrics that use a combination of both survival and reproductive success such as annual fitness (Bonnet & Postma, 2018), can also be used. While we mostly discuss temporal common environments such as year, we emphasise that other common environments, such as location, can also be included. Our method can also be used to determine consistency with causality not just for selection gradients (i.e. where the response is a fitness measure of an individual), but any relationship between two traits estimated by linear regression if both traits have been measured repeatedly across individuals and/or common environments. For example, Ravindran et al. (2022) decomposed the correlation between annual measures of telomere length and female reproductive performance among and within individuals and among years in a wild population of Soay sheep *Ovis aries* to shed light on the processes underlying their association.

Our method does not need genetic information to be used, although without it confounding via permanent environment and genetic correlations with traits under selection cannot be ruled out. This makes our method suitable to a large number of studies from wild populations where genetic data is not available but have repeated measures of fitness and trait phenotypes across individuals and common environments, where paternity or fitness related data can be inferred from other sources.

While there is considerable scope to apply our methods across a range of different situations, there are some limitations and considerations that should be discussed. There are issues of power in selection analyses, and many studies on natural populations may lack the power required to accurately estimate selection (Hersch & Phillips, 2004). Unfortunately, the multi-response model requires more power than the Lande-Arnold method, as it involves estimating covariances and variances at different hierarchical levels. Estimating the power required for the multi-response model approach is convoluted and dependent on a number of facets - the strength of selection acting on the trait, the number of traits, the number of hierarchical levels specified, the variation of the trait at different hierarchical levels, the direction and strength of correlations between traits at each hierarchical level, the number of groups in each hierarchical level and the number of samples in each group. Estimating the power required for a specific analysis can be done using simulations using packages such as squidSim (Pick, 2025) in R.

We propose that the main limiting factor in wild population studies is likely the number of common environments (e.g. locations or years) sampled. Generally, there are few studies where the number of years or the locations sampled exceed 10 (Hughes et al., 2017; Siepielski et al., 2009), as keeping individual-based wild studies funded long term or across large locations is challenging (Clutton-Brock & Sheldon, 2010; Festa-Bianchet et al., 2017; Pemberton et al., 2022; Sheldon et al., 2022). Additionally, the more complex the models become, i.e. more traits or hierarchical levels included, the more power (i.e. higher sample sizes) that is required.

Despite these issues with power, we argue that decomposing selection gradients is informative, and many studies do have appropriate data. As in our empirical example, qualitative comparison of the magnitude of the different gradients can provide information on likely sources of environmental confounding and potential unaccounted for trait correlations even if statistical comparison is inconclusive as a result of the level of uncertainty in some parameter estimates. Additionally, if soft selection is thought to be likely for a trait, our method can be used to investigate if the data supports soft selection and to correct the selection gradient for soft selection to an extent.

We have not discussed measuring quadratic, correlational or fluctuating selection in much detail in our paper so far. Our framework does not allow for selection to vary across common environments and may produce biased selection gradient estimates if selection does vary (Westneat Paper). The theory we propose behind decomposed selection gradients does apply to decomposed quadratic and correlational selection. However, we do not recommend using multi-response to estimate quadratic and correlational selection gradients, as Dingemanse et al. (2021) showed using simulations that quadratic and correlation selection gradients estimated this way were very imprecise and prone to inaccuracies. Latent-variable models can be used to estimate quadratic, correlational and fluctuating selection (Dingemanse et al., 2021), but they require more complex statistical programming skills to implement. There are a few packages that are currently available for fitting these models, but few tutorials exist on how to fit them properly (Dingemanse et al., 2021). Therefore, for measuring directional selection where selection does not fluctuate, which is the scope of our paper, multi-response models are an easily accessible and implementable option to decompose selection gradients, with many tutorials available to help fit them.

In conclusion, in this article we have highlighted that, when trait and fitness are measured repeatedly across individuals or common environments, the association between a trait and fitness can be decomposed across different hierarchical levels. This approach can provide additional insight into whether the association between a trait and fitness is consistent with a causal relationship and whether selection on a trait is likely to be hard or soft. We demonstrate how this decomposition can be achieved with the readily available and well documented statistical approach of multi-response models. This type of repeated measures data is likely to exist for many wild populations and for associations other than those between traits and fitness. Application of our approach to such data should provide more information about the sources of the correlation between traits and between traits and fitness, which can better inform our conclusions about the ecology of natural selection and the evolution of traits in wild populations.

Bell, D. A., Kovach, R. P., Robinson, Z. L., Whiteley, A. R., & Reed, T. E. (2021). The ecological causes and consequences of hard and soft selection. *Ecology Letters*, *24*(7), 1505-1521. https://doi.org/https://doi.org/10.1111/ele.13754

Bernstein, D. S. (2009). *Matrix mathematics: theory, facts, and formulas*. Princeton university press.

Bonnet, T., & Postma, E. (2018). Fluctuating selection and its (elusive) evolutionary consequences in a wild rodent population. *Journal of Evolutionary Biology*, *31*(4), 572-586. https://doi.org/https://doi.org/10.1111/jeb.13246

Brookfield, J. F. Y. (2016). Why are estimates of the strength and direction of natural selection from wild populations not congruent with observed rates of phenotypic change? *Bioessays*, *38*(9), 927-934. https://doi.org/10.1002/bies.201600017

Caruso, C. M., Martin, R. A., Sletvold, N., Morrissey, M. B., Wade, M. J., Augustine, K. E., Carlson, S. M., MacColl, A. D. C., Siepielski, A. M., & Kingsolver, J. G. (2017). What Are the Environmental Determinants of Phenotypic Selection? A Meta-analysis of Experimental Studies. *American Naturalist*, *190*(3), 363-376. https://doi.org/10.1086/692760

Charmantier, A., Burkhard, T., Gervais, L., Perrier, C., Schulte-Hostedde, A. I., & Thompson, M. J. (2024). How does urbanization affect natural selection? *Functional Ecology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/1365-2435.14667

Charmantier, A., Garant, D., & Kruuk, L. E. (2014). *Quantitative genetics in the wild*. OUP Oxford.

Clements, M. N., Clutton-Brock, T. H., Guinness, F. E., Pemberton, J. M., & Kruuk, L. E. B. (2011). VARIANCES AND COVARIANCES OF PHENOLOGICAL TRAITS IN A WILD MAMMAL POPULATION. *Evolution*, *65*(3), 788-801. https://doi.org/10.1111/j.1558-5646.2010.01161.x

Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, *25*(10), 562-573. https://doi.org/10.1016/j.tree.2010.08.002

Clutton-Brock, T. H., Albon, S. D., Gibson, R. M., & Guinness, F. E. (1979). LOGICAL STAG - ADAPTIVE ASPECTS OF FIGHTING IN RED DEER (CERVUS-ELAPHUS L). *Animal Behaviour*, *27*(FEB), 211-225. https://doi.org/10.1016/0003-3472(79)90141-6

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray.

de Villemereuil, P., Morrissey, M. B., Nakagawa, S., & Schielzeth, H. (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: issues and solutions. *Journal of Evolutionary Biology*, *31*(4), 621-632. https://doi.org/10.1111/jeb.13232

Dingemanse, N. J., Araya-Ajoy, Y. G., & Westneat, D. F. (2021). Most published selection gradients are underestimated: Why this is and how to fix it. *Evolution*, *75*(4), 806-818. https://doi.org/https://doi.org/10.1111/evo.14198

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39-54. https://doi.org/https://doi.org/10.1111/1365-2656.12013

Dobson, S., Dunning, J., Burke, T., Chik, H. Y. J., & Schroeder, J. (2023). Indirect genetic effects increase heritability estimates for male and female extra-pair reproduction. *Evolution*, *77*(8), 1893-1901. https://doi.org/10.1093/evolut/qpad100

Falconer, D., & Mackay, T. (1996). Introduction to Quantitative Genetics (Harlow, UK: Longman Group Ltd.).

Festa-Bianchet, M., Douhard, M., Gaillard, J.-M., & Pelletier, F. (2017). Successes and challenges of long-term field studies of marked ungulates. *Journal of Mammalogy*, *98*(3), 612-620.

Froy, H., Sparks, A. M., Watt, K., Sinclair, R., Bach, F., Pilkington, J. G., Pemberton, J. M., McNeilly, T. N., & Nussey, D. H. (2019). Senescence in immunity against helminth parasites predicts adult mortality in a wild mammal. *Science*, *365*(6459), 1296-1298. https://doi.org/doi:10.1126/science.aaw5822

Gauzere, J., Walling, C. A., Pick, J. L., Watt, K., Jack, P., Morris, A., Morris, S., & Pemberton, J. M. (2021). The role of maternally transferred antibodies in maternal performance in red deer. *Ecology Letters*, *24*(10), 2065-2076. https://doi.org/https://doi.org/10.1111/ele.13834

Hadfield, J. D. (2008). Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 723-734. https://doi.org/10.1098/rspb.2007.1013

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, *33*(2), 1-22. https://doi.org/10.18637/jss.v033.i02

Hadfield, J. D., Heap, E. A., Bayer, F., Mittell, E. A., & Crouch, N. M. A. (2013). DISENTANGLING GENETIC AND PRENATAL SOURCES OF FAMILIAL RESEMBLANCE ACROSS ONTOGENY IN A WILD PASSERINE. *Evolution*, *67*(9), 2701-2713. https://doi.org/https://doi.org/10.1111/evo.12144

Hansen, T. F., & Houle, D. (2004). Evolvability, stabilizing selection, and the problem of stasis. *Phenotypic integration*, 130-152.

Hersch, E. I., & Phillips, P. C. (2004). Power and Potential Bias in Field Studies of Natural Selection. *Evolution*, *58*(3), 479-485. http://www.jstor.org/stable/3449241

Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., Beerli, P., & Kingsolver, J. G. (2001). Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences*, *98*(16), 9157-9160. https://doi.org/doi:10.1073/pnas.161281098

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse of BLUP in behavioural ecology. *Behavioral Ecology*, *28*(4), 948-952. https://doi.org/10.1093/beheco/arx023

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., de Nesnera, K. L., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., . . . Carr, M. H. (2017). Long-Term Studies Contribute Disproportionately to Ecology and Policy. *BioScience*, *67*(3), 271-281. https://doi.org/10.1093/biosci/biw185

Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., & Carlson, S. M. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, *26*, 1101-1118.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., & Beerli, P. (2001). The Strength of Phenotypic Selection in Natural Populations. *The American Naturalist*, *157*(3), 245-261. https://doi.org/10.1086/319193

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the ‘animal model’. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1446), 873-890. https://doi.org/doi:10.1098/rstb.2003.1437

Kruuk, L. E. B., Clutton-Brock, T., & Pemberton, J. M. (2014). *Case study: quantitative genetics and sexual selection of weaponry in a wild ungulate*. <Go to ISI>://WOS:000354449000011

Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F., & Clutton-Brock, T. (2002). ANTLER SIZE IN RED DEER: HERITABILITY AND SELECTION BUT NO EVOLUTION. *Evolution*, *56*(8), 1683-1695. https://doi.org/10.1111/j.0014-3820.2002.tb01480.x

Lande, R., & Arnold, S. J. (1983). THE MEASUREMENT OF SELECTION ON CORRELATED CHARACTERS. *Evolution*, *37*(6), 1210-1226. https://doi.org/10.2307/2408842

Merilä, J., Sheldon, B. C., & Kruuk, L. E. B. (2001). Explaining stasis: microevolutionary studies in natural populations. *Genetica*, *112*(1), 199-222. https://doi.org/10.1023/A:1013391806317

Mitchell-Olds, T., & Shaw, R. G. (1987). REGRESSION ANALYSIS OF NATURAL SELECTION: STATISTICAL INFERENCE AND BIOLOGICAL INTERPRETATION. *Evolution*, *41*(6), 1149-1161. https://doi.org/https://doi.org/10.1111/j.1558-5646.1987.tb02457.x

Mittell, E. A., Mandaliya, P., Pemberton, J. M., Morris, A., Morris, S., Johnston, S. E., & Kruuk, L. E. B. (2024). Antler size in red deer: declining selection and increasing genetic variance with age, but little senescence. *Journal of Evolutionary Biology*. https://doi.org/10.1093/jeb/voae112

Morrissey, M. B. (2014). SELECTION AND EVOLUTION OF CAUSALLY COVARYING TRAITS. *Evolution*, *68*(6), 1748-1761. https://doi.org/10.1111/evo.12385

Morrissey, M. B., & Hadfield, J. D. (2012). DIRECTIONAL SELECTION IN TEMPORALLY REPLICATED STUDIES IS REMARKABLY CONSISTENT. *Evolution*, *66*(2), 435-442. https://doi.org/10.1111/j.1558-5646.2011.01444.x

Morrissey, M. B., & Henshaw, J. M. (2022). Phenotypic selection analysis and confounding environmental variables. *bioRxiv*, 2022.2006.2015.496257. https://doi.org/10.1101/2022.06.15.496257

Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, *23*(11), 2277-2288. https://doi.org/10.1111/j.1420-9101.2010.02084.x

Morrissey, M. B., Parker, D. J., Korsten, P., Pemberton, J. M., Kruuk, L. E. B., & Wilson, A. J. (2012). THE PREDICTION OF ADAPTIVE EVOLUTION: EMPIRICAL APPLICATION OF THE SECONDARY THEOREM OF SELECTION AND COMPARISON TO THE BREEDER'S EQUATION. *Evolution*, *66*(8), 2399-2410. https://doi.org/10.1111/j.1558-5646.2012.01632.x

Moyes, K., Nussey, D. H., Clements, M. N., Guinness, F. E., Morris, A., Morris, S., Pemberton, J. M., Kruuk, L. E. B., & Clutton-Brock, T. H. (2011). Advancing breeding phenology in response to environmental change in a wild red deer population. *Global Change Biology*, *17*(7), 2455-2469. https://doi.org/10.1111/j.1365-2486.2010.02382.x

Pemberton, J. M. (2010). Evolution of quantitative traits in the wild: mind the ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2431-2438. https://doi.org/doi:10.1098/rstb.2010.0108

Pemberton, J. M., Kruuk, L. E. B., & Clutton-Brock, T. (2022). The Unusual Value of Long-Term Studies of Individuals: The Example of the Isle of Rum Red Deer Project. *Annual Review of Ecology Evolution and Systematics*, *53*, 327-351. https://doi.org/10.1146/annurev-ecolsys-012722-024041

Phillimore, A. B., Hadfield, J. D., Jones, O. R., & Smithers, R. J. (2010). Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences*, *107*(18), 8292-8297. https://doi.org/doi:10.1073/pnas.0913792107

Pick, J., Allegue, H.,. (2025). *squidSim: Flexible and reproducible structured simulations*. In (Version R package version 0.2.3 https://github.com/squidgroup/squidSim)

Pick, J. L., Ebneter, C., Hutter, P., & Tschirren, B. (2016). Disentangling Genetic and Prenatal Maternal Effects on Offspring Size and Survival. *The American Naturalist*, *188*(6), 628-639. https://doi.org/10.1086/688918

Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P., Roux, F., Scotti, I., Teplitsky, C., Thomson, C. E., & Winney, I. (2018). The Missing Response to Selection in the Wild. *Trends in Ecology & Evolution*, *33*(5), 337-346. https://doi.org/10.1016/j.tree.2018.02.007

R Core Team. (2025). *R: A Language and Environment for Statistical Computing.* . In R Foundation for Statistical Computing. <https://www.R-project.org/>

Rausher, M. D. (1992). THE MEASUREMENT OF SELECTION ON QUANTITATIVE TRAITS: BIASES DUE TO ENVIRONMENTAL COVARIANCES BETWEEN TRAITS AND FITNESS. *Evolution*, *46*(3), 616-626. https://doi.org/10.1111/j.1558-5646.1992.tb02070.x

Ravindran, S., Froy, H., Underwood, S. L., Dorrens, J., Seeker, L. A., Watt, K., Wilbourn, R. V., Pilkington, J. G., Harrington, L., Pemberton, J. M., & Nussey, D. H. (2022). The association between female reproductive performance and leukocyte telomere length in wild Soay sheep. *Molecular Ecology*, *31*(23), 6184-6196. https://doi.org/https://doi.org/10.1111/mec.16175

Reid, J. M. (2012). Predicting evolutionary responses to selection on polyandry in the wild: additive genetic covariances with female extra-pair reproduction. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1747), 4652-4660. https://doi.org/doi:10.1098/rspb.2012.1835

Reid, J. M., & Wolak, M. E. (2018). Is there indirect selection on female extra-pair reproduction through cross-sex genetic correlations with male reproductive fitness? *Evolution Letters*, *2*(3), 159-168. https://doi.org/10.1002/evl3.56

Sheldon, B., Kruuk, L. E. B., & Alberts, S. C. (2022). The expanding value of long-term studies of individuals in the wild. *Nature Ecology & Evolution*, *6*(12), 1799-1801. https://doi.org/10.1038/s41559-022-01940-7

Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, *12*(11), 1261-1276. https://doi.org/10.1111/j.1461-0248.2009.01381.x

Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. A., Arnold, S. J., & Blows, M. W. (2008). ESTIMATING NONLINEAR SELECTION GRADIENTS USING QUADRATIC REGRESSION COEFFICIENTS: DOUBLE OR NOTHING? *Evolution*, *62*(9), 2435-2440. https://doi.org/https://doi.org/10.1111/j.1558-5646.2008.00449.x

Stinchcombe, John R., Rutter, Matthew T., Burdick, Donald S., Tiffin, P., Rausher, Mark D., & Mauricio, R. (2002). Testing for Environmentally Induced Bias in Phenotypic Estimates of Natural Selection: Theory and Practice. . *The American Naturalist*, *160*(4), 511-523. https://doi.org/10.1086/342069

Svensson, E. I. (2023). Phenotypic selection in natural populations: what have we learned in 40 years? *Evolution*, *77*(7), 1493-1504. https://doi.org/10.1093/evolut/qpad077

Thomson, C. E., Bayer, F., Crouch, N., Farrell, S., Heap, E., Mittell, E., Zurita-Cassinello, M., & Hadfield, J. D. (2017). Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. *Evolution*, *71*(3), 716-732. https://doi.org/https://doi.org/10.1111/evo.13169

Wallace, B. (1975). Hard and soft selection revisited. *Evolution*, 465-473.

Wallace, B., & Lewontin, R. (1968). Population biology and evolution. In (pp. 87-108): Syracuse University Press.

Walling, C. A., Pemberton, J. M., Hadfield, J. D., & Kruuk, L. E. B. (2010). Comparing parentage inference software: reanalysis of a red deer pedigree. *Molecular Ecology*, *19*(9), 1914-1928. https://doi.org/10.1111/j.1365-294X.2010.04604.x

Supplementary materials

S1 Derivation of equation [9]

The Lande-Arnold selection gradient β_z can be estimated by dividing the total phenotypic covariance between trait and relative fitness ($\sigma_{w,z}$) by the total phenotypic variance of the trait (σ_z^2) as follows.

$$\beta_z = \frac{\sigma_{w,z}}{\sigma_z^2} \qquad [S1]$$

Where $\sigma_{w,z}$ is composed of the covariance between relative fitness and trait among individuals ($\sigma_{u_{w,z}}$), among common environments ($\sigma_{c_{w,z}}$) and within individuals ($\sigma_{e_{w,z}}$).

$$\sigma_{w,z} = \sigma_{u_{w,z}} + \sigma_{c_{w,z}} + \sigma_{e_{w,z}}. \qquad [S2]$$

σ_z^2 is composed of the variance of the trait among individuals ($\sigma_{u_z}^2$), among common environments ($\sigma_{c_z}^2$) and within individuals ($\sigma_{e_z}^2$).

$$\sigma_z^2 = \sigma_{u_z}^2 + \sigma_{c_z}^2 + \sigma_{e_z}^2 \qquad [S3]$$

Therefore, decomposed selection gradients β_u, β_c and β_e are calculated by dividing the decomposed covariances by their respective decomposed trait variances as follows:

$$\beta_u = \frac{\sigma_{u_{w,z}}}{\sigma_{u_z}^2} \qquad [s4a]$$

$$\beta_c = \frac{\sigma_{c_{w,z}}}{\sigma_{c_z}^2} \qquad [s4b]$$

$$\beta_e = \frac{\sigma_{e_{w,z}}}{\sigma_{e_z}^2} \qquad [s4c]$$

We rearrange and insert equation [S1] into [S2] to get:

$$\beta_z \sigma_z^2 = \sigma_{u_{w,z}} + \sigma_{c_{w,z}} + \sigma_{e_{w,z}} \qquad [s5]$$

Which we derive to get:

$$\beta_z = \frac{\sigma_{u_{w,z}} + \sigma_{c_{w,z}} + \sigma_{e_{w,z}}}{\sigma_z^2} \qquad [S6]$$

We then rearrange and insert equations [S4a – c] into equation [S6a] as follows:

$$\beta_z = \frac{\beta_u \sigma_{u_z}^2 + \beta_c \sigma_{c_z}^2 + \beta_e \sigma_{e_z}^2}{\sigma_z^2} \qquad [S6a]$$

$$\beta_z = \beta_u \frac{\sigma_{u_z}^2}{\sigma_z^2} + \beta_c \frac{\sigma_{c_z}^2}{\sigma_z^2} + \beta_e \frac{\sigma_{e_z}^2}{\sigma_z^2} \qquad [S6b]$$

S2 How different ways of standardising fitness and traits affects selection estimates depending on whether selection on the trait is soft or hard

When explaining how fitness is relativised and traits are standardised when measuring selection, Lande and Arnold (1983) stated that fitness should be relativised by the "mean of the population". Lande and Arnold were considering a single population at a single time point, which allows no way to distinguish hard and soft selection and makes this definition reasonably clear in this context. However, this definition is somewhat vague if a population is measured across multiple environments or time points. Under these scenarios, what constitutes a population becomes more ambiguous and may be context dependent, making hard and soft selection more distinct mechanisms in this context. Importantly, this can result in different ways of relativising fitness and standardising traits that can introduce unintentional biases and error into estimates of selection.

Studies measuring selection can standardise fitness and trait phenotypes either by the global mean of the data (as described in the main text) or by the mean within common environments (e.g., within each year in populations with annual breeding cycles). A review by (De Lisle & Svensson, 2017) found that most studies measuring fitness and traits across several environments or 'groups' tended to mostly standardise fitness and traits 'globally' without much justification. As mentioned in the main text, if fitness and trait phenotypes are standardised by the global mean selection may be underestimated if selection is soft. A way of dealing with this would be to standardise fitness and trait phenotypes within common environments (ie. within groups) prior to analysis with the Lande-Arnold method. This would theoretically correct Lande-Arnold derived selection gradients for soft selection because individuals compete with other individuals that exist in the same space and/or time, and this approach relativises their fitness and trait phenotype to the mean of the individuals they compete against. Indeed, this is how (De Lisle & Svensson, 2017) recommend to standardise fitness and trait measurements when soft selection is occurring. However, relativising fitness and trait phenotypes within common environments implicitly assumes soft selection is occurring across those common environments, which can lead to erroneous selection gradients if that is not the case. It can also be difficult to understand in advance of the analysis whether selection on a trait is likely to be soft or hard. For example, when a population occurs across different areas with different environments, it can be difficult to determine how connected these subpopulations are (De Lisle & Svensson, 2017). If soft selection on a trait is occurring and subpopulations are completely isolated, selection would appear soft, but increasing connectivity is likely to give soft selection the appearance of hard selection. For this scenario, relativising fitness and trait phenotypes within common environments would result in the selection gradient being misestimated. Additionally, common environment fitness and trait means are likely to be erroneous, particularly if samples sizes within common environments are small, which may cause underestimation of the error in the means within environments, and therefore, the selection gradient estimates (Lüdtke et al., 2008). Plotting the relationship between a trait and absolute fitness separately for each level of the common environment to determine whether soft or hard selection is occurring may also lead to spurious conclusions being made, as these associations are not tested statistically.

Aside from our method, whether selection is hard or soft can also be explicitly tested for by expanding the Lande-Arnold method to include the within common environment trait mean as a covariate and comparing the slopes (i.e. contextual analysis)(Goodnight et al., 1992), or by including within common environment mean and an individual’s deviation from the mean and comparing their slopes (i.e. within and among group centring) (Kreft et al., 1995;

Westneat et al., 2020). It is worth noting that these methods are reparameterizations of each other and are consistent with our approach of decomposing selection gradients. However, the statistical implementation of the contextual analysis/ within and among group centring approach differs from our approach, as they suffer from the issues mentioned above when samples sizes within common environments are small.

Our approach of decomposing selection gradients into different hierarchical levels using multi-response models allows for hard vs soft selection to be explicitly tested in a statistically robust framework, while accounting for uncertainty in estimation of the means of common environments with small sample sizes. Our method also simultaneously assesses consistency with causality and allows selection gradient estimates to be corrected when selection is soft without any of the drawbacks associated without having to first assume soft or hard selection is occurring.

S3 Additional information on measurements used in our empirical example of selection on antler size in male red deer

In the Rum red deer study population, the identity of almost every individual is known from natural markings, collars, ear tags and punches (Pemberton et al., 2022). Females are sexually receptive for 1-2 days each year during the mating season (September – October) and give birth May-June of the following year. Male reproductive success is primarily determined via precopulatory male-male competition during the mating season, where males will attract females into their harem, which they then mate with once they come into oestrus. Males guard their harem against other males, and contests frequently escalate into dangerous fights to establish dominance and possession of females (Clutton-Brock et al., 1979). We expect selection on antler size to be consistent with causality because males use their antlers when fighting to gain access to females and defend their harems, and males with larger antlers are likely to win more fights and mate more females (Clutton-Brock et al., 1979;

Kruuk et al., 2002; Mittell et al., 2024). Selection on antler size is likely be soft as males would only need bigger antlers compared to their opponents be able to win fights, and the limiting number of niches to be occupied would be the ratio of females to males during the mating season (Clutton-Brock et al., 1979; Kruuk et al., 2002).

Antler size has been a major focus in the Rum study population, with much already investigated in terms of both phenology (Clements et al., 2010; Moyes et al., 2011) and its association with male fitness (Clutton-Brock et al., 1979; Kruuk et al., 2014; Kruuk et al., 2002), making it an ideal trait to demonstrate our methods. Antlers start to grow in the spring, encased in blood supplying tissue called velvet which is rubbed off in summer (Clements et al., 2010). Antlers are then cast off the following spring and the cycle restarts (Clements et al., 2010). Cast antlers are collected and measured, and the unique shape of the antler is used to identify which individuals they belonged to (Kruuk et al., 2002).

The paternity of calves born each year were determined via tissue samples collected at birth (Guinness et al., 1982; Pemberton et al., 2022), which were then used to infer male annual breeding success measures as the number of calves males were assigned paternity to each year. The age of males born in the study area were calculated from their date of birth, whereas the ages of immigrant males who were born outside the study area but rutted within the study area were estimated from appearance (Walling et al., 2010). We included males aged 5 and over in our analysis as males are considered to be sexually mature at this age (Pemberton et al., 2022). We then mean centre male age by subtracting mean male age from each male age.

S4 Analysis specifications for our empirical example

Our data comprised of 818 observations from 246 male deer across 44 years. The pedigree used for the animal multi-response model was pruned down to 564 individuals who either had phenotypic data available or where related to individuals who had phenotypic data.

We estimated the Lande-Arnold derived selection gradient for antler size using the equation specified below:

$$w_{ijk}\sim \mu_0 + \ \beta_z z_{ijk} + \beta_m m_{ijk} + \gamma_m m_{ijk}^2 + \beta_b b_{ijk} + \ u_j + c_k + e_{ijk} \qquad [S7]$$

Where w is the relative annual breeding success of a male, μ_0 is the global intercept for relative fitness, z is male antler weight standardised to a variance of 1 and mean of 0, m is the mean centred age of males, m^2 is the mean centred age of males squared, *b* is the year in which both trait and fitness were measured, u, c and *e* are identity and year random effects and any residual variation in *w* not explained by the model. β_z, β_m, γ_m and β_b are the effects of z_{ijk}, m, m^2 and *b* on w respectively, and β_z is the Lande-Arnold derived directional selection gradient.

We estimated decomposed selection gradients for antler size with no genetic information using the equations specified below:

$$w_{ijk}\sim \mu_{w_0} + \beta_{w_m} m_{w_{ijk}} + \gamma_{w_m} m_{w_{ijk}}^2 + \beta_{w_b} b_{w_{ijk}} + u_{w_j} + c_{w_k} + e_{w_{ijk}} \qquad [S8a]$$

$$z_{ijk}\sim \mu_{z_0} + u_{z_j} + c_{z_k} + e_{z_{ijk}} \qquad [S8b]$$

Where $\mu_{w_0}$is the global intercept for relative annual breeding success, β_{w_m}, $Y_{w_m}$and β_{w_m} are the effects of m, m^2 and b on W and u_w, c_w and e_w are identity and year random effects for *w* and any residual variation in *w* not explained by the model respectively. μ_{z_0} is the global intercept for standardised antler weight, and u_z, c_z and e_z are identity and year random effects for *z* and any residual variation in *z* not explained by the model respectively.

We estimated decomposed selection gradients for antler size including genetic information using the equations specified below:

$$w_{ijk} \sim \mu_{w_0} + \beta_{w_m} m_{w_i} + \gamma_{w_m} m_{w_i}^2 + \beta_{w_b} b_{w_i} + a_{w_j} + pe_{w_j} + c_{w_j} + e_{w_{ijk}} \qquad [S9a]$$

$$z_{ijk} \sim \mu_{z_0} + a_{z_j} + pe_{z_j} + c_{z_k} + e_{z_{ijk}} \qquad [S9b]$$

Where a_w and pe_w are genetic and permanent random effects for w and a_z and pe_z are genetic and permanent random effects for z.

We then followed procedures outlined in section 3.1 to estimate variances-covariances matrices and eventually the decomposed selection gradients from both multi-response models.

S5 Additional results and model outputs from our empirical example

Table S1: Fixed effects estimated from models of selection on antler in the rum study population, including a univariate (Lande-Arnold derived) model and multi-response models both with and without genetic information. Effects are the global intercept for relative breeding success (μ_0 or μ_{w_0}), the global intercept for standardised antler size (μ_{z_0}), the effect of mean-centred age (β_m or β_{w_m}), the quadratic effect of mean-centred age (γ_m or γ_{w_m}) and the effect of year (β_b or β_{w_b}). Estimates are posterior means and 95% credible intervals. Terms are highlighted in bold when pMCMC values are lower than 0.05. Note we do not include estimates of β_z from equation [S1] as these are included in table S3.

	Lande-Arnold derived		Multi-response no genetic information		Multi-response with genetic information	
Outputs	Post.mean (+95 cis)	pMCMC	Post.mean (+95 cis)	pMCMC	Mean (+95 CIs)	pMCMC
Intercept- relative fitness μ_0 or μ_{w_0}	-24.2(-49.9 – 1.72)	0.07	**-44.3 (-80.4 – -8.9)**	**0.001**	-37.4 (-78.9 – 0.74)	0.06
Intercept - Trait μ_{z_0}	--	--	0.02 (-0.18 – 0.24)	0.851	0.11 (-0.15 – 0.35)	0.39
Age- relative fitness β_m or β_{w_m}	**-0.13 (-0.19 – -0.05)**	**0.0008**	**-0.12 (-0.19 – -0.04)**	**0.002**	**-0.13 (-0.2 -0.05)**	**0.002**
Age^2- relative fitness γ_m or γ_{w_m}	**-0.04 (-0.06 – -0.03)**	**0.0008**	**-0.04 (-0.06 – -0.03)**	**0.0008**	**-0.04 (-0.06 – -0.03)**	**0.0008**
Year- relative fitness β_b or β_{w_b}	0.013 (-1.83e-04 – 0.03)	0.06	**0.02 (0.005 – 0.04)**	**0.014**	0.02 (2.4e-04 – 0.04)	0.06

Table S2: Random effect estimates from models of the Lande Arnold derived selection gradient (univariate model) and multi-response models without and with genetic information. Outputs include the variance of relative breeding success among individuals (σ_u^2 or $\sigma_{u_w}^2$), permanent environment ($\sigma_{pe_w}^2$), additive genetic variance ($\sigma_{a_w}^2$), among years (σ_c^2 or $\sigma_{c_w}^2$) and within individuals (σ_e^2 or $\sigma_{e_w}^2$); the variance of standardised antler weight among individuals ($\sigma_{u_z}^2$), permanent environment ($\sigma_{pe_z}^2$), additive genetic variance ($\sigma_{a_z}^2$), among years ($\sigma_{c_z}^2$) and within individuals ($\sigma_{e_z}^2$); and the covariance between relative breeding success and standardised antler size among individuals ($\sigma_{u_{w,z}}$), permanent environment ($\sigma_{pe_{w,z}}$), additive genetic variance ($\sigma_{a_{w,z}}$), among years ($\sigma_{c_{w,z}}$) and within individuals ($\sigma_{e_{w,c}}$). Estimates are posterior means with 95% credible intervals in parentheses.

	Lande-Arnold derived	Multi-response no genetic information	Multi-response with genetic information
Relative breeding success – Individual σ_u^2 or $\sigma_{u_w}^2$	0.75 (0.55 – 0.96)	0.94 (0.67 – 1.22)	--
Relative breeding success – Permanent Environment $\sigma_{pe_w}^2$	--	--	0.8 (0.51 – 1.07)
Relative breeding success – Genetic $\sigma_{a_w}^2$	--	--	0.26 (0.12 – 0.43)
Relative breeding success – Year σ_c^2 or $\sigma_{c_w}^2$	0.07 (0.01 – 0.12)	0.18 (0.09 – 0.29)	0.19 (0.09 – 0.32)
Relative breeding success – Residual σ_e^2 or $\sigma_{e_w}^2$	1.01 (0.89 – 1.13)	1.09 (0.96 – 1.25)	1.1 (0.96 – 1.26)
Antler size – Individual $\sigma_{u_z}^2$	--	0.46 (0.34 -0.59)	--
Antler size- Permanent Environment $\sigma_{pe_z}^2$	--	--	0.32 (0.19 – 0.44)
Antler size- Genetic $\sigma_{a_z}^2$	--	--	0.23 (0.12 – 0.34)
Antler size – Year $\sigma_{c_z}^2$	--	0.41 (0.24 – 0.66)	0.45 (0.25 – 0.73)
Antler size – Residual $\sigma_{e_z}^2$	--	0.37 (0.32 –0.42)	0.36 (0.31 – 0.41)

Covariance – Individual $\sigma_{u_{w,z}}$	--	0.3 (0.14 – 0.43)	--
Covariance – Genetic $\sigma_{a_{w,z}}$	--	--	0.03 (-0.08 – 0.14)
Covariance – Permanent Environment $\sigma_{pe_{w,z}}$	--	--	0.29 (0.16 – 0.45)
Covariance – Year $\sigma_{c_{w,z}}$	--	0.08 (-0.06 – 0.23)	0.1 (-0.08 – 0.25)
Covariance – Residual $\sigma_{e_{w,c}}$	--	0.17 (0.08 – 0.26)	0.17 (0.09 –0.27)

Table S3: selection gradients for antler size in the Rum study population, derived from a univariate (Land-Arnold method) model, a multi-response model with no genetic term and a multi-response model with a genetic term. Outputs are the Lande-Arnold derived selection gradient or the equivalent of a Lande-Arnold derived selection gradient estimated form a multi-response model (β_z), the among individual selection gradient (β_u), the genetic selection gradient (β_a), the permanent environment selection gradient (β_{pe}), the among year selection gradient (β_c) and the within individual selection gradient (β_e). Estimates are extracted from posterior means and 95% credible intervals.

Selection gradient	Lande-Arnold derived	Multi-response no genetic information	Multi-response with genetic information
Lande-Arnold derived/ equivalent β_z	0.51 (0.33 –0.67)	0.44 (0.27 – 0.64)	0.45 (0.27 – 0.62)
Among individual β_u	--	0.64 (0.39 –0.94)	--
Genetic β_a	--	--	0.45 (0.27 – 0.62)
Permanent environment β_{pe}	--	--	0.94 (0.57 –1.30)
Among common environment (year) β_c	--	0.20 (-0.14 –0.51)	0.22 (-0.16 –0.52)
Within individual β_e	--	0.46 (0.23 –0.7)	0.50 (0.24 – 0.72)

Bell, D. A., Kovach, R. P., Robinson, Z. L., Whiteley, A. R., & Reed, T. E. (2021). The ecological causes and consequences of hard and soft selection. *Ecology Letters*, *24*(7), 1505-1521. https://doi.org/https://doi.org/10.1111/ele.13754

Bernstein, D. S. (2009). *Matrix mathematics: theory, facts, and formulas*. Princeton university press.

Bonnet, T., & Postma, E. (2018). Fluctuating selection and its (elusive) evolutionary consequences in a wild rodent population. *Journal of Evolutionary Biology*, *31*(4), 572-586. https://doi.org/https://doi.org/10.1111/jeb.13246

Brookfield, J. F. Y. (2016). Why are estimates of the strength and direction of natural selection from wild populations not congruent with observed rates of phenotypic change? *Bioessays*, *38*(9), 927-934. https://doi.org/10.1002/bies.201600017

Caruso, C. M., Martin, R. A., Sletvold, N., Morrissey, M. B., Wade, M. J., Augustine, K. E., Carlson, S. M., MacColl, A. D. C., Siepielski, A. M., & Kingsolver, J. G. (2017). What Are the Environmental Determinants of Phenotypic Selection? A Meta-analysis of Experimental Studies. *American Naturalist*, *190*(3), 363-376. https://doi.org/10.1086/692760

Charmantier, A., Burkhard, T., Gervais, L., Perrier, C., Schulte-Hostedde, A. I., & Thompson, M. J. (2024). How does urbanization affect natural selection? *Functional Ecology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/1365-2435.14667

Charmantier, A., Garant, D., & Kruuk, L. E. (2014). *Quantitative genetics in the wild*. OUP Oxford.

Clements, M. N., Clutton-Brock, T. H., Albon, S. D., Pemberton, J. M., & Kruuk, L. E. B. (2010). Getting the timing right: antler growth phenology and sexual selection in a wild red deer population. *Oecologia*, *164*(2), 357-368. https://doi.org/10.1007/s00442-010-1656-7

Clements, M. N., Clutton-Brock, T. H., Guinness, F. E., Pemberton, J. M., & Kruuk, L. E. B. (2011). VARIANCES AND COVARIANCES OF PHENOLOGICAL TRAITS IN A WILD MAMMAL POPULATION. *Evolution*, *65*(3), 788-801. https://doi.org/10.1111/j.1558-5646.2010.01161.x

Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, *25*(10), 562-573. https://doi.org/10.1016/j.tree.2010.08.002

Clutton-Brock, T. H., Albon, S. D., Gibson, R. M., & Guinness, F. E. (1979). LOGICAL STAG - ADAPTIVE ASPECTS OF FIGHTING IN RED DEER (CERVUS-ELAPHUS L). *Animal Behaviour*, *27*(FEB), 211-225. https://doi.org/10.1016/0003-3472(79)90141-6

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray.

De Lisle, S. P., & Svensson, E. I. (2017). On the standardization of fitness and traits in comparative studies of phenotypic selection. *Evolution*, *71*(10), 2313-2326. https://doi.org/10.1111/evo.13325

de Villemereuil, P., Morrissey, M. B., Nakagawa, S., & Schielzeth, H. (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: issues and solutions. *Journal of Evolutionary Biology*, *31*(4), 621-632. https://doi.org/10.1111/jeb.13232

Dingemanse, N. J., Araya-Ajoy, Y. G., & Westneat, D. F. (2021). Most published selection gradients are underestimated: Why this is and how to fix it. *Evolution*, *75*(4), 806-818. https://doi.org/https://doi.org/10.1111/evo.14198

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39-54. https://doi.org/https://doi.org/10.1111/1365-2656.12013

Dobson, S., Dunning, J., Burke, T., Chik, H. Y. J., & Schroeder, J. (2023). Indirect genetic effects increase heritability estimates for male and female extra-pair reproduction. *Evolution*, *77*(8), 1893-1901. https://doi.org/10.1093/evolut/qpad100

Falconer, D., & Mackay, T. (1996). Introduction to Quantitative Genetics (Harlow, UK: Longman Group Ltd.).

Festa-Bianchet, M., Douhard, M., Gaillard, J.-M., & Pelletier, F. (2017). Successes and challenges of long-term field studies of marked ungulates. *Journal of Mammalogy*, *98*(3), 612-620.

Froy, H., Sparks, A. M., Watt, K., Sinclair, R., Bach, F., Pilkington, J. G., Pemberton, J. M., McNeilly, T. N., & Nussey, D. H. (2019). Senescence in immunity against helminth parasites predicts adult mortality in a wild mammal. *Science*, *365*(6459), 1296-1298. https://doi.org/doi:10.1126/science.aaw5822

Gauzere, J., Walling, C. A., Pick, J. L., Watt, K., Jack, P., Morris, A., Morris, S., & Pemberton, J. M. (2021). The role of maternally transferred antibodies in maternal performance in red deer. *Ecology Letters*, *24*(10), 2065-2076. https://doi.org/https://doi.org/10.1111/ele.13834

Goodnight, C. J., Schwartz, J. M., & Stevens, L. (1992). CONTEXTUAL ANALYSIS OF MODELS OF GROUP SELECTION, SOFT SELECTION, HARD SELECTION, AND THE EVOLUTION OF ALTRUISM. *American Naturalist*, *140*(5), 743-761. https://doi.org/10.1086/285438

Guinness, F., Albon, S., & Clutton-Brock, T. (1982). *Red deer: behavior and ecology of two sexes*. Edinburgh University Press.

Hadfield, J. D. (2008). Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 723-734. https://doi.org/10.1098/rspb.2007.1013

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, *33*(2), 1-22. https://doi.org/10.18637/jss.v033.i02

Hadfield, J. D., Heap, E. A., Bayer, F., Mittell, E. A., & Crouch, N. M. A. (2013). DISENTANGLING GENETIC AND PRENATAL SOURCES OF FAMILIAL RESEMBLANCE ACROSS ONTOGENY IN A WILD PASSERINE. *Evolution*, *67*(9), 2701-2713. https://doi.org/https://doi.org/10.1111/evo.12144

Hansen, T. F., & Houle, D. (2004). Evolvability, stabilizing selection, and the problem of stasis. *Phenotypic integration*, 130-152.

Hersch, E. I., & Phillips, P. C. (2004). Power and Potential Bias in Field Studies of Natural Selection. *Evolution*, *58*(3), 479-485. http://www.jstor.org/stable/3449241

Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., Beerli, P., & Kingsolver, J. G. (2001). Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences*, *98*(16), 9157-9160. https://doi.org/doi:10.1073/pnas.161281098

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse of BLUP in behavioural ecology. *Behavioral Ecology*, *28*(4), 948-952. https://doi.org/10.1093/beheco/arx023

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., de Nesnera, K. L., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., . . . Carr, M. H. (2017). Long-Term Studies Contribute Disproportionately to Ecology and Policy. *BioScience*, *67*(3), 271-281. https://doi.org/10.1093/biosci/biw185

Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., & Carlson, S. M. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, *26*, 1101-1118.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., & Beerli, P. (2001). The Strength of Phenotypic Selection in Natural Populations. *The American Naturalist*, *157*(3), 245-261. https://doi.org/10.1086/319193

Kreft, I. G. G., Deleeuw, J., & Aiken, L. S. (1995). THE EFFECT OF DIFFERENT FORMS OF CENTERING IN HIERARCHICAL LINEAR-MODELS. *Multivariate Behavioral Research*, *30*(1), 1-21. https://doi.org/10.1207/s15327906mbr3001_1

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the ‘animal model’. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1446), 873-890. https://doi.org/doi:10.1098/rstb.2003.1437

Kruuk, L. E. B., Clutton-Brock, T., & Pemberton, J. M. (2014). *Case study: quantitative genetics and sexual selection of weaponry in a wild ungulate*. <Go to ISI>://WOS:000354449000011

Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F., & Clutton-Brock, T. (2002). ANTLER SIZE IN RED DEER: HERITABILITY AND SELECTION BUT NO EVOLUTION. *Evolution*, *56*(8), 1683-1695. https://doi.org/10.1111/j.0014-3820.2002.tb01480.x

Lande, R., & Arnold, S. J. (1983). THE MEASUREMENT OF SELECTION ON CORRELATED CHARACTERS. *Evolution*, *37*(6), 1210-1226. https://doi.org/10.2307/2408842

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229. https://doi.org/10.1037/a0012869

Merilä, J., Sheldon, B. C., & Kruuk, L. E. B. (2001). Explaining stasis: microevolutionary studies in natural populations. *Genetica*, *112*(1), 199-222. https://doi.org/10.1023/A:1013391806317

Mitchell-Olds, T., & Shaw, R. G. (1987). REGRESSION ANALYSIS OF NATURAL SELECTION: STATISTICAL INFERENCE AND BIOLOGICAL INTERPRETATION. *Evolution*, *41*(6), 1149-1161. https://doi.org/https://doi.org/10.1111/j.1558-5646.1987.tb02457.x

Mittell, E. A., Mandaliya, P., Pemberton, J. M., Morris, A., Morris, S., Johnston, S. E., & Kruuk, L. E. B. (2024). Antler size in red deer: declining selection and increasing genetic variance with age, but little senescence. *Journal of Evolutionary Biology*. https://doi.org/10.1093/jeb/voae112

Morrissey, M. B. (2014). SELECTION AND EVOLUTION OF CAUSALLY COVARYING TRAITS. *Evolution*, *68*(6), 1748-1761. https://doi.org/10.1111/evo.12385

Morrissey, M. B., & Hadfield, J. D. (2012). DIRECTIONAL SELECTION IN TEMPORALLY REPLICATED STUDIES IS REMARKABLY CONSISTENT. *Evolution*, *66*(2), 435-442. https://doi.org/10.1111/j.1558-5646.2011.01444.x

Morrissey, M. B., & Henshaw, J. M. (2022). Phenotypic selection analysis and confounding environmental variables. *bioRxiv*, 2022.2006.2015.496257. https://doi.org/10.1101/2022.06.15.496257

Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, *23*(11), 2277-2288. https://doi.org/10.1111/j.1420-9101.2010.02084.x

Morrissey, M. B., Parker, D. J., Korsten, P., Pemberton, J. M., Kruuk, L. E. B., & Wilson, A. J. (2012). THE PREDICTION OF ADAPTIVE EVOLUTION: EMPIRICAL APPLICATION OF THE SECONDARY THEOREM OF SELECTION AND COMPARISON TO THE BREEDER'S EQUATION. *Evolution*, *66*(8), 2399-2410. https://doi.org/10.1111/j.1558-5646.2012.01632.x

Moyes, K., Nussey, D. H., Clements, M. N., Guinness, F. E., Morris, A., Morris, S., Pemberton, J. M., Kruuk, L. E. B., & Clutton-Brock, T. H. (2011). Advancing breeding phenology in response to environmental change in a wild red deer population. *Global Change Biology*, *17*(7), 2455-2469. https://doi.org/10.1111/j.1365-2486.2010.02382.x

Pemberton, J. M. (2010). Evolution of quantitative traits in the wild: mind the ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2431-2438. https://doi.org/doi:10.1098/rstb.2010.0108

Pemberton, J. M., Kruuk, L. E. B., & Clutton-Brock, T. (2022). The Unusual Value of Long-Term Studies of Individuals: The Example of the Isle of Rum Red Deer Project. *Annual Review of Ecology Evolution and Systematics*, *53*, 327-351. https://doi.org/10.1146/annurev-ecolsys-012722-024041

Phillimore, A. B., Hadfield, J. D., Jones, O. R., & Smithers, R. J. (2010). Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences*, *107*(18), 8292-8297. https://doi.org/doi:10.1073/pnas.0913792107

Pick, J., Allegue, H.,. (2025). *squidSim: Flexible and reproducible structured simulations*. In (Version R package version 0.2.3 https://github.com/squidgroup/squidSim)

Pick, J. L., Ebneter, C., Hutter, P., & Tschirren, B. (2016). Disentangling Genetic and Prenatal Maternal Effects on Offspring Size and Survival. *The American Naturalist*, *188*(6), 628-639. https://doi.org/10.1086/688918

Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P., Roux, F., Scotti, I., Teplitsky, C., Thomson, C. E., & Winney, I. (2018). The Missing Response to Selection in the Wild. *Trends in Ecology & Evolution*, *33*(5), 337-346. https://doi.org/10.1016/j.tree.2018.02.007

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. . In R Foundation for Statistical Computing. <https://www.R-project.org/>

Rausher, M. D. (1992). THE MEASUREMENT OF SELECTION ON QUANTITATIVE TRAITS: BIASES DUE TO ENVIRONMENTAL COVARIANCES BETWEEN TRAITS AND FITNESS. *Evolution*, *46*(3), 616-626. https://doi.org/10.1111/j.1558-5646.1992.tb02070.x

Ravindran, S., Froy, H., Underwood, S. L., Dorrens, J., Seeker, L. A., Watt, K., Wilbourn, R. V., Pilkington, J. G., Harrington, L., Pemberton, J. M., & Nussey, D. H. (2022). The association between female reproductive performance and leukocyte telomere length in wild Soay sheep. *Molecular Ecology*, *31*(23), 6184-6196. https://doi.org/https://doi.org/10.1111/mec.16175

Reid, J. M. (2012). Predicting evolutionary responses to selection on polyandry in the wild: additive genetic covariances with female extra-pair reproduction. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1747), 4652-4660. https://doi.org/doi:10.1098/rspb.2012.1835

Reid, J. M., & Wolak, M. E. (2018). Is there indirect selection on female extra-pair reproduction through cross-sex genetic correlations with male reproductive fitness? *Evolution Letters*, *2*(3), 159-168. https://doi.org/10.1002/evl3.56

Sheldon, B., Kruuk, L. E. B., & Alberts, S. C. (2022). The expanding value of long-term studies of individuals in the wild. *Nature Ecology & Evolution*, *6*(12), 1799-1801. https://doi.org/10.1038/s41559-022-01940-7

Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, *12*(11), 1261-1276. https://doi.org/10.1111/j.1461-0248.2009.01381.x

Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. A., Arnold, S. J., & Blows, M. W. (2008). ESTIMATING NONLINEAR SELECTION GRADIENTS USING QUADRATIC REGRESSION COEFFICIENTS: DOUBLE OR NOTHING? *Evolution*, *62*(9), 2435-2440. https://doi.org/https://doi.org/10.1111/j.1558-5646.2008.00449.x

Stinchcombe, John R., Rutter, Matthew T., Burdick, Donald S., Tiffin, P., Rausher, Mark D., & Mauricio, R. (2002). Testing for Environmentally Induced Bias in Phenotypic Estimates of Natural Selection: Theory and Practice. . *The American Naturalist*, *160*(4), 511-523. https://doi.org/10.1086/342069

Svensson, E. I. (2023). Phenotypic selection in natural populations: what have we learned in 40 years? *Evolution*, *77*(7), 1493-1504. https://doi.org/10.1093/evolut/qpad077

Thomson, C. E., Bayer, F., Crouch, N., Farrell, S., Heap, E., Mittell, E., Zurita-Cassinello, M., & Hadfield, J. D. (2017). Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. *Evolution*, *71*(3), 716-732. https://doi.org/https://doi.org/10.1111/evo.13169

Wallace, B. (1975). Hard and soft selection revisited. *Evolution*, 465-473.

Wallace, B., & Lewontin, R. (1968). Population biology and evolution. In (pp. 87-108): Syracuse University Press.

Walling, C. A., Pemberton, J. M., Hadfield, J. D., & Kruuk, L. E. B. (2010). Comparing parentage inference software: reanalysis of a red deer pedigree. *Molecular Ecology*, *19*(9), 1914-1928. https://doi.org/10.1111/j.1365-294X.2010.04604.x

Westneat, D. F., Araya-Ajoy, Y. G., Allegue, H., Class, B., Dingemanse, N., Dochtermann, N. A., Garamszegi, L. Z., Martin, J. G. A., Nakagawa, S., Réale, D., & Schielzeth, H. (2020). Collision between biological process and statistical analysis revealed by mean centring.

Journal of Animal Ecology, *89*(12), 2813-2824. https://doi.org/https://doi.org/10.1111/1365-2656.13360