20

Environmental Effects and the Response in Natural Populations

Associations between phenotype and fitness, however appealing, will give a misleading impression of the potential for evolution in a trait if the true target of selection is unmeasured or immeasurable. — Kruuk et al. 2002

Version 27 June 2013

Under artificial selection (animal/plant breeding and laboratory selection experiments), the breeder's equation machinery developed in Chapters 13 - 19 for the prediction and analysis of response generally works well. However, there is considerable angst as to whether this is also true for natural populations (e.g., Merilä et al. 2001c; Pemberton 2010; Morrissey et al. 2010, 2012). There are two principal reasons for this. First, under artificial selection, individuals are strictly chosen by the phenotypic value of their focal trait. By contrast, in natural populations, one *infers* the target of selection, typically by looking for changes in the mean and/or variance of certain candidate traits. The problem is that the moments of an unselected trait can change if it is correlated with another trait under selection. A within-generation change (a nonzero selection differential) occurs if an unselected trait is phenotypically correlated to a selected one (Equation 13.25c), while a between-generation change (a response) occurs if they are genetically correlated (Equation 13.26c).

The second complication is lack of control over the environment. With artificial selection, there is generally considerable environmental control, in part due to husbandry/cultivation methods designed to standardize rearing/growing conditions. This is certainly not the case when attempting to track selection in natural populations. Indeed, we are largely ignorant of which environmental factors may be important, much less being able to control them. Further, artificial selection experiments generally impose considerable control over the *biotic* (as opposed to the physical) environment in which an organism finds itself. In natural populations, the biotic environment is both absolutely critical and largely uncontrollable. One potential consequence of lack of environment control are fitness-trait correlations (and hence a selection differential) caused by unmeasured environmental factors. Changes in the environment can also mask underlying genetic changes and can lead to significant changes in the nature of selection from one generation to the next. Finally, changes in the environment can modify genetic and/or environmental variances, altering the heritability.

This chapter addresses these concerns in two parts. The first is largely theoretical, centering on the bias caused by selection on unmeasured features. We initially frame this within the context of correlated characters, and then focus on the special case where an unmeasured variable is entirely environmental, which can generate a selection different but no response. This leads the classic analysis by Price et al. (1988) on the apparent lack of response to selection for earlier breeding dates in birds. Next, we by extend the univariate breeder's equation to account for all of the possible biases that arise from any unmeasured traits influencing the focal trait. We conclude by recasting response under Robertson's secondary theorem (Chapter 6), $R = \sigma(A_z, A_w)$ or $\sigma(A_z, w)$, as opposed to the breeder's equation $R = h^2 S$ framework. Contrasts between these two predictions suggests tests for accessing whether a focal trait is

the sole target of selection.

The second part of our treatment is largely empirical, examining the advantages, and pitfalls, of applying mixed models (Chapter 19) to natural populations. During the first decade of 2000, there was much excitement that BLUP predictions of breeding values would offer powerful insight into the nature of response in natural populations. A rash of results, many initially viewed as classic, quickly appeared from the analysis of several pedigreed vertebrate populations under long-term observation (reviewed in Kruuk et al. 2008). However, problems with using BLUP estimates of breeding values in natural settings were initially suggested by Postma (2006) and then shown to be often fatal by Hadfield (2008) and Hadfield et al. (2010). Hence, many of these initial results need to be seriously reconsidered. However, certain aspects of the animal model, in particular REML estimates of specific covariances (such as that between the breeding values of a trait and fitness), remain powerful approaches. We first review how a BLUP analysis in a natural population proceeds (with a specific focus on estimating the relationship matrix $\mathbf{A}$), and then show where flaws can appear, and finally examine how certain features of the animal model can still prove useful. The rapid rise, even quicker demise, and then phoenix-like reorientation of animal model applications in natural populations can be quite confusing to the novice reading the historical literature, so we try to carefully navigate the reader through the shoals of confusion. We conclude by reviewing a number of examples of response (or lack thereof) in natural populations, using the developed theoretical and statistical machinery to highlight problems that arise when attempting to predict response in a natural population.

RESPONSE IN NATURAL POPULATIONS: WHAT IS THE TRAGET OF SELECTION?

While there are many assumptions underlying the breeder's equation (Chapter 6, Table 13.2), the core one that is both most likely to fail in natural populations, and is often the most challenging to test, is that of *causality* – the phenotype of the focal trait is the sole target of selection in the sense that it is genetically and phenotypically uncorrelated with any other traits under selection. Trait value z entirely governs fitness and transmission of the resulting change in z to the next generation is entirely given by h^2. The conceptual beauty of the breeder's equation is that it partitions evolution into separate and distinct ecological (S) and genetical (h^2) processes, allowing ecologists to focus on the former (forces of selection) and geneticists on the later (inheritance of the trait). If we incorrectly assign the target of selection, the breeder's equation will give misleading results.

Direct and Correlated Responses

Bias from correlated traits can be removed by using the multivariate breeder's equation. This expresses the vector $\mathbf{R}$ of responses (changes in means) as a function of the genetic (breeding value) $\mathbf{G}$ and phenotypic $\mathbf{P}$ covariance matrices for the traits of interest and the vector $\mathbf{S}$ of their selection differentials. From Equations 13.23b and 13.25b,

$$\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S} = \mathbf{G}\,\boldsymbol{\beta}$$

where the selection gradient $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$ controls for any phenotypic correlations among the *measured* traits, returning the amount of direct selection acting on each particular character (LW Chapter 8; Chapter 29). Using the bivariate version of this equation provides insight into some of the complications than can arise by incorrectly identifying the target traits(s) by ignoring selection on correlated traits. Suppose we are following a trait z_1, influenced by a second (and unmeasured) trait z_2. Noting that $\mathbf{S} = \mathbf{P}\boldsymbol{\beta}$, the selection differential on trait 1 becomes

$$S_1 = P_{11}\beta_1 + P_{12}\beta_2 = \sigma^2(z_1)\beta_1 + \sigma(z_1, z_2)\beta_2 \tag{20.1a}$$

A within-generation change ($S_1 \neq 0$) in trait one occurs from (i) direct selection on trait 1 ($\beta_1 \neq 0$) and/or (ii) indirect selection from a *phenotypically* correlated trait under directional selection ($\beta_2 \neq 0$ *and* $\sigma(z_1, z_2) \neq 0$). As a result, the signs of S_1 and β_1 can differ, and even strong direct selection ($\beta_1 \neq 0$) on a trait can still result in a net selection differential of nearly zero. Turning to the expected response in trait 1,

$$R_1 = \Delta\mu_1 = G_{11}\beta_1 + G_{12}\beta_2 = \sigma^2(A_1)\beta_1 + \sigma(A_1, A_2)\beta_2 \qquad (20.1b)$$

Again, trait 1 can change from direct selection (given it has heritable variation) and/or a correlated response from direct selection on another *genetically* correlated trait (the breeding values of the two traits are correlated within individuals). The use of selection gradients results in conceptually, and notationally, elegant expressions for response. As Equation 13.24c shows, this expression is more cumbersome when given in terms of the observed selection differentials, S_1 and S_2.

Example 20.1. Alatalo et al. (1989) examined tarsus length in a population of collared flycatchers (*Ficedula albicollis*) residing in the southern part of the island of Gotland in the Baltic Sea. Measurements of lifetime fitness on this isolated bird population were possible since most surviving offspring (which are tagged before **fledging**, i.e., leaving the nest) return to breed in the area they were reared as offspring. In addition to tarsus length, fledging weight was also measured and a Pearson-Lande-Arnold regression (Chapter 29; LW Chapter 8) performed to compute the amounts of direct selection (the estimated selection gradients) on both characters, yielding

Year	Observed S on tarsus length	Estimated selection gradients, $\widehat{\beta}$ tarsus length	fledging weight
1981	0.19**	0.01	0.25*
1983	0.08	-0.01	0.21*
1984	0.20**	0.12	0.33***
1985	0.02	-0.06	0.27***
pooled	0.12**	0.03	0.27***

$* = p < 0.05$, $** = p < 0.01$, $*** = p < 0.001$

Although there is a significant selection differential on tarsus length in two of the years (and in the pooled data), there is no significant direct selection on tarsus length itself (estimated gradients not significantly different from zero). Rather, direct selection is on fledging weight. While there is a significant phenotypic correlation between tarsus length and fledging weight ($r = 0.32$, $p < 0.001$), it appears to be entirely due to within-individual correlations of environmental effects, as there is no correlation between offspring weight and parental tarsus length ($r = -0.01$, $p > 0.1$). Hence, the observed selection on tarsus length is a consequence of selection on fledging weight, which has no genetic correlation with tarsus length and hence no response in tarsus length is expected.

As this example highlights, some of the best studies of the response to selection in natural populations come from birds (reviewed by Merilä and Sheldon 2001, Merilä et al. 2001b, Kruuk et al. 2008, Gienapp et al. 2008). In certain setting (such as isolated islands), the entire population can be banded and all nests located (often through the use of nest boxes), allowing for accurate measurement of individual fitness (Chapter 28).

Example 20.2. As reviewed in Grant and Grant (1995, 2002, and references therein), one of the best documented cases of natural selection is on body size and bill morphology in Darwin's finches (*Geospiza fortis*) on the Galápagos island of Daphne Major. Two strong episodes of selection were observed during their long-term study, due to serious droughts in 1976–1977 (where the population crashed from 634 birds down to 95, a 15% survival rate), and in 1984–1986 (556 birds down to 180, a 32% survival rate). Six morphological traits were followed through both episodes, and (after rescaling all traits to have unit variances), the following selection differentials S and gradients β for the two episodes were as follows (where $* = p < 0.05$):

	1976 - 1977		1984 - 1986	
Trait	S	β	S	β
Weight	0.74^*	0.477^*	-0.11	-0.040
Wing length	0.72^*	0.436^*	-0.08	-0.015
Tarsus length	0.43^*	0.005	-0.09	-0.047
Bill length	0.54^*	-0.144	-0.03	0.245^*
Bill depth	0.63^*	0.528^*	-0.16^*	-0.135
Bill width	0.53^*	-0.450^*	-0.17^*	-0.152

Two striking features are apparent. First, the observed (within-generation) change in mean (S) is not a good predictor of the actual amount of direct selection (β) on a trait, and can even be of the wrong sign (e.g., bill length). Second, the nature of selection changed over the two drought periods. During the 1976 – 1977 drought, larger individuals were favored, and there was selection on bill shape (increased bill depth while decreasing bill width). A change in the dominant food supply during a subsequent drought from 1984 – 1986 resulted in selection favoring smaller birds. Hence, the two episodes of selection were in opposite directions (at least in terms of body size).

Grant and Grant used the multivariate breeder's equation to examine how well responses were predicted (given below as means following response). Response was well predicted in 1976, but over-predicted in three of the six traits in the 1984 episode. Grant and Grant suggest that the main reason for these discrepancies was a change in the biotic environment. Higher population densities for offspring in 1984 retarded growth, resulting in an over-prediction of size-related response.

	1976-1977		1984-1986	
Character	Predicted	Observed	Predicted	Observed
Weight	17.39 ± 0.22	17.52 ± 0.25	16.82 ± 0.13	$15.48 \pm 0.08^*$
Wing length	69.98 ± 0.39	69.65 ± 0.35	67.93 ± 0.17	$67.21 \pm 0.11^{***}$
Tarsus length	19.45 ± 0.09	19.32 ± 0.14	19.02 ± 0.04	19.02 ± 0.04
Bill length	11.14 ± 0.10	11.06 ± 0.11	10.86 ± 0.05	10.96 ± 0.03
Bill depth	9.83 ± 0.12	9.94 ± 0.09	9.51 ± 0.06	$9.32 \pm 0.03^{**}$
Bill width	8.96 ± 0.08	8.97 ± 0.08	8.77 ± 0.04	8.70 ± 0.03

where $* = p < 0.05$, $** = p < 0.01$, $*** = p < 0.001$

Environmental Correlations Between Fitness and Traits

In natural populations, an environmental factor (or factors) can influence both an individual's trait value and their fitness. This generates a correlation between the trait and fitness, and hence a selection differential on a trait $S = \sigma(z, w)$, even if there is no direct selection on

the trait itself. Consider the following example, suggested by Rauscher (1992). Suppose soil nitrate concentration influences both fitness (seed production) and the amount of alkaloids (secondary plant chemicals) in the foliage of a plant. As Figure 20.1 shows, if we are able to partition individuals from a population into high and low nitrate environments, within each group we would find no association between alkaloid concentration and fitness. However, if we ignore this partition and simply consider all individuals as a single group, there is a positive covariance between alkaloid concentration and fitness. An investigator unaware of this difference in soil nitrates might conclude a fitness effect from the presence of alkaloids (for example, as an insect deterrent), when in fact the correlation between trait and fitness arises solely because both are influenced by a third, and unmeasured, variable. In this setting, the breeder's equation assumes that alkaloid levels are the causative agent of fitness differences, rather than the actual environmental agent (nitrate levels), which is ignored.

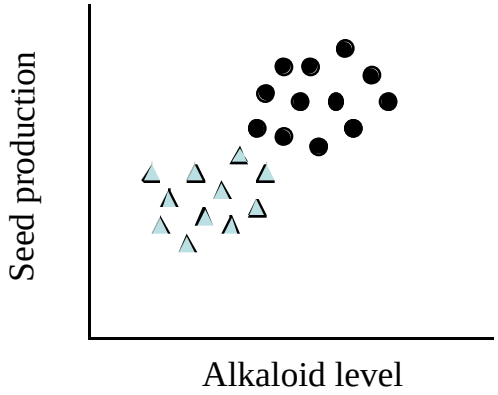


Figure 20.1. An environmental variable (soil nitrate) influences both fitness and trait value (alkaloid levels), creating a covariance between the trait and fitness, when in fact the trait value is not causal to fitness. In low nitrate soils (triangles), plants have low fitness and low levels of alkaloids. In high nitrate soils (circles), plants have high fitness and high levels of alkaloids. Within each of the two environments, there is no association between trait and fitness. If one ignores the environmental effects and simply lumps all individuals together, there is a strong association between fitness and the trait value. Figure based on Mauricio and Mojonnier (1997) and Rausher (1992).

Rausher's example introduces the concept of selection on a **non-heritable environmental component** of a trait. While soil nitrate is the primary determinant of plant fitness, it is a non-heritable environmental trait. While plants with high alkaloid levels (which are heritable) have higher fitness, alkaloid levels do not increase in the next generation. Plants with high *environmental* values for alkaloids were selected, not plants with high *genetic* values, and hence no response. Put another way, there is selection on the *environmental* value of the trait, *not* on its breeding value.

Example 20.3. Price and Liou (1989), considering the evolution of **clutch size** (z_1) in birds (number of eggs laid in a particular episode), suggested that fitness is largely determined by the nutritional state z_2 of a mother, which also influences her clutch size. They assumed this nutritional state can be treated as a non-heritable environmental factor. Equation 20.1a

implies that, even if there is no direct selection on clutch size *per se* ($\beta_1 = 0$), we would still observe a selection differential on clutch size if it is phenotypically-correlated with nutritional state which is itself under selection ($\beta_2 \neq 0$), as

$$S_1 = \beta_2\, \sigma(z_1, z_2) \neq 0$$

However, if there is no additive-genetic variance in the nutritional state, then $\sigma^2(A_2) = 0$ and hence $\sigma(A_1, A_2) = 0$. Thus, in the absence of direct selection on clutch size ($\beta_1 = 0$), any response in clutch size arises from a correlated response on nutritional state, with Equation 20.1b giving the response as

$$\Delta\mu_1 = \beta_2\, \sigma(A_1, A_2),$$

This contribution is zero because $\sigma(A_1, A_2)$ is zero since z_2 has no heritable variance. The result is apparent directional selection on clutch size ($S_1 \neq 0$), but no response ($R_1 = 0$).

The notion of some nutritional status, or other well-being measure, of an organism is often referred to as its **condition** by ecologists (Le Cren 1951). This term is often used fairly loosely without any formal definition. One common operational measure is the residual of a regression of weight on some measure of body size (i.e., size-adjusted weight). The motivation for this metric is that individuals in good condition will be heavier than expected given their size, while individuals in poor condition will be underweight. Jakob et al. (1996), Green (2001), and Schulte-Hostedde et al. (2005) discuss the merits of this metric. While condition is often treated entirely as a product of the environment, as with any standard quantitative trait, it is reasonable to assume it also has some genetic component as well (e.g., Gosler and Harper 2000, Merilä et al. 2001a).

THE FISHER-PRICE-KIRKPATRICK-ARNOLD MODEL FOR EVOLUTION OF BREEDING DATE IN BIRDS

The idea of a nonheritable environmental factor being the target of selection dates back to Fisher and Darwin. Fisher (1958), based on observations by Darwin (1871), suggested that the condition of a bird influences both her clutch size and the date at which she breeds, with healthier females breeding earlier and having larger clutch sizes. Price, Kirkpatrick and Arnold (1988) used Fisher's idea as an explanation for the apparent lack of selection response for breeding date in many birds in the temperate zone. Birds that reproduce early have higher fitness than those who breed later in the season, and hence S for breeding date is negative (selection to move the breeding date earlier). Further, where examined, breeding date has moderate to high heritability. Since both h^2 and S are nonzero, the breeder's equation suggests that we expect a response to selection resulting in a decrease in breeding date, but this is not seem.

The model of Price et al. (1988) is shown in Figure 20.2. For brevity, we refer to the **Fisher-Price-Kirkpatrick-Arnold model** as simply **Fisher's model** (we will resist the temptation of referring to this as the Fisher-Price toy model). The model is as follows: assume that the breeding date z of an female has three components,

$$z = A - n + e \tag{20.2}$$

A is the breeding value for date, e the environmental value, and n the nutritional state (or condition) of the female. Equation 20.2 shows that females with a higher value of n (higher nutritional status) breed earlier. Price et al. (following Fisher) treat n as a nonheritable

component of the environment, but one could also model n as a heritable trait, changing this to a multivariate selection problem. The three components of Equation 20.2 are assumed uncorrelated and normally distributed with variances σ_A^2, σ_n^2, and σ_e^2. Let μ be the current mean breeding value, while we take the mean of n and e to both be zero.

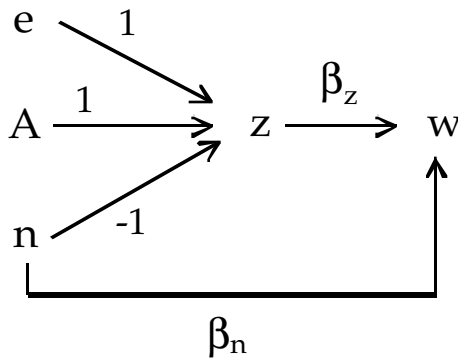


Figure 20.2. A path diagram (LW Appendix 2) of the components in the Fisher-Price-Kirkpatrick-Arnold model, showing the connections between breeding data z, nutritional state n, and fitness w. The breeding value A, general environmental value e, and nutritional state n all influence breeding date z, which itself influences fitness (path coefficient β_z). Likewise, there is a second path to fitness directly from nutritional state (β_n) which represents the direct contribution to w once the contribution of n through breeding date z is removed. We assume A, e, and n are all uncorrelated, and hence not connected by any paths. (After Price et al. 1988.)

Price et al. impose selection by considering two separate components of fitness. First, they assume there is an optimum breeding date θ, so that z is under stabilizing selection. Recall from Equation 16.17 that a standard model for stabilizing selection in natural populations is nor-optimal or normalizing selection (Weldon 1895, Haldane 1954),

$$W(z) = \exp\left(-\frac{(z-\theta)^2}{2\,\omega^2} \right) \tag{20.3a}$$

This fitness function, giving the expected fitness of an individual with phenotypic value z, has the same form as a normal distribution, with the highest fitness at the optimal phenotypic value $z = \theta$, declining as one moves away from θ. The strength of selection is given by ω^2, the "variance" of the fitness function. If $\omega^2 \gg \sigma_z^2$, selection is weak, while if $\omega^2 \ll \sigma_z^2$ selection is strong. One advantage of this fitness function is that if z is normally distributed before selection, it remains normally distributed following selection, and expressions for the new mean and variance are easily obtained (Equation 16.18a). Second, Price et al. assumed that fitness increases with the nutritional value n. One way to express this is to assume

$$W(n) = \exp(\alpha n), \quad \text{for } \alpha > 0 \tag{20.3b}$$

Note that if $|\,\alpha\, n\,| \ll 1$, then $W(n)$ is approximately $1 + \alpha n$. As with the nor-optimal fitness function, traits normally-distributed before selection remain normal following selection. The resulting joint fitness given both breeding date z and nutritional value n is

$$W(n,z) = W(z) \cdot W(n) = \exp\left(\alpha n - \frac{(z-\theta)^2}{2\,\omega^2} \right) \tag{20.4a}$$

Expressed in terms of the components of our model,

$$W(n, e, A) = \exp\left(\alpha n - \frac{(A - n + e - \theta)^2}{2\,\omega^2}\right) \tag{20.4b}$$

Assuming A is normally distributed with mean μ, while n and e are normally distributed with mean zero, Heywood (2005) gives the change in mean as

$$R = \sigma_A^2 \left(\frac{\theta - \mu + \alpha\sigma_n^2}{\omega^2 + \sigma_z^2}\right) \tag{20.5}$$

At equilibrium ($R = 0$), the mean breeding date is

$$\widehat{\mu} = \theta + \alpha\sigma_n^2, \tag{20.6}$$

which is later than the optimal breeding date θ. Price et al. comment that this displacement beyond θ occurs because females in good nutritional condition are constrained (Equation 20.2) to breed earlier than the mean breeding value at equilibrium. Since $\widehat{\mu} > \theta$, females in good nutritional condition ($n > 0$) have a mean breeding date ($A + e - n$) closer to the optimal value θ. Price et al. note that this model may also apply to clutch size in birds and might also be a reasonable model for seed germination time, especially if there is a significant nonheritable nutritional contribution from the maternal endosperm.

Additional insight into Fisher's model was offered by Heywood (2005), who used his decomposition of the Price Equation (Chapter 6) to partition the response into linear and spurious response terms plus a transmission bias term (Equation 6.32),

$$R = \beta_{\overline{z},z}\, S + \sigma(w, \overline{z}\cdot\ z) + E[\,\overline{\delta}\,]$$

Recall, from the notation in Chapter 6, that $\overline{z}$ is the offspring value, with $\beta_{\overline{z},z}$ the slope of the parent-offspring regression, so that the linear response term is the breeder's equation analog. Since the basic Fisher model assumes random mating, no epistasis nor cross-generational environmental effects, the offspring mean in the absence of selection is the same as the current parental mean and hence the transmission bias term $E[\,\overline{\delta}\,] = 0$. Heywood shows that the linear response to selection is given by

$$\beta_{\overline{z},z} S = \sigma_A^2\, \frac{\theta - \mu - \omega^2\alpha\sigma_n^2/\sigma_z^2}{\omega^2 + \sigma_z^2} \tag{20.7a}$$

The linear response component is zero when

$$\widehat{\mu}_L = \theta - \omega^2\alpha\sigma_n^2/\sigma_z^2 \tag{20.7b}$$

This is the value for which fitness is optimized with respect to breeding date, and represents the balance between high n values increasing fitness (while decreasing z) being countered by stabilizing selection trying to move z towards θ (note that $\widehat{\mu}_L = \theta$ when $\sigma_n^2 = 0$). Likewise, the partial covariance between fitness (w) and offspring value ($\overline{z}$) when the (linear) effect of parent breeding date z is removed is

$$\sigma(w, \overline{z}\cdot\ z) = \alpha\sigma_n^2\sigma_A^2/\sigma_z^2 \tag{20.8}$$

As discussed in Chapter 6, this partial covariance term can be thought of as a spurious response, because it is the residual response after the linear effect of z on response is removed.

At equilibrium, $\widehat{\mu} = \theta + \alpha\sigma_n^2$ (Equation 20.6), and the linear response to selection (Equation 20.7a) becomes

$$\sigma_A^2 \, \frac{\theta - (\theta + \alpha\sigma_n^2) - \omega^2\alpha\sigma_n^2/\sigma_z^2}{\omega^2 + \sigma_z^2} = \sigma_A^2 \, \frac{-\alpha\sigma_n^2(\sigma_z^2 + \omega^2)/\sigma_z^2}{\omega^2 + \sigma_z^2} = -\alpha\sigma_n^2\sigma_A^2/\sigma_z^2,$$

which is exactly canceled by the spurious response (Equation 20.8), given a net total response of zero. The reason for a non-zero partial covariance between w and offspring value $\overline{z}$ is that fitness is proportional to nutritional status and (due to the exponential nature of the fitness function), residuals of the (linear) fitness regression of w on z are non-random, increasing with z. Likewise, when offspring mean $\overline{z}$ is regressed on z, residuals are proportion to n (since $\overline{z}$ is related to the parental breeding value $A = z + n - e$). As a result, the residuals of both the regressions of $\overline{z}$ on z and w on z covary with n, generating a conditional covariance (Chapter 6). The spurious response arises because both regressions are influenced by a common variable, nutritional status n. Conversely, when $\sigma_n^2 = 0$, there is no spurious response and the linear response term is zero when $\mu = \theta$.

MODIFYING THE BREEDER'S EQUATION FOR NATURAL POPULATIONS

As the above examples show, one of the most serious limitations in applying the breeder's equation in natural populations is that selection can occur on unmeasured characters and/or environmental features. Additionally, genotype-environment *correlations* can be a concern, as (for example) larger individuals may be able to obtain better environments. In artificial selection and breeding situations, this is less of a concern because there is usually some attempt to randomize individuals over (obvious) environments. Here we develop a general expression for the response in a single trait when all of these factors are in play. We do so by first assuming a static environment (absence of environmental change), namely that the distribution of environmental effects within the population remains constant over the generation of response being predicted.

Complications in the Absence of Environmental Change

Just how these complications bias the breeder's equation was examined by van Tienderen and de Jong (1994), with similar analysis (under a multivariate breeder's equation framework) by Hadfield (2008). van Tienderen and de Jong assume complete additivity (no dominance or epistasis), multivariate normality, and linear parent-offspring regressions. As shown in Figure 20.3, they use a path analysis argument (LW Appendix 2) to explore the relationship between response R and the selection differential S when complications such as selection on correlated characters and genotype-environment correlations exist.

To present their analysis, decompose the phenotype z as

$$z = A + E + e_s$$

where A is the additive genetic value, E the general environmental value (for example, the average value for a particular macrohabitat) and e_s the special (or residual) environmental value unique to each individual (LW Chapter 6). By construction, e_s is independent of the other variables (so that the total environmental variance is $\sigma_E^2 + \sigma_{e_s}^2$), but A and E may be correlated. Consider Figure 20.3, which shows possible paths of how the environmental value E, the genotypic value A, and the phenotypic value z can influence fitness. Figure 20.3A shows the breeder's equation assumption that E and A influence fitness only through the phenotypic value z. Figure 20.3B shows the general situation where E and A can influence fitness independent of (or in addition to) their effects on z, as can occur if the focal trait is

phenotypically and/or genetically correlated with other characters under selection (Figure 20.3B). If fitness is entirely determined by the phenotypic value of the focal trait, there should be no expected differences in the fitness of individuals with the same phenotypic value z but different underlying genetic (A) or environmental values (E). For example, suppose two individuals both have a phenotypic value $z = 100$, but individual one has $A = 80, E = 20$, while individual two arrives at this phenotypic value by $A = 10, E = 90$. If selection is entirely on phenotype, both individuals have the same expected fitness, but their expected fitnesses may differ if there is addition selection on A and/or E beyond that based on any selection on z. For example, if correlated characters are under selection, then individuals with the same z value can have different fitnesses due to correlations between their A and/or E values with the genetic and/or environmental values at other traits that influence fitness.

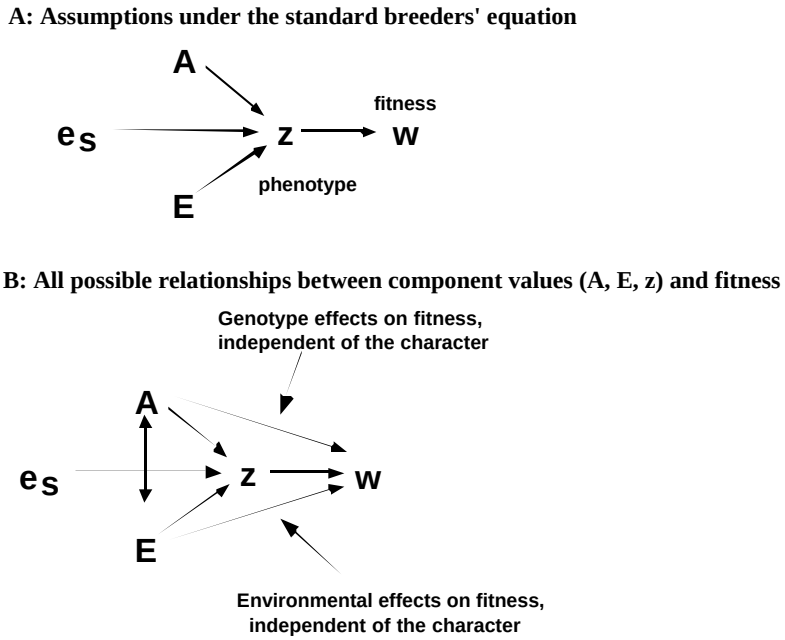


Figure 20.3. The pathways by which the components of a character (phenotype z, additive genetic value A, common environmental effect E, and special environmental effect e_s) influence fitness w. **A**. The breeder's equation assumes that only the phenotype (z) of a character casually influences fitness. This is not an unreasonable starting assumption for artifical selection, wherein the breeder directly chooses individuals on the basis of phenotypes and randomizes environments with respect to phenotypes. **B**. Other pathways by which the components of a character can influence fitness. Either (or both) of the additive genetic and/or environmental values can influence fitness *independent* of their influence on phenotype. For example, an environmental value can influence both the character of interest and independently influence fitness. The influence of genetically correlated characters is through A, as the impact of other traits also under selection whose breedings values are correlated with our focal trait, appears through A and not through z. Similarly, the effect of shared environmental factors on phenotypic correlations appear through E. Finally, genotypic and environmental values may be correlated ($\sigma(A, E) \neq 0$), which is indicated by the double-headed arrows connecting A and E.

To quantify the effects from these different paths influencing fitness, van Tienderen and

de Jong consider the multiple regression of relative fitness w as a function of z, A, and E,

$$w = \alpha + b_z\, z + b_A\, A + b_E\, E + \epsilon \tag{20.10a}$$

The partial regression coefficients b represent the expected change in fitness holding the other variables constant (LW Chapter 8). For example, b_z is the expected change in fitness from a unit change in the phenotype z, holding the other variables (A and E) constant. In particular, if selection is *entirely* based on phenotypic value, then $b_A = b_E = 0$. As shown by Quinn (1992), the condition for this to occur is that the partial correlation (given z) of breeding value and fitness, or of environmental value and fitness, is zero,

$$r_{A,w\cdot z} = r_{E,w\cdot z} = 0 \tag{20.10b}$$

We return to this observation in the next section.

A notational aside is that we use b here (instead of β) for the partial regression coefficients, as we use β in the next section for the univariate slopes of the fitness regression based on either A, or E, or z separately. For example, for the regression $w = 1 + \beta_z z + \epsilon$, the slope is β_z, while b_z is the fitness regression slope when A and E are also included (Equation 20.10a). As developed in the next section, comparing appropriate β and b values provides information on whether unmeasured variables are potentially influencing response.

From multiple regression theory (LW Chapter 8), the partial regression coefficients satisfy

$$\begin{pmatrix} \sigma(w,z) \\ \sigma(w,A) \\ \sigma(w,E) \end{pmatrix} = \begin{pmatrix} S \\ R \\ \sigma(w,E) \end{pmatrix} = \begin{pmatrix} \sigma_z^2 & \sigma(z,A) & \sigma(z,E) \\ \sigma(z,A) & \sigma_A^2 & \sigma(A,E) \\ \sigma(z,E) & \sigma(A,E) & \sigma_E^2 \end{pmatrix} \begin{pmatrix} b_z \\ b_A \\ b_E \end{pmatrix} \tag{20.11}$$

The first vector contains the covariances between relative fitness w and the predictor variables (z, A, and E), with $S = \sigma(w,z)$ following from the Robertson-Price identity (Equation 6.10) and $R = \sigma(w,A)$ from Robertson's secondary theorem (Equation 6.25a). The matrix is the variance-covariance matrix for these predictor variables, where

$$\sigma(z,A) = \sigma(A + E + e_s, A) = \sigma_A^2 + \sigma(E,A)$$

Likewise, $\sigma(z,E) = \sigma_E^2 + \sigma(E,A)$. Using these identities and considering the first two rows of Equation 20.11 gives the within-generation change as

$$\begin{aligned} S &= \sigma_z^2\, b_z + \left[\, \sigma_A^2 + \sigma(E,A) \,\right] b_A + \left[\, \sigma_E^2 + \sigma(E,A) \,\right] b_E \\ &= \sigma_z^2\, b_z + \sigma_A^2\, b_A + \sigma_E^2\, b_E + \sigma(E,A)\ (b_A + b_E)\,, \end{aligned} \tag{20.12a}$$

and the response as

$$\begin{aligned} R &= \left[\, \sigma_A^2 + \sigma(E,A) \,\right] b_z + \sigma_A^2\, b_A + \sigma(E,A)\, b_E \\ &= \sigma_A^2\, b_z + \sigma_A^2\, b_A + \sigma(E,A)\ (\, b_z + b_E)\,. \end{aligned} \tag{20.12b}$$

If there are no genotype-environment correlations [$\sigma(E,A) = 0$],

$$R = \sigma_A^2\ (b_z + b_A) \tag{20.13a}$$

and

$$S = \sigma_z^2\, b_z + \sigma_A^2\, b_A + \sigma_E^2\, b_E \tag{20.13b}$$

It is worth noting the connection between the above expression for S (based on following a single focal trait) and Equation 20.1a, where the differential on a focal trait is expressed in terms of direct selection on that trait plus indirect selection from a correlated trait. Equating the two, we find that $\sigma(z_1, z_2)\beta_2 = \sigma_A^2\, b_A + \sigma_E^2\, b_E$, where z_2 is the unmeasured correlated trait (and $z_1 = z$ the focal trait). Writing $z_1 = A + E$,

$$\sigma(z_1, z_2)\beta_2 = \sigma(A + E, z_2)\beta_2 = \sigma(A, z_2)\beta_2 + \sigma(E, z_2)\beta_2$$

so that with a single unmeasured trait,

$$\sigma(A, z_2)\beta_2 = \sigma_A^2\, b_A, \qquad \text{and} \qquad \sigma(E, z_2)\beta_2 = \sigma_E^2\, b_E$$

Equation 20.12 is much more general than the single correlated trait expression given by Equation 20.1b, as b_A and b_E encompass *all* of the effects of *any* genetically and/or phenotypically traits. Under the formulation for S given by Equation 10.12a, selection on any phenotypically-correlated traits appears as a nonzero value of b_A (if the phenotypic correlations are, at least in part, due to correlated breeding values) and/or b_E (if the correlations are due, at least in part, to shared environmental values).

Multiplying both sides of Equation 20.13b by h^2 and rearranging gives

$$h^2\, \sigma_z^2\, b_z = \sigma_A^2\, b_z = h^2 \left(S - \left[\sigma_A^2\, b_A + \sigma_E^2\, b_E \right] \right).$$

Substituting into Equation 20.13a yields

$$R = h^2\, S + \sigma_A^2 (1 - h^2)\, b_A - h^2\, \sigma_E^2\, b_E \tag{20.13c}$$

Hence, any extra (positive) selection on additive genetic values ($b_A > 0$) inflates response over the breeder's equation, while extra selection on environmental values ($b_E > 0$) decreases response. Following this same approach gives the general response [when $\sigma(E, A) \neq 0$] as

$$R = h^2\, S + \sigma_A^2 (1 - h^2)\, b_A - h^2 \sigma_E^2\, b_E + \sigma(E, A)\, \left(b_z - h^2\, b_A + [1 - h^2]\, b_E \right) \tag{20.14}$$

If selection acts only on the phenotype of the character being considered, then $b_A = b_E = 0$ and Equation 20.12a reduces to $S = \sigma_z^2\, b_z$, implying $b_z = S/\sigma_z^2$. Substituting into Equation 20.12b gives the response as

$$R = b_z \left[\sigma_z^2 + \sigma(E, A) \right] = \left(h^2 + \frac{\sigma(E, A)}{\sigma_z^2} \right) S, \tag{20.15a}$$

which (as expected) reduces to the breeder's equation when there is no genotype-environment correlation. As discussed in Chapter 15, unless the correlation between E and A is perfect, this component of response will be transient, decaying to zero once selection stops.

Under artifical selection, it is generally assumed that individual fitness is entirely based on the phenotype of the character of interest, specifically those phenotypes chosen by the breeder. In this case, the partial regression coefficients of fitness on genotype and environmental values are zero, as phenotype entirely determines fitness. In natural populations, we do not have this luxury and another possibility is that there is no natural selection on the character of interest (its phenotype, by itself, has no effect on fitness so that $b_z = 0$), but rather selection occurs on characters correlated with the one we are following. If these characters under selection are only connected to the character we are following through its breeding

value (i.e., no environmental correlation between characters), then $b_A \neq 0$ while $b_z = b_E = 0$. In this case, using Equation 20.12a to express b_A in terms of S gives the response as

$$R = b_A\,\sigma_A^2 = S\,\frac{\sigma_A^2}{\sigma_A^2 + \sigma(E,A)} \tag{20.15b}$$

which reduces to $R = S$ in the absence of a genotype-environment correlation for the focal trait. The reason for this strong response is that (in this setting) all of the selection is on the breeding value. With selection on a phenotype (the breeder's equation), only a fraction (h^2) translates into selection on breeding values.

A final possibility is that the only correlation between characters under selection and our focal character is through shared environmental effects, giving $b_E \neq 0$ while $b_A = b_z = 0$. In this case the response becomes

$$R = b_E\,\sigma(E,A) = S\,\frac{\sigma(E,A)}{\sigma_E^2 + \sigma(E,A)}, \tag{20.15c}$$

which equals zero unless a genotype-environment correlation exists. Again, in the absence of a perfect correlation between E and A, this response is transient (Chapter 15).

Additional Complications from Environmental Change

The above analysis considers the complications from uncontrolled, but static, environmental effects. A further layer of complications arises when the environment (more formally, the distribution of possible environments) changes from year to year. First, the target of selection can radically change from one year to the next (e.g., Example 20.2). How significant such temporal variation in selection is remains an unresolved question. Siepielski et al. (2009) claimed it is rather common, and changes in sign are not unexpected. Conversely, a reanalysis of the same dataset by Morrissey and Hadfield (2011) concludes that the strength of directional selection in these studies is actually remarkably consistent after accounting for sampling variation. A related question is whether evolution is largely shaped by relatively rare, but major events, (e.g., Example 20.2), or by more gradual, but constant, pressures with less temporal variation.

A second complication is that a major shift in the environment can result in a shift in the trait mean even in the absence of any genetic change. As we will see later, a deterioration in the environment can mask significant underlying genetic change, leading to the appearance of status. Finally, changes in the environment can result in changes in genetic (and environmental) variance components, and hence in h^2.

IS A FOCAL TRAIT THE DIRECT TARGET OF SELECTION?

Causality (the phenotypic value of a focal trait is the sole target of selection) is a critical assumption in applying the breeder's equation to natural populations. As we have seen, an observed selection differential can be generated by direct selection on a trait, direct selection on phenotypically correlated traits, an environmental covariance between the focal trait and fitness, or a combination of all of these (Equations 20.1a, 20.12a). One approach to control for phenotypically correlated traits is to include them in the analysis and then compute the vector $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$ of selection gradients (Equation 13.25a, Chapter 29). However, how does one ascertain if all relevant traits are included in the analysis? Further, many such "traits" could be environmental variables with no heritability, which could easily be missed in even the most careful analysis. One approach to assess causality (initially suggested by Rausher and Simms 1989, Rausher 1992, and Queller 1992), is intimately connected with Robertson's secondary theorem of natural selection (Chapter 6). If the prediction from Robertson's theorem

is consistent with that from the breeder's equation, one has significantly increased confidence that the phenotypic value of the focal trait is the target of selection. We refer to this basic strategy, and its variants, as **Robertson consistency tests**. This approach can also be framed as comparing how much of a selection differential is associated with a trait's breeding versus environmental values, and whether these values are inconsistent with selection based solely on phenotypic value z.

What is the advantage of a consistency test versus simply comparing the realized response in a natural population with its prediction for the breeder's equation? A lack of fit, by itself, is not informative as to which assumption (or assumptions) failed. By contrast, if a consistency test fails, this strongly suggests that selection is acting on more than just the phenotype of our focal trait.

Robertson's Theorem: Response Prediction Without Regard to the Target of Selection

Recall from Chapter 6 that the breeder's equation is not the only expression for predicting response (Table 6.1). Exact (but largely unusable) expression follow from Price's theorem (Equations 6.8, 6.38, 6.39). Under the assumption that parental breeding values are excellent predictors of offspring mean, Robertson's secondary theorem of natural selection (Equations 6.24a, 6.25a) provides an alternative expression for response. As discussed in Chapter 6, there is some confusion in the literature on the secondary theorem, as Robertson actually suggested two slightly different versions. Robertson (1966) suggested $R = \sigma(A_z, w)$, namely that response in a specific trait is the covariance between the breeding value of that trait and relative fitness, while Robertson (1968) suggested $R = \sigma(A_z, A_w)$, where relative fitness w is replaced by its breeding value A_w. The relationship between the 1966 and 1968 versions follows (Equation 6.25c) by noting

$$\sigma(A_z, w) = \sigma(A_z, A_w + e_w) = \sigma(A_z, A_w) + \sigma(A_z, e_w),$$

showing that while the 1966 version is more general, the two are equal when $\sigma(A_z, e_w) = 0$. There is no biological reason to suggest that this covariance should generally be zero, as it simply states that there is a covariance between the residual component of fitness (once breeding value has been removed) and the breeding value of the trait itself.

Example 20.4. One of the earliest application of Robertson's theorem to natural populations is by van Noordwijk (1988), who examined mean nestling weight and offspring survival in Great tits (*Parus major*). By performing regressions of parental nestling weigh on offspring survival, the correlation in breeding values between the trait and this measure of fitness could be estimated. Likewise, estimates of S and h^2 for nestling weight were also computed. The results for 1975 to 1978 were as follows:

Year	S	h^2	$R = h^2 S$	$R = \sigma(A_z, A_w)$
1975	0.24	0.28	0.07	0.00
1976	0.68	0.47	0.32	0.03
1977	0.16	0.26	0.04	0.06
1978	0.53	0.29	0.15	0.05
mean	0.40	0.35	0.14	0.03

The breeder's equation significantly overpredicts response relative to Robertson's theorem, suggesting that factors correlated with nestling weight also influence selection.

Robertson Consistency Tests

Rausher (1992), Queller (1992), and Morrissey et al. (2010, 2012) all suggest that an analysis which jointly estimates required parameters for the breeder's equation and Robertson's theorem can provide insight on whether the breeder's equation should predict response. If the two estimates of response agree, this suggests that the phenotypic value z is largely causative as the target of selection. If they are significantly different, other forces besides selection on z are involved. Note that this analysis simply checks the *static environment* prediction. The consequences from generational changes in E — shifting selection target, shift in trait mean from entirely environmental factors, changes in variance components due to $G \times E$ — can all cause Robertson's theorem (as well as the breeder's equation) to fail. Likewise, if our standard breeding value model is not a good approximation of trait heritability, both Robertson's theorem and the breeder's equation can also fail (Chapter 6).

Under what conditions should the breeder's equation and Robertson's theorem predict the same reponse? Using the more general 1966 version, the two predicted responses are equal when

$$R = h^2 S = \frac{\sigma^2(A_z)}{\sigma^2(z)}\,\sigma(z,w) = \sigma(A_z,w) \tag{20.16a}$$

Rearranging the last equality yields the result of Queller (1992, also Hadfield 2008),

$$\frac{\sigma(z,w)}{\sigma^2(z)} = \frac{\sigma(A_z,w)}{\sigma^2(A_z)} = \frac{\sigma(A_z,A_w)+\sigma(A_z,e_w)}{\sigma^2(A_z)} \tag{20.16b}$$

The left hand side is the slope of the linear regression of w on z, while the middle equality is the regression of w on the breeding value A_z of the trait. If the two sides of Equation 20.16b are very similar, then the breeder's equation is likely to hold (subject to the assumptions of a static environment and the infinitesimal model). However if the two sides are significantly different, there are additional targets of selection besides the phenotypic value of the focal trait that influence the response of the focal trait.

Rausher's Consistency Criteria

Rausher (1992) obtained the multivariate version of Equation 20.16b by equating the vector $\mathbf{R}$ of responses under Robertson's theorem with the multivariate breeder's equation (Equation 13.26a),

$$\mathbf{R} = \sigma(\mathbf{z}_A, w) = \mathbf{G}\boldsymbol{\beta} \tag{20.17a}$$

with the ith component of the vector $\sigma(\mathbf{z}_A, w)$ is $\sigma(z_{A_i}, w)$, the covariance between the breeding value of trait i and relative fitness. Multipling the left side by the identity matrix $\mathbf{I} = \mathbf{G}\mathbf{G}^{-1}$ gives

$$\mathbf{G}\mathbf{G}^{-1}\sigma(\mathbf{z}_A, w) = \mathbf{G}\mathbf{B} = \mathbf{G}\boldsymbol{\beta}$$

where

$$\mathbf{B} = \mathbf{G}^{-1}\sigma(\mathbf{z}_A, w) \tag{20.17b}$$

is the vector of coefficients of the regression of relative fitness on the breeding values of $\mathbf{z}$. **Rausher's consistency condition** is $\mathbf{B} = \boldsymbol{\beta}$ or $\mathbf{G}^{-1}\sigma(\mathbf{z}_A, w) = \mathbf{P}^{-1}\sigma(\mathbf{z}, w)$, where we have used the multivariate version of the Robertson-Price identity (Equation 6.10), $\mathbf{S} = \sigma(\mathbf{z}, w)$. Rearranging yields

$$\mathbf{G}\mathbf{P}^{-1}\sigma(\mathbf{z}, w) = \sigma(\mathbf{z}_A, w) \tag{20.17c}$$

This a slight generation of Rausher's (1992) result, which assumed the 1968 version of Robertson's theorem, with $\sigma(\mathbf{z}_A, A_w)$ replacing $\sigma(\mathbf{z}_A, w)$. For a univariate trait, $h^2\sigma(z,w) = \sigma(z_A, w)$, which directly implies

$$h^2 S_z = S_A \tag{20.18}$$

Namely, that the selection differential S_A based on the breeding value of a trait is just h^2 times the phenotypic selection differential S_z. Although Rausher's condition (Equation 20.17c) directly leads to Equation 20.18, the formal regression test he proposed (Rausher and Simms 1989, Rausher 1992) is slightly different, and thus Equation 20.18 is referred to as **Postma's test** (Postma 2006).

Rausher framed his formal consistency test in terms of the relative strengths of selection on the *components* of the focal trait phenotypic value, namely its breeding A and environmental E value (or more generally its residual value, as E can include nonadditive genetic terms). Consider the slopes β_z and β_A of the *univariate* regressions of relative fitness on phenotype z and breeding value A. Here β_z is just the selection gradient $\beta = \sigma(z,w)/\sigma_z^2$ (Equation 13.8b), while (from the definition of regression slope, LW Equation 3.14b) $\beta_A = \sigma(A,w)/\sigma_A^2$. When Equation 20.18 holds,

$$\beta_A = \frac{\sigma(A,w)}{\sigma_A^2} = \frac{h^2\sigma(z,w)}{h^2\sigma_z^2} = \frac{\sigma(z,w)}{\sigma_z^2} = \beta_z \tag{20.19a}$$

Similarly, for the regression of fitness on E, $\beta_E = \sigma(E,w)/\sigma_E^2$. Noting that $\sigma_A^2/h^2 = \sigma_z^2$ and $\sigma_E^2/(1-h^2) = \sigma_z^2$, we can relate the univariate fitness regression slope of z with those for A and E as follow,

$$\begin{aligned}\beta_z = \frac{\sigma(w,A)+\sigma(w,E)}{\sigma_z^2} &= \frac{\sigma(w,A)}{\sigma_A^2}h^2 + \frac{\sigma(w,E)}{\sigma_E^2}(1-h^2) \\ &= \beta_A h^2 + \beta_E(1-h^2)\end{aligned} \tag{20.19b}$$

When Equation 20.18 holds, $\beta_z = \beta_A$, in which case Equation 20.19b implies that $\beta_z = \beta_E = \beta_A$ (provided $\sigma(A,E) = 0$). This observation suggests a test for an environmentally-induced fitness-trait correlation. Following Rausher and Simms (1989) and Rausher (1992), compute the multiple regression of fitness on both the breeding value A and the environmental deviation E,

$$w = 1 + b_A A + b_E E + e \tag{20.20}$$

If z is the sole target of selection, then $b_A = b_E$, which can be tested in a straightforward fashion using standard results from regression theory (LW Chapter 8). A closely related test is whether b_A is significantly different from zero, as this indicates that at least some of the selection is translated into selection on the breeding value of the focal trait. The equality test of $b_A = b_E$ is stricter, in that it assumes equal selection on both breeding value and environmental deviation. In particular, it tests whether selection is strictly a function of phenotypic value, no matter how that phenotype is obtained (e.g., individuals with high breeding value vs. high environmental deviation, but the same phenotype, experience the same amount of selection). When $b_A \neq b_E$, the reader might be included to assume that the b with the larger magnitude implies more selection on that component (outside of their common effects on z). To see that this is misleading, suppose selection on phenotype influences both components by 0.4, while additional selection on A adds -0.35, and additional selection on E adds 0.05, resulting in $b_A = 0.05$ and $b_E = 0.45$. While this suggests more selection on E, the additional component of selection (beyond that due to z) was much stronger on A.

These tests were suggested before the application of animal models to natural populations, leading to the critical issue of how to estimate breeding values. Rausher and Simms (1989) and Stinchcombe et al. (2002), focusing on plants, replicated genotypes (when clones are available) or sibs (half, full, or selfed) over environments, estimating the genotypic value of a clone by its average over the sampled environments and likewise assigning all sibs the same breeding value, namely their family mean. Since sib (or clone) means replicated over environments are used for the breeding/genotypic values, there are different sample sizes

associated with b_A (number of families) and b_E (number of individuals). Stinchcombe et al. (2002) discuss how to deal with this issue. Using this regression approach, Stinchcombe et al. and Scheiner et al. (2002) compared genotypic and phenotypic estimates of selection for six plant species grown on experimental plots (and hence stricter environmental control than expected in populations fully in nature). Even in these settings, these authors found that a significant fraction (around 25%) of the traits appeared to show selection on factors other than z (b_A significantly different from b_E). While this bias rarely resulted in a change in sign, it often significantly impacted the magnitude of the estimated selection

While tests based on replicated genotypes/sibs are an important conceptual breakthrough, their actual application is very limited. In particular, they *critically* depend upon randomization of genotypes over environments. The estimated breeding value assigned to all members of a sibship is their family effect, which is a function of the mean breeding value of their parents but also of maternal effects and common family environmental values. In the extreme case, if environments are not randomized, a common family environment could influence both the trait and fitness, and this would appear in the family effect. In this case, b_A could be significantly different from zero, but this is a reflection of selection on common family environmental values (including maternal effects), not on breeding values.

The realization in the early 2000's that animal models (Chapter 19) could return estimates of trait breeding values for *individuals* lead to a brief period (with numerous publications) where BLUP estimated breeding values were used to test for associations between trait breeding value and fitness (e.g., Kruuk 2004). While potentially much more powerful than clone or family studies (using individual, rather than group, breeding values), as we detail below, given the structure of most natural pedigrees, BLUP estimated breeding values have a strong environmental bias, and were eventually realized to be highly unreliable for these sort of studies (Postma 2006, Postma and Charmantier 2007, Hadfield 2008, Hadfield et al. 2010, Wilson et al. 2010). However, the power of an animal model analysis can still be used through a direct REML estimate of either $\sigma(A_z, A_w)$ or $\sigma(A_z, w)$. By directly estimating this covariance under a bivariate animal model (as opposed to regressing fitness on predicting breeding values), the pitfalls of using individually predicted breeding values can be avoided (Hadfield et al. 2010). We examine all these issues in detail shortly.

Morrissey et al.'s Consistency Criteria

An alternate expression for consistency can be obtained as follows. Writing

$$\sigma(z, w) = \sigma(A_z + e_z, A_w + e_w) = \sigma(A_z, A_w) + \sigma(e_z, e_w) + \sigma(A_z, e_w) + \sigma(e_z, A_w),$$

the consistency condition given by Equation 20.16b becomes

$$\frac{\sigma(A_z, A_w) + \sigma(e_z, e_w) + \sigma(A_z, e_w) + \sigma(e_z, A_w)}{\sigma^2(A_z) + \sigma^2(e_z)} = \frac{\sigma(A_z, A_w) + \sigma(A_z, e_w)}{\sigma^2(A_z)}. \qquad (20.21a)$$

This can be expressed as

$$1 + \frac{\sigma(e_z, e_w) + \sigma(e_z, A_w)}{\sigma(A_z, A_w) + \sigma(A_z, e_w)} = 1 + \frac{\sigma^2(e_z)}{\sigma^2(A_z)},$$

implying

$$\frac{\sigma(e_z, e_w) + \sigma(e_z, A_w)}{\sigma(A_z, A_w) + \sigma(A_z, e_w)} = \frac{\sigma^2(e_z)}{\sigma^2(A_z)}.$$

Finally, this rearranges to yield an equivalent consistency condition,

$$\frac{\sigma(e_z, e_w) + \sigma(e_z, A_w)}{\sigma^2(e_z)} = \frac{\sigma(A_z, A_w) + \sigma(A_z, e_w)}{\sigma^2(A_z)} \qquad (20.21b)$$

Morrissey et al. (2010, 2012) arrived at the slightly different condition

$$\frac{\sigma(e_z, e_w)}{\sigma^2(e_z)} = \frac{\sigma(A_z, A_w)}{\sigma^2(A_z)}, \tag{20.21c}$$

by assuming Robertson's 1968 version (and hence $\sigma(A_z, e_w) = 0$) and also that $\sigma(e_z, A_w) = 0$.

Example 20.5. Morrissey el al. (2012) used a bivariate animal model (the focal trait plus fitness as the second trait) to estimate the variance components required for Equation 20.21c. They examined four morphological traits in Soay sheep (*Ovis aries*) on the island of St. Kilda. Body size was of special interest, because parameter estimates suggest a positive response using the breeder's equation, yet the sheep are, if anything, getting smaller. By contrast, the expected response under the secondary theorem is slightly negative (but not significant different from zero). A joint test as to whether the consistency condition given by Equation 20.21c holds was significant at $p = 0.048$, showing the two sides are significantly different. Thus, failure of response (as predicted by the breeder's equation) is likely a result of unmeasured factors upwardly biasing selection on the phenotype that does not influence selection on the breeding value.

The Breeder's Equation vs. the Secondary Theorem

As mentioned, the elegance of the breeder's equation is that it fully separates ecology S from genetics h^2. Queller (1992) noted that when Equation 20.16b is satisfied, this separation occurs. More formally, **Queller's separation condition** states that the partial correlation of A and w given z is zero, $r_{Aw \cdot z} = 0$ (Equation 20.10b, which also implies $r_{Ew \cdot z} = 0$, see Queller 1992). This is simply another way of interpreting Equation 20.16b, and states that the residual values of A and w (following their regression on z) are uncorrelated (e.g., Equation 6.31a). Thus, after accounting for the phenotypic value, there is no residual correlation between breeding value and fitness. Note that when the separation condition holds, Heywood's spurious response term (Chapter 6) is zero.

By constant, the secondary theorem fully *confounds* (rather than separates) selection and genetics, and says absolutely nothing about the nature of selection on the phenotype. Rather, it simply does the accounting and asks what fraction of selection translates into direct selection on the breeding value. The secondary theorem is thus largely about genetics (van Tienderen and de Jong 1994, Morrissey et al. 2012) and largely devoid of ecology. As such, it is generally expected to be more predictive than the breeder's equation as it ignores the actual target of selection (but, as mentioned, can still fail). When the two responses agree, we have some confidence that we have accurately found the target of selection — z is causal, and thus the breeder's equation is not compromised by unmeasured variables.

Which approach, the breeder's equation or the secondary theorem, should be used by an investigator? In large part, it depends on the question. In a conservation biology setting, such as trying to predict if a species has sufficient genetic variation to withstand a major environmental change, response is the major issue of concern. An example of this is the early paper by Etterson and Shaw (2001), who used Robertson's theorem to show that there are significant constraints in response to selection from climate change in a native annual legume from the Great Plains region. While targets of selection are always of interest, the more pressing concern here was whether the population could mount a successful response. Robertson's theorem can address this question without any bias due to unmeasured characters influencing the focal traits. Conversely, the targets of selection are always of interest

to ecologists and evolutionary biologists, and the joint use of the breeder's equation and Robertson's theorem can help clarify the importance of candidate traits.

APPLYING ANIMAL MODELS TO NATURAL POPULATIONS: BASICS

Recall from Chapter 19 that mixed models offer a very flexible platform for genetic analysis in the presence of multiple fixed effects and multigenerational relatives. In particular, the general animal model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \sum_{i=1}^{k} \mathbf{Z}_i\mathbf{u}_i + \mathbf{e} \tag{20.22a}$$

has been widely used in animal breeding since the 1970's. Here $\boldsymbol{\beta}$ is the vector of unknown fixed effects, $\mathbf{a}$ the random vector of breeding values, $\mathbf{e}$ the random vector of residuals, and possibly k other vectors of random effects $\mathbf{u}_i$. These additional random effects can accommodate permanent environmental effects under a repeated records design, common family and/or maternal effects, and other concerns that can complicate the residual error structure (Chapter 19). In these models, $\mathbf{y}$ is an observed vector, while $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{Z}_i$ are matrices of known constants. The power of a mixed model is its ability to borrow information across observations. This is done for fixed effects through $\mathbf{X}$, which keeps track of which observations contribute information on a particular fixed effect. For random effects, this is done through their covariance structure, which determines the strength of additional information provided by correlated observations. The vector $\mathbf{a}$ of breeding values has a covariance structure given by the (assumed known) design matrix $\mathbf{A}$, and we assume $\mathbf{a}$ and $\mathbf{e}$ are uncorrelated,

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{e} \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_A^2\,\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2\,\mathbf{I} \end{pmatrix} \tag{20.22b}$$

Similar assumptions are made about the covariance structures for any additional random effects (Chapter 19), and the covariance structure plus Equation 20.22a fully specifies the model. REML is usually used to estimate variance components, while the vector $\widehat{\mathbf{a}}$ of **predicted breeding values** (**PBVs**) is determined by BLUP. These are also called **estimated breeding values** (**EBVs**) in the literature, but our preference is to use predicted for random effects and estimated for fixed effects. As we will see, while REML estimates are appropriate when using animal models in wild populations, using PBVs is not (unless within an appropriate Bayesian framework). Indeed, Hadfield et al. (2010) "discourage future use of BLUP as an inferential tool in the fields of ecology and evolutionary biology". Why this is the case, and solutions to some of these issues, are examined in the next major section. Here, our focus is on limitations faced when trying to construct an animal model for a wild population, with analysis issues examined later.

Animal Model Analysis in Natural Populations: Overview

Given sufficiently strong pedigrees, animal models can help separate genetic from environmental trends (Chapter 19), and their ability to estimate *individual* breeding values offers the possibility of more accurate Robertson consistency tests. Given these features, it is surprising that their application to natural populations is rather recent, starting with Konigsberg and Cheverud (1992) and Cheverud and Dittus (1992), who applied them to free-living primate populations. These papers went somewhat unnoticed, and a second wave of applications to ungulate mammals and nesting birds started in 1999 (Réale et al. 1999) and has been a rapid growth industry ever since (Kruuk 2004, Kruuk and Hadfield 2007, Postma and Charmantier 2007, Kruuk and Hill 2008, Wilson et al. 2010, Hadfield et al. 2010).

Table 20.1. Design limitations when applying animal models to natural populations.

Relationship Matrix $\mathbf{A}$ must be estimated
Pedigree errors result in bias and lower power.
Open population structure
Immigration from outside the study area complicates interpretation of model results.
Lack of sufficient size or depth of the sampled pedigree (low connectedness)
Low variation in sample relatedness results in low power, potential confounding of parameters of interest.
Lack of power complicates model selection
Low power to detect variance components can complicate results.
Inclusion of additional random effects (because of significant variance components) can results in a different interpretations of key parameters relative to models excluding them. When power is low, significant issues about which model is appropriate.

The animal model has generally been quite successful in the analysis of artificial selection experiments/breeding programs (Chapter 19). However, natural populations differ in fundamental ways from these more controlled settings, which leads to a number of design issues (Table 20.1). First, in natural populations the relationship matrix must be estimated, and this is usually done with a bias towards finding mother-offspring (maternal) links, while missing father-offspring (paternal) connections. Second, artificial selection experiments, and many breeding programs, involve **closed populations**, with little to no immigration from outside sources once selection has started, with most (if not all) animals in the population included in the analysis. Further, when immigration occurs, it is usually controlled and hence immigrants can be identified in the pedigree. This is *highly* problematic in natural populations, where in most settings immigration from outside the study area is the norm, not the exception. Immigrants potentially bring in a different distribution of breeding values, and sufficiently high immigration can both remove any signal of local genetic change or falsely create such a signal when none is present. Likewise, analysis in natural populations is usually based on a somewhat random sample of individuals, rather than information from the entire population, which is usually the case for artifical selection experiments.

Finally, an important consequence of the more open structure of natural populations is that the connectiveness (number of links, i.e., relatives) is lower, often substantially so, than for breeding programs. The $\mathbf{A}$ matrices from breeding programs tend to be denser than for natural populations (more, and larger, off-diagonal elements), as most individuals in the sample have measured relatives in previous generations. This is *not* ensured for random samples from a natural population. To see the significance, consider an individual unlinked to any others in our pedigree, whose PBV is simply the estimated heritability times their phenotypic value (adjusted for fixed effects). In the simplest case of a single fixed effect (the mean μ), the PBV is just $\widehat{a} = h^2(z - \mu)$. When an individual has links (via $\mathbf{A}$) to other members in the sample, BLUP uses this covariance information to obtain an improved estimate of their PBV, making it less dependent on just their phenotypic value (which has some environmental influence). In large pedigrees with many links (such as many breeding programs), this additional information can be substantial, and significantly improves the accuracy of PBVs. By contrast, in natural populations the number of links may be far less, in which case an individual's predicted breeding value is largely determined by their phenotype alone. In such cases, PBVs can be strongly influenced by the environment, a point we return to below. Because of the fragility of individual PBV estimates (especially when pedigree links are sparse), their use is now strongly discouraged in such settings (Postma 2006, Hadfield

2008, Hadfield et al. 2010).

One important observation is that, to date, most estimates of heritability based on mixed-model analysis of wild populations are *lower* than more traditional estimates based on parent-offspring regressions (Kruuk 2004). As mentioned in Chapter 19, the opposite is expected. Heritability under a mixed model is defined as the ratio of additive variance to the sum of all variance components, the later computed *after* fixed effects (and thus a significant source of variation) are removed (Wilson 2008). Phrased another way, once fixed effects are removed, the sum of all variance components is less than the trait variance σ_z^2 when individuals differ in their fixed effects. Thus, parent-offspring regression vs. mixed-model h^2 estimates are looking at slightly different quantities, with the later estimate expected to be larger. Why does this not seem to be the case? Given that the vast majority of these mixed-model estimates are for species with extensive parental care (birds and mammals), part of this difference may arise from lack of control over maternal effects. Most parent-offspring regressions in the wild are mother-offspring, confounding direct and material effects, inflating regression-based estimates of h^2. Mixed models that include maternal effects remove this bias.

Obtaining the Relationship Matrix: Direct Observation of the Social Pedigree

The central difficulty in applying the animal model to free-living populations is obtaining the relationship matrix $\mathbf{A}$ for the measured sample of individuals. In natural populations, the required pedigree information can be extremely difficult to obtain. One source of information are **social pedigrees** based on field observations. If we observe a mother nursing an offspring, we have fairly high confidence that the offspring is from that mother. Accessing paternity is more difficult. Again, field observations may be useful, for example which male visits the nest in pair-bonded birds, or which appears to be the dominant male in other social settings.

Of course, none of these observations is fool-proof. **Intraspecific brood parasitism** can occur where a female lays an egg in the nest of another female. Likewise, even with (apparently) pair-bonded birds, **extra-pair paternities** can occur. This occurs at around 15% in the collared flycatchers (*Ficedula albicollis*) discussed in Example 20.1 (Sheldon and Ellegren 1999). In other species, it can be upwards of 50%, so the simple observation of a male helping at the nest does not imply they are the father. Similarly, determining the dominant male in a harem is no guarantee that he sired all the offspring. Because of this intrinsic bias towards determining the mother, pedigrees from wild populations often show an excess of **maternal linkages**. This has implications when maternal effects are present, as the pedigree must contain a sufficient number of **paternal linkages** to disentangle direct from maternal effects (Clément et al. 2001, Kruuk 2004). Pedigree errors can be high even in systems with apparently strong control over matings. Visscher at al. (2002) estimated a sire error rate of around 10% for UK dairy cattle, despite very widespread use of artificial insemination. Recording errors and the biological vagaries of organisms should never be underestimated!

Obtaining the Relationship Matrix: Marker Data

A second source of information is provided from polymorphic molecular markers. These can be grouped into two categories: those that are hypothesis-driven (e.g., tests for paternity of candidate males or that individuals are full sibs) and those that make no a prior assumptions about relatedness. Put another way, a focus on categorical **relationships** (assigning pairs of individuals into discrete classes such as parent-offspring, full- or half-sibs) versus continuous measures of **relatedness** (estimates of the coefficient of coancestry). A number of methods to estimate the pairwise relatedness have been proposed (reviewed by Ritland 2000, van de Casteele et al. 2001, Blouin 2003, Garant and Kruuk 2005, Thomas 2005, Csilléry et al. 2006, Oliehoek et al. 2006, Frentiu et al. 2008, Pemberton 2008). At first blush, one might think to simply use one of these to estimate the pairwise relatedness between all sampled individuals

and use these as the elements of **A**. There are numerous problems with this approach. First, there are high sampling variances with these estimates (see the above reviews). Second, such a procedure typically does not result in the molecular-based relationship matrix **R** used to estimate **A** being positive-definite (Frentiu et al. 2008), and hence is not a covariance matrix. There is also the issue that some pairwise methods may return negative estimates of relatedness for unrelated individuals. These are typically set to zero, but this introduces a bias akin to that introduced by setting negative variance estimates to zero.

Unless one has a very dense set of markers (e.g., Example 20.6), a better approach is to ignore more distant relationships that must be inferred solely by molecular markers and instead use markers to confirm (or find) sets of close relatives, such as assigning parentage (Blouin 2003, Jones and Ardren 2003, Jones et al. 2010, Walling et al. 2010) or individuals into sibships (Thomas and Hill 2000). Much of the power in a mixed model comes from data on such close relatives, and the initial focus should be on detecting, and confirming, such close linkages. Another reason for this focus is that studies currently do not use a significant number of markers, which limits the resolution of ancestry to just one or two generations back. Most studies using markers to infer paternity or to assign individuals to sibships typically use no more than a couple of dozen of microsatellite loci. These are highly polymorphic markers, and hence just a few have significant power to assign very recent ancestry. While powerful for detecting first-degree relatives (which share half their alleles IBD), the expected fraction of alleles shared between two relatives with a common ancestor k generations in the past is $(1/2)^{2k-1}$. Relatives with a common ancestor two generations in the past thus share only $1/8$ of their alleles IBD, greatly reducing the power for a relatively small number of markers to detect this degree of ancestry (or smaller) with any power.

With a modest number of markers, the most powerful approach is to combine marker data with additional information, such as ranges of specific individuals and their behavior (i.e., apparent position in a dominance hierarchy). Hadfield et al. (2006) presents such an analysis, wrapped within a Bayesian framework, so that uncertainty in relationship estimates is fully captured in the posterior uncertainty of parameter estimates. Further, in a full Bayesian analysis, information at different levels can inform each other (O'Hara et al. 2008). Consider a setting where individuals A and B are, based on marker information, almost equally likely to be the father of C, but since A has a slightly higher probability, it is used in a standard likelihood analysis. Phenotypic data on whether C is closer to A or B certainly provides additional information on which is the father, but is ignored in a sequential likelihood analysis (estimate relationships first, then use these to estimate genetic parameters). Under a Bayesian analysis, this additional information influences the posterior paternity estimates.

Example 20.6. While dense SNP Chips (platforms scoring hundreds of thousands to millions of SNPs in a single pass) at the time of this writing exist only for humans and a few important domesticated and model species, they are expected to become widespread (or replaced by other techniques such as whole-genome sequencing). SNP chips (or similar high-density polymorphism data) offer a very simple approach for obtaining the relationship matrix **A**. Given their very low mutation rates, two SNP alleles that are alike in state (AIS) can be viewed as being identical by descent (IBD), allowing us to compute the coefficient of coancestry θ_{ij} (LW Chapter 7) directly from the SNP data, and hence the entry $A_{ij} = 2\theta_{ij}$ in the relationship matrix. Recall that θ_{ij} is simply the probability that a randomly-drawn allele from individual i and a randomly-drawn allele from individual j are IBD, or in our case alike in state at a given SNP. Coding the two alleles at a SNP locus as **0**/**1**, if (at a given SNP locus) individual i is **11** while j is **10**, then all random draws from i are allele **1**, while half the random draws from j are also **1**, giving (for that locus) $\theta_{ij} = 1/2$. If i is **11** and j is **00**, then θ_{ij} (at this locus) is zero. Likewise if both are **10**, then with probability $(1/2)(1/2)$

= 1/4, SNP allele **1** is drawn from both, while with probability 1/4, SNP allele **0** is drawn from both, while all other draws do not match, giving 1/4 + 1/4 = 1/2 as the coefficient of coancestry. The resulting coefficients of coancestry for all possible combinations for a biallelic SNP becomes

	Genotype of i		
Genotype of j	**11**	**10**	**00**
11	1	0.5	0
10	0.5	0.5	0.5
00	0	0.5	1

One computes the coefficient of coancestry for each SNP, taking the average value over all loci as the coefficient of coancestry for that pair of individuals. Toro et al. (2002) refer to this as **molecular coancestry**. Note that we can compare an individual with itself ($i = j$), which returns 1 for each homozygous locus and 1/2 for each heterozygous loci.

This approach turns out to be significantly more important than simply a way to obtain $\mathbf{A}$ without a pedigree. With sufficiently large number of SNPs, it actually returns a *better and more accurate* estimate of relatedness than is obtained through a pedigree. To see this, suppose individuals 1, 2, and 3 are full sibs. In a pedigree approach, all pairwise θ_{ij} values would be the same, 1/2. However, under Mendelian sampling, there is variation in the number of alleles shared (Risch and Lange 1979, Suarez et al. 1979, Stam 1980, Guo 1996,Visscher et al. 2006), so that (for example) 1 and 2 may share 0.55 of their alleles, while 1 and 3 only 0.42. The above approach captures this variation, giving more accurate weights when combining information from relatives. This is the basis of the genomic selection method known as **G-BLUP** (genomic-BLUP), where a marker-estimated matrix is used in place of a pedigree matrix for $\mathbf{A}$ to improve the BLUP estimates (e.g., Hayes et al. 2009). Volume 3 examines this in detail.

Oliehoek al. (2006) give a nice discussion of the connection between this molecular coancestry estimator and other more elaborate estimators that require knowledge of individual allele frequencies. Denote the two alleles (which may be alike in state) in x by a and b and similarly by c and d in y. The molecular similarity at locus ℓ between x and y is defined by

$$S_{xy,\ell} = \frac{I_{ac} + I_{ad} + I_{bc} + I_{bd}}{4}$$

where I_{ad} is an indicator function which equals one if a and d are AIS, otherwise is zero. This generalizes our estimator to an arbitrary number of alleles, which again is given by the average of $S_{xy,\ell}$ over all loci. More generally, one can consider some base (or reference) population, with s_ℓ denoting the probability that two randomly-drawn individuals in that population are AIS. Obviously, s_ℓ (at a minimum) is a function of the allele frequencies at ℓ. As shown by Lynch (1988), the expected value for $S_{xy,\ell}$ is given by

$$E[\, S_{xy,\ell}\,] = \theta_{xy} + (1 - \theta_{xy})s_\ell,$$

which suggests a more general estimator

$$\widehat{\theta}_{xy} = \frac{1}{L}\sum_{\ell=1}^{L} \frac{S_{xy,\ell} - s_\ell}{1 - s_\ell}$$

Oliehoek al. show that our simple estimator (taking $s_\ell = 0$ for all ℓ, which corresponds to assuming a very ancient base population) is rather robust, and (for a large number of loci) performs equally well, or better, than other estimates that require information on allele frequencies to infer s_ℓ. Microsatellites are expected to be much more unreliable in this setting

than SNPs, as they have high mutation rates, allowing unrelated alleles to easily mutate to the same state (number of repeats). Hence, two alleles that are AIS can easily trace back to unrelated alleles. Conversely, SNP alleles have very low mutation rates, so two alleles AIS very likely had a common ancestor (albeit potentially quite ancient). Finally, negative estimates of θ can arise when $S_{xy,\ell} < s_\ell$ over a larger number of loci, implying these individuals are more unrelated than expected by change. Assuming $s_\ell = 0$ eliminates this problem.

In the few cases where the estimates from a pedigree plus marker-information study were compared to an entirely marker-inferred pedigree, erratic behavior in the variance components is seen (e.g., Thomas et al. 2002, Coltman 2005, Frentiu et al. 2008, Pemberton 2008). In part, this is likely do to the very low resolution offered by using up to a few dozen markers to estimate relationships. As Example 20.6 suggests, this may be less of a limitation when one scores tens of thousands of markers. Given the potential of very dense marker information to more accurately infer relationships, it has been suggested that most wild populations will soon have the potential for animal-model style analysis, with $\mathbf{A}$ directly estimated from marker information (Moore and Kukuk 2002).

Does this mean that, in the near future, animal-model analyses will be widely applicable to many/most natural populations? The answer is likely no, as even if $\mathbf{A}$ is estimated with complete accuracy, any analysis is still limited by the variance among relationships in the sample. If the sample lacks sufficient links between relatives (low variance), it will contain little information for an animal-model analysis (Thomas and Hill 2000, Thomas et al. 2002, Csilléry et al 2006). For example, if one randomly samples a very large population over multiple generations, there is a reasonable expectation that very few, if any, relatives will be found. This could still generate a very large sample (and hence the suggestion of significant power), but if $\mathbf{A}$ is of the form $\mathbf{A} = c\mathbf{I} + \epsilon\,\mathbf{B}$, where $\epsilon \ll 1$, then practically speaking, the sample consists of unrelated individuals ($\mathbf{A}$ is essentially a diagonal matrix). With little information from relatives, the power of a mixed-model analysis vanishes, as most breeding values are simply estimated from that individual's phenotype alone. Balancing this pessimistic view are two studies on free-living fish that spend at least part of their time in the open ocean, which found sampled individuals are enriched for close relatives (Thériault et al. 2007, DiBattista et al. 2009).

Consequences of Pedigree Errors

Since the strength of a mixed-model analysis arises from accurately borrowing information from relatives, pedigree errors result in both bias and a loss of power. Generally speaking, there are two types of such errors — missing a link and incorrectly linking unrelated individuals together. In natural populations, there is a strong bias towards maternal connections, in that most mother/offspring connections will be found and incorrect assignment of a mother to the wrong cohort of offspring is unlikely (but certainly possible). By contrast, assigning fathers is much more problematic, in that incorrect fathers can be assigned to some offspring, while other offspring will have unassigned fathers, even when their true fathers are in the sample. Missed/incorrect paternal assignments have the effect of making what the model assumes to be unrelated individuals (father in the sample, but not assigned to offspring) more similar and related individuals (incorrect father) less similar. This generally reduces heritability estimates. More importantly, there are significant implications for the detection of maternal effects. If most pedigree errors are paternal, them offspring will tend to resemble their mothers more than their fathers, resulting in a false signal of maternal effects (Postma and Charmantier 2007). Proper estimation of maternal effects requires a significant

number of correct paternal links, as these are what allows direct and maternal effects to be disentangled (Chapter 22).

The consequences of misassigned paternities are a function of the pedigree structure and trait heritability. In animal breeding designs, where there is a great excess of mothers (dams) over fathers (sires), the effects can substantial. In beef cattle, Lee and Pollak (1997) observed a significant reduction in the estimated heritability (0.1 versus the true value of 0.3) when 20% of the sires were misidentified. Their pedigree structure had roughly 2% sires, 22% dams, and the rest were nonparents. In contrast, simulation studies by Charmantier and Réale (2005) with roughly equal percentage of sires and dams found that if the rate of extra-pair paternity was under 20%, then the biases in h^2 were modest (less than 15%). They also found that bias introduced by misassigned paternities increases with h^2. Interestingly, they also found cases where misassigned paternity can actually inflate estimates of h^2. Keller et al. (2001), working with morphological traits in Darwin's finches (*Geospiza fortis*), found that not accounting for maternal effects introduced a much greater bias than did extra-pair matings. Milner at al. (2000), working with Soay sheep (*Ovic aries*), found that variance estimates decreased between a pedigree with 95% confidence on a paternity versus one with 80% confidence. Finally, incorrect pedigree links have an even greater impact on estimates of individual PBVs, which are key in accessing direct selection on breeding values. As seen with the prediction of a breeding value from a single phenotype, $\widehat{a} = h^2(z - \mu)$, the effect of the heritability is to shrink an estimate backs towards the mean μ. If h^2 is large, most the phenotypic deviation is kept by the estimate, while if h^2 is close to zero, all estimates are shrunk back to very near zero. Since pedigree errors typically result in underestimated heritabilities, this results in excessive shrinkage of PBVs (Geldermann et al. 1986, Israel and Weller 2000). As a result, true extreme breeding values are underestimated and low breeding values overestimated (Figure 20.4 below).

How does an investigator deal with all of this pedigree uncertainty? Henderson (1988) suggested if a father is not known with certainty, it may be more efficient to include all possible sires (weighted by their paternal probabilities) than to simply not include any sire-offspring linkages in the pedigree. He introduced the idea of an average numerator relationship matrix to accomplish this. Konigsberg and Cheverud (1992) applied this approach to estimate craniometric traits on a macaque colony on Cayo Santiagos. Here, mothers were known with certainty, but sires are unknown. However, field and social data can be used to exclude many males as possible sires, leading to a subset of potential sires for each offspring. If there were k possible sires, Konigsberg and Cheverud weighted them with equal probability $(1/k)$ and applied Henderson's method. The natural (and more sophisticated) extension of this idea is a fully Bayesian approach, where uncertainty in the pedigree estimates is directly built into the model, and the resulting marginal posteriors fully incorporate all of this uncertainty.

Performing a sensitivity analysis is critical before applying an animal model. Following Morrissey et al. (2007), the investigator first assumes a rough pedigree framework for their study population and then incorporates the types of pedigree errors suspected given the biological system in question. Simulation studies can then be used to examine power (the ability to detect parameters) and sensitivity (how robust these estimates are in the face of pedigree errors). Software is given by Morrissey and Wilson (2010). Using a framework pedigree for Soay sheep (*Ovis aries*), Morrissey et al. found that the simple animal model (breeding values are the only random effect) was relatively robust to pedigree errors, but that when maternal effects were included, the results were more fragile. This is not surprising as separating maternal and direct effects is fairly sensitive to the types of links in the pedigree. Quinn et al. (2006), using the pedigrees for two bird species, suggested a rough rule of thumb that at least three generations (years) and 100 individuals/year is required to estimate heritability with confidence. These numbers should be treated as lower bounds, as other

sampled natural populations may contain fewer relatives than do samples from these species, significantly reducing power.

In summary, although there appears to be a wealth of tools for using molecular markers to assign relationships, using them as the sole means to reconstruct **A** is suspect at best unless the marker density is extremely high. Rather, the combination of using field observations to first suggest potential linkages, and then molecular markers to confirm these, should provide fairly reliable **A** matrices (albeit culled for more distant relatives). With a multigenerational study, this approach can provide links across generations and connecting these links over several generations can largely fill out the important entries in **A**. A further caveat is that, as mentioned in Chapter 19, BLUP and REML methods can be compromised by previous selection, which is exactly what is expected in natural populations. Given the generally smaller size and depth (connectiveness) of wild pedigrees relative to pedigrees from much larger breeding programs, a full and formal accounting of all uncertainty is critical, and our recommendation is that Bayesian approaches be used whenever possible.

Model Selection

A final concern when using wild pedigrees is the always delicate issue of model selection — which additional random effects (if any) should be incorporated and how does one select the most appropriate model among a collection of candidates? Typically one incorporates additional random effects when their associated variance component is significant or when the model fit is improved by some criteria (likelihood ratio test or model selection statistics such as AIC or BIC). One interesting consequence of a Bayesian analysis is that unless a variance prior has some point mass at zero, the resulting posterior confidence interval will always exclude zero, resulting in all variance components (for which the MCMC convergences) being significant in the sense of zero being excluded. In such settings, goodness of fit such as BIC should be the criteria used for model selection.

Given the low power associated with wild pedigrees, especially when a full accounting of uncertainty is done, our ability to detect potentially important, but small, variance components may be rather limited. For example, the genetic variance in relative fitness is quite important, but generally expected to be small (Chapter 6). While such additional variance components may be small (or worse, almost fully confounded with other components, e.g., Kruuk and Hadfield 2007, Ovaskainen et al. 2008), their incorporation may significantly change the interpretation of a key feature (such as the additive variance) of a model. The problem faced by an investigator is when to include such additional factors, given that the power to declare a variance component significant is likely small. With the expected low power for many wild populations, the best strategy may be to present analyses under a series of models (even if the resulting variance components may not be significant), if they result in substantially different interpretations of key results.

APPLYING ANIMAL MODELS TO NATURAL POPULATIONS: BEST PRACTICES

Animal models have been used to address two important questions regarding the nature of response on a focal trait. First, is their evidence of selection on the breeding values of that trait, and if so, is it consistent with selection largely on the phenotype? Second, is there any the genetic trend (change in mean breeding values) in the data? Historically (the first decade of the 2000's), these tasks were accomplished by regressing relative fitness on PBVs for the former and population mean PBVs on time/generation for the later. As detailed below, both of these approaches are flawed when simply using the PBVs directly, but both questions can be safely addressed within an animal-model framework with appropriate adjustments (Table 20.2). These best-practices, and their motivation, are the subject of this section.

Table 20.2. Summary of best practices for examining common evolutionary questions on response with the animal-model framework. Full details in text.

Task: Robertson consistency tests based on BV-fitness associations	
Problem: The variance of PBVs is less than the additive variance (Postma 2006). Biases Rausher's test (Equation 20.20), as slope $\beta_{\widehat{a}}$ of fitness-PBV regressions overestimates the slope β_A of the fitness - BV regressions (Equation 20.24b).	
Solution: Use a bivariate (trait, relative fitness) animal model and frame tests in terms of REML variance components, e.g., Morrissey's test (Equation 20.21c).	
Problem: If trait BVs estimated with a univariate animal model, $\sigma(\widehat{a}, w)$ is a biased estimate of $\sigma(A, w)$, as correlations between BVs and w are not incorporated into the model (Hadfield 2008). Biases Postma's test (Equation 20.18).	
Solution: Use a bivariate animal model and estimate S_z and S_A directly from REML variance components (Equations 20.27, 20.28).	
Task: Detecting genetic trends using temporal regressions based on PBVs	
Problem: The error structure associated with PBVs is GLS, not OLS, with heteroscedatic and correlated residuals. Results in strongly anticonservative tests, with p values highly biased towards smaller values (Hadfield et al. 2010).	
Solution: Use a Bayesian posterior for the regression slopes (Figure 20.5, Example 20.7).	

Consistency Tests: Accuracy, Reliability, and Caveats with Using PBVs

Many of the initial applications of animal models in natural populations used PBVs in tests of selection on breeding versus environmental values (Equations 20.18, 20.20). For example, Rausher's regression is fit using PBVs, $w = 1 + b_A\,\widehat{a} + b_E\,\widehat{e} + \epsilon$. Postma (2006) noted that there are significant, albeit subtle, problems with using an estimated value $\widehat{a}$ in place of the true breeding value A. His central point was that when the pedigree adds little additional information, the PBV for an individual is almost entirely determined by their phenotype. These resulting PBVs are biased by environmental values (as these influence the phenotype of an individual), which confounds their ability to separate A from E. Further, as shown by Hadfield (2008, Hadfield et al. 2010), PBVs are correlated, and this correlation structure must be taken into account for proper inference.

To quantify these concerns, we first need to consider several related measures of the uncertainty of predicted breeding values. Their **accuracy** ρ is the correlation between the predicted $(\widehat{a})$ and actual (A) breeding value (Chapter 13), while **reliability** ρ^2 is the fraction of variation in A accounted for by the PBVs. When using just the phenotype of an individual to obtain their PBV, $\rho = h$ (Chapter 13). The difference $\rho^2 - h^2$ between the reliability of a particular PBV and the heritability is a measure of how much additional information is provided from relatives. Writing a PBV as its expected value (the true breeding value) plus an uncorrelated residual, $\widehat{a} = A + \epsilon$, then $\sigma^2(\widehat{a} - A) = \sigma^2(\epsilon)$ is the **prediction error variance** (**PEV**, Chapter 19). When the PEV is small, the PBVs will be very close to the true breeding values.

These measures of uncertainty in PBVs are connected as follows. By construction, A and the residual ϵ are uncorrelated, implying $\sigma(\widehat{a}, A) = \sigma(\widehat{a}, \widehat{a} - \epsilon) = \sigma^2(\widehat{a})$. This allows us to express the PEV as

$$\sigma^2(\widehat{a} - A) = \sigma^2(\widehat{a}) - 2\sigma(\widehat{a}, A) + \sigma^2(A) = \sigma^2(A) - \sigma^2(\widehat{a}) \tag{20.23a}$$

Second, the definition of a correlation gives the accuracy as

$$\rho = \frac{\sigma(\widehat{a}, A)}{\sqrt{\sigma^2(\widehat{a})\,\sigma^2(A)}} = \sqrt{\frac{\sigma^2(\widehat{a})}{\sigma^2(A)}} \tag{20.23b}$$

Hence,

$$\sigma^2(\widehat{a}) = \rho^2\sigma^2(A) \tag{20.23c}$$

from which it immediately follows that $\sigma^2(\,\widehat{a}\,) \leq \sigma^2(A)$. We have already seen a hint of this result, in that BPV are shrunk back towards their expected values A, reducing their variance relative to the variance of their true values (Figure 20.4). These results imply that the correlation between PBV and phenotype is greater than the correlation between A and z,

$$|\,\rho(\widehat{a}, z)\,| = \frac{|\,\sigma(\widehat{a}, z)\,|}{\sqrt{\sigma^2(\widehat{a})\,\sigma_z^2}} = \frac{|\,\sigma(A, z)\,|}{\sqrt{\sigma^2(\widehat{a})\,\sigma_z^2}} = \frac{|\,\sigma(A, z)\,|}{\rho\sqrt{\sigma^2(A)\,\sigma_z^2}} = \frac{|\,\rho(A, z)\,|}{\rho} \geq |\,\rho(A, z)\,|$$

Equations 20.23a and 23c give the connection between ρ and the PEV as,

$$\sigma^2(\widehat{a} - A) = \left(1 - \rho^2\right)\sigma^2(A) \tag{20.23d}$$

Finally, for $z = A+E$, Postma (2006) shows that the prediction error variance is the covariance between the PBV and the environmental deviation E,

$$\text{PEV} = \sigma(\widehat{a}, E) = \left(1 - \rho^2\right)\sigma^2(A) \tag{20.23e}$$

The implication of these results is that PBVs resemble the phenotype more than do true breeding values (Figure 20.4), and hence they are biased by environmental factors that influence the phenotype. In particular, Equation 20.23e shows that the prediction error is the influence of an individual's environmental value on their PBV. When $\rho^2 \sim 1$, environmental value has no impact, but when ρ^2 is modest to small, it has a significant impact, resulting in the PBV being a mixture of the true breeding value A and plus an error reflecting environmental effects on the phenotype.

Figure 20.4 illustrates this phenomena in a plot of predicted breeding values versus phenotypic values. When breeding value is predicted based solely on the phenotype z of a single individual (open circles), there is no variation about its predicted value of $h^2(z - \mu)$, as all individuals with the same z value have the same predicted breeding value. In this case, while there is no residual variance in the PBV for a given z (as all PBVs are the same), the residual variance between z and the *true* breeding value can be considerable, $\sigma^2(\epsilon) = \left(1 - h^2\right)\sigma^2(A)$. With a standard heritability of around 0.3, this amounts to 70% of the additive variance not being accounted for by the PBVs. As additional pedigree information influences the PBV, the reliability ρ^2 exceeds h^2, and (influenced by information from relatives) predicted values start to vary about their mean value $h^2(z - \mu)$ for a given z, giving a more correct picture of their true values. The influence of relatives is to make an individual's predicted breeding value less dependent on their phenotypic (and hence environmental) value. The residual variance for the true breeding value for a given phenotypic value z is $(\rho^2 - h^2)\sigma^2$, which is maximized at $(1 - h^2)\sigma_A^2$ when the reliability is complete ($\rho = 1$). Postma and Charmantier (2007) note that PEVs are often around 0.5 for wild pedigrees, so that roughly half of the estimate of a typical PBV is influenced by the environment.

The prediction error variance for any specific PBV can be obtained from Equations 19.5c and 19.5d, which give the covariance matrix of the PEVs for each of the estimated breeding values (also see Meyer 1989, Tosh and Wilton 1994). The i-th diagonal element gives the prediction error variance for individual i as

$$PEV_{ii} = (1 - \rho_i^2)\sigma_A^2, \tag{20.23e}$$

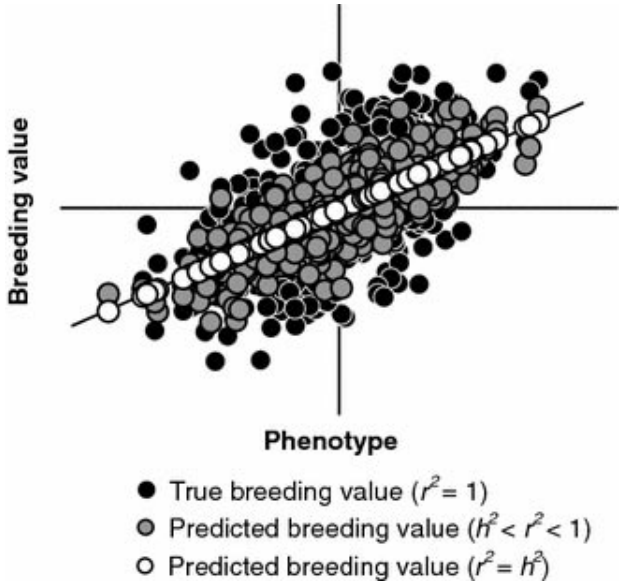


Figure 20.4. Comparison of predicted breeding values (PBVs) as a function of phenotypic value z based on different amounts of information from relatives. With only a single observation (i.e., no relatives), PBVs show no variation about their predicted value $\widehat{a} = h^2(z-\mu)$, and $\rho^2 = h^2$. As more information from relatives is added, $\rho^2 > h^2$, and PBVs become less dependent on an individual's phenotype, showing greater spread about the regression. For a given phenotypic value z, the residual variance of a PBV around its mean predicted value $h^2(z-\mu)$ is $(\rho^2 - h^2)\sigma_A^2$. When $\rho = 1$ (PBV = true BV) this spread around z is $(1-h^2)\sigma_A^2$, which can be substantial when heritability is low (z is poor predictor of A). After Postma and Charmantier (2007).

and the predicted reliability as

$$\rho_i^2 = 1 - PEV_{ii}/\sigma_A^2 \tag{20.23f}$$

The amount by which this exceed h^2 is a measure of the amount of additional information on that individual provided by the pedigree. A more subtle, but equally important, point is that the PEV matrix is not diagonal, rather PBVs for relatives are correlated. Hence, the residual error structure for Rausher's regression (Equation 20.18) is complex, requiring GLS, not OLS regression (LW Chapter 8). We return to this point, which has very significant consequences for detecting genetic trends.

These results have important implications for populations under selection. If an individual is lost before it leaves offspring, (i.e., its fitness is zero), it will have fewer links in the pedigree than individuals who survive to leave offspring. Individuals of low fitness are thus expected to have lower reliabilities than individuals with higher fitness (Postma 2006), resulting in their PBVs being more influenced by environmental values than are higher-fitness individuals. Obviously, this can bias estimates of the amount of selection on breeding value.

Besides the likelihood of differential bias for individuals with different fitness, Postma points out that a critical component of Rausher's regression (Equation 20.20) does not hold for when the regression is based on $\widehat{a}$. Recall that when selection is entirely on phenotypic value (Equation 20.16b is satisfied), then

$$\beta_A = \frac{\sigma(w,A)}{\sigma^2(A)} = \frac{h^2\sigma(w,z)}{h^2\sigma^2(z)} = \beta_z \tag{20.24a}$$

However, when using predicted breeding values $\widehat{a}$ in place of A,

$$\beta_{\widehat{a}} = \frac{\sigma(w,\widehat{a})}{\sigma^2(\widehat{a})} = \frac{h^2\sigma(w,z)}{\rho^2 h^2\sigma^2(z)} = \frac{\beta_z}{\rho^2} \tag{20.24b}$$

When $\rho^2 < 1$ (as would be expected in wild pedigrees), $\beta_A < \beta_{\widehat{a}}$ and gradients based on predicted breeding values overestimate the gradient expected for true breeding values, compromising Rausher's $b_A = b_E$ test for selection entirely on phenotype.

To circumvent this problem, Postma suggests that the consistency test $h^2 S_z = S_A$ (Equation 20.18) be used instead, as when selection is entirely on the phenotype of the focal trait,

$$S_A = \sigma(A,w) = S_{\widehat{a}} = \sigma(\widehat{a},w)$$

However, Hadfield (2008) shows that $\sigma(\widehat{a},w)$ is a biased estimate of $\sigma(A,w)$ in a univariate animal model, because the breeding values are estimated in a model *separately* from fitnesses. This estimate is unbiased only when Equation 20.16b is satisfied (i.e., exactly one of the Robertson consistency conditions being testing). Fortunately, the solution to both these problems is to *jointly* model the trait and fitness in a bivariate animal model, and then use REML-estimated variance components for consistency tests.

Bivariate Animal Models: REML Estimates of $\sigma(A_z, A_w)$, S_A, and S_z

LW Chapter 26 introduced the multivariate animal model, where a series of potentially correlated traits are measured in a set of pedigreed individuals. For trait i, a standard animal model is fit,

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_j\mathbf{a}_i + \mathbf{e}_j \tag{20.25a}$$

where

$$\begin{pmatrix}\mathbf{a}_j\\ \mathbf{e}_j\end{pmatrix} \sim \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\sigma^2_{A_j}\,\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \sigma^2_{e_j}\,\mathbf{I}\end{pmatrix} \tag{20.25b}$$

Let $\mathbf{z}$ denote the vector of phenotypes for the trait of interest and $\mathbf{w}$ the vector of corresponding relative fitnesses, so that z_i, w_i are the value of the focal trait and relative fitness for individual i. The resulting bivariate mixed model becomes

$$\begin{pmatrix}\mathbf{z}\\ \mathbf{w}\end{pmatrix} = \begin{pmatrix}\mathbf{X}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_w\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_z\\ \boldsymbol{\beta}_w\end{pmatrix} + \begin{pmatrix}\mathbf{Z}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_w\end{pmatrix}\begin{pmatrix}\mathbf{a}_z\\ \mathbf{a}_w\end{pmatrix} + \begin{pmatrix}\mathbf{e}_z\\ \mathbf{e}_w\end{pmatrix} \tag{20.26a}$$

Note that the structure of this model allows the trait and fitness to have different fixed effects. The model can logically be extended to include more complex designs, such as maternal effects and/or repeated measured (e.g., Morrissey et al. 2012). The resulting covariance structure for the stacked vector of breeding values is

$$\boldsymbol{\sigma}\begin{pmatrix}\mathbf{a}_z\\ \mathbf{a}_w\end{pmatrix} = \begin{pmatrix}\sigma^2(A_z)\mathbf{A} & \sigma(A_z,A_w)\mathbf{A}\\ \sigma(A_z,A_w)\mathbf{A} & \sigma^2(A_w)\mathbf{A}\end{pmatrix} = \mathbf{G}\otimes\mathbf{A} \tag{20.26b}$$

where $\otimes$ denotes the Kronecker (or direct) product (LW Chapter 26) and

$$\mathbf{G} = \begin{pmatrix}\sigma^2(A_z) & \sigma(A_z,A_w)\\ \sigma(A_z,A_w) & \sigma^2(A_w)\end{pmatrix} \tag{20.26c}$$

is the matrix of genetic covariances of interest. Similarly, the covariance structure for the stacked vectors of residuals is

$$\boldsymbol{\sigma}\begin{pmatrix}\mathbf{e}_z\\ \mathbf{e}_w\end{pmatrix} = \mathbf{E}\otimes\mathbf{I}, \quad \text{where} \quad \mathbf{E} = \begin{pmatrix}\sigma^2(e_z) & \sigma(e_z,e_w)\\ \sigma(e_z,e_w) & \sigma^2(e_w)\end{pmatrix} \tag{20.26d}$$

Finally, we need to specify any covariances between $\mathbf{a}$ and $\mathbf{e}$. By construction $\sigma(a_z, e_z) = \sigma(a_w, e_w) = 0$, while the standard assumption is $\sigma(A_z, e_w) = \sigma(A_w, e_z) = 0$, giving the covariance structure as

$$\boldsymbol{\sigma}\begin{pmatrix}\mathbf{a}_z\\ \mathbf{a}_w\\ \mathbf{e}_z\\ \mathbf{e}_w\end{pmatrix} = \begin{pmatrix}\mathbf{G}\otimes\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}\otimes\mathbf{I}\end{pmatrix} \tag{20.26e}$$

The resulting six variance components in $\mathbf{G}$ (Equation 20.26b) and $\mathbf{E}$ (Equation 20.26c) are estimated by REML, and this is our main interest in the model. For example, $\sigma(A_z, A_w)$ is the expected response under the 1968 version of Robertson's theorem. From the assumption that $\sigma(A_z, e_w) = \sigma(A_w, e_z) = 0$,

$$S_z = \sigma(z, w) = \sigma(A_z + e_z, A_w + e_w) = \sigma(A_z, A_w) + \sigma(e_z, e_w) \tag{20.27a}$$

is a direct estimate of the selection differential on z, while

$$S_A = \sigma(A_z, w) = \sigma(A_z, A_w + e_w) = \sigma(A_z, A_w) \tag{20.27b}$$

is an estimate of the selection differential on the breeding value of z. By using the variance components directly we avoid the pitfalls associated with working with PBVs. Postma's (Equation 20.18) and Morrissey's (Equation 20.21c) consistency conditions are easily computed from these variance components. Ideally, this is done within a Bayesian setting, so that the posterior distribution reflects all of the model uncertainly (from pedigree estimation on down), and draws from the sampler can be used to compute the distribution of (say) $S_A - h^2 S_z$, and thus a test of whether there is sufficient support to include (consistent) or exclude (inconsistent) zero.

Removing the assumption that $\sigma(A_z, e_w) = \sigma(A_w, e_z) = 0$, the covariance structure given by Equation 20.26e is replaced by

$$\boldsymbol{\sigma}\begin{pmatrix}\mathbf{a}_z\\ \mathbf{a}_w\\ \mathbf{e}_z\\ \mathbf{e}_w\end{pmatrix} = \begin{pmatrix}\mathbf{G}\otimes\mathbf{A} & \mathbf{C}\\ \mathbf{C}^T & \mathbf{E}\otimes\mathbf{I}\end{pmatrix} \tag{20.28a}$$

where

$$\mathbf{C} = \begin{pmatrix}\boldsymbol{\sigma}(\mathbf{a}_z, \mathbf{e}_z) & \boldsymbol{\sigma}(\mathbf{a}_z, \mathbf{e}_w)\\ \boldsymbol{\sigma}(\mathbf{a}_w, \mathbf{e}_z) & \boldsymbol{\sigma}(\mathbf{a}_w, \mathbf{e}_w)\end{pmatrix} = \begin{pmatrix}\mathbf{0} & \sigma(A_z, e_w)\mathbf{A}\\ \sigma(A_w, e_z)\mathbf{A} & \mathbf{0}\end{pmatrix} \tag{20.28b}$$

These two additional variance components allow us to use the more general consistency condition given by Equation 20.21b. Likewise, to apply Postma's test (Equation 20.18), the phenotypic and breeding value selection differentials become

$$S_z = \sigma(A_z, A_w) + \sigma(e_z, e_w) + \sigma(A_z, e_w) + \sigma(A_w, e_z) \tag{20.28c}$$

and

$$S_A = \sigma(A_z, A_w) + \sigma(A_z, e_w) \tag{20.28d}$$

One major point of discussion is that while normality assumptions for a trait are often reasonable, this is not the case for fitness (Chapters 28, 29). For example, if fitness is number of offspring, an overdispersed Poisson (Poisson with an additional point mass at zero, e.g,, Chapter 14) might be a more reasonable distribution. In such cases, mixed models are replaced by generalized mixed models (Bolker et al. 2009; Chapter 29), see Example 20.7.

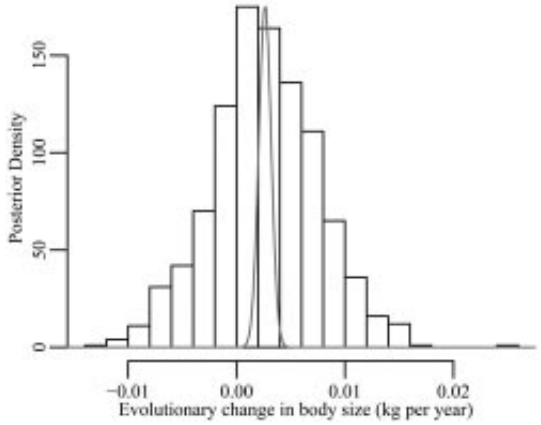


Figure 20.5. The anticonservative nature of using PBVs regressed on time to detect a genetic trend. Here, the trait is body weight in Soay sheep, with the thin smooth curve centered slightly to the right of zero being the distribution of slope values using BLUP-PBVs, where the individual PBVs were assumed to be uncorrelated. This results in a slope that is significantly positive. Conversely, the histogram gives samples from the posterior distribution of slopes from a Bayesian analysis. While this has the same mean as the BLUP-based slope estimates, its variance is significantly greater. Indeed, 0.283 of the probability mass is less than zero, showing that this estimate of the slope, which more fully accounts for the uncertainty and correlation among individual estimates, is not significant different from zero. (After Hadfield et al. 2010).

Detecting Genetic Trends

The gold standard for detecting a genetic change is to grow two populations in a common garden experiment, ensuring that any change is genetic, rather than environmental. Unfortunately, with the exception of the use of remnant seed, contemporaneous comparisons of genetic composition of different generations is not possible. As we saw in Chapter 19, it is often possible to do this comparison statistically though the animal model, provided the population is sufficiently connected across the generations by sampled relatives. Due their pedigree depth and connectedness, this is reasonable for most breeding programs and artificially-selected populations. The reliability of PBVs in such settings is fairly high, reducing any environmental influence on PBVs. Further, PBVs are unbiased (provided individuals are randomized over environments), with their average smoothing out some of the environmental noise, allowing a plot of PBVs over time to show a trend (or lack thereof). In theory, this same approach can be applied to natural populations, provided that genotype-environment correlations are largely ignorable.

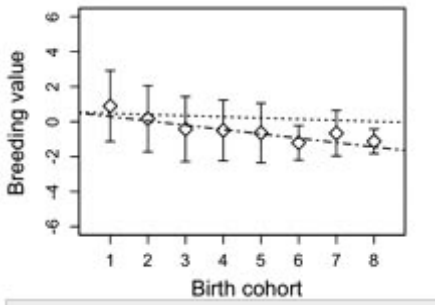
The more delicate issue is that of *inference*. As we saw in Chapter 18, both OLS and GLS regressions of response on selection differential gave unbiased *estimates* of the realized heritability, but since the residuals are heteroscedastic and correlated, the sampling variances under OLS are far too small relative to the correct values under GLS (see Example 18.6). As a result, OLS-based tests are **anticonservative**, namely the p values are heavily biased towards smaller values (i.e., more significant values). The same issue arises with the error structure of PBVs (Hadfield et al. 2010), as Figure 20.5 illustrates.

Fortunately, there is a simple solution (Hadfield al. 2010), again within the powerful Bayesian framework that (additionally) accounts for all the levels of uncertainty in our analysis. The vector $\mathbf{a}$ of PBVs from the posterior distribution is drawn from a given iteration of the MCMC sampler, and the resulting series of PBVs is used to regress (OLS) mean PBV

on generation, and the resulting slope recorded. Multiple slopes are sampled this way to construct an empirical histogram of the posterior slope distribution, which is then used for inference (see Figure 20.5). The resulting empirical distribution of slopes fully accounts for the correlated structure among the PBVs.

Detection of a significant genetic trend is not, by itself, evidence of a selection response, as this could arise simply from drift. A powerful feature of the MCMC approach is that it allows for an easy test of drift versus selection (Hadfield et al. 2010). First, a value for σ_A^2 is sampled from the marginal posterior. Breeding values (with a mean of zero, and this sampled variance) are then drawn from a normal and assigned to individuals in the pedigree with no past relatives. These parental BVs, coupled with Mendelian sampling (Chapter 16), are then used to generate random breeding values (RBP) for their downstream relatives, generating a set of RBVs over the specific pedigree under the assumption of drift (as the BVs are chosen at random). With a set of RBP in hand, a genetic trend is computed as above (regression of mean RBPs over generations), and the slope recorded. Extracting many such samples (redrawing σ_A^2 at the start of each) generates an empirical histogram of the posterior slope distribution under drift, which can be compared with those generated from the MCMC using the full data.

Example 20.7. Milot et al. (2011) examined the evolution of age at first reproduction (AFR) in the isolated island of Île aux Coudres in Québec. This island was settled by thirty families between 1720 and 1772. Because of careful church records, it has a very detailed record of births, marriages, and deaths, allowing the authors to construct a bivariate animal model for AFR and lifetime reproductive success (LRS). The later is a proxy for total fitness and was defined in this study as the total number of offspring from a woman that reach age 15. The authors assumed a normal distribution for AFR, but a latent Poisson model for LRS. Under this model (discussed in Chapter 14), breeding value was defined on an underlying latent scale, with the distribution of LRS for an individual with a latent score of $y = \mu_w + A_w + e_w$ following a Poisson distribution with mean e^y. MCMC methods were used in a Bayesian analysis of this bivariate model, which found significant heritability in both AFR and LRS. Further, these two traits showed a significant negative genetic correlation (posterior mode of -0.81, 95% credible interval of -0.97 to -0.48). Thus, Robertson's theorem suggests direct selection to reduce age of first reproduction, which declined from roughly age 26 to roughly age 22 over a 140 year period. The regression of predicted breeding value over time (measured as eight 20-year cohorts) is shown below, with the diamond representing the average of 1,000 MCMC samples from the marginal posterior ($\pm$ their standard errors).



The authors then tested whether this trend could be due to drift, using the approach suggested by Hadfield et al. (2010). Given the posterior estimate of the additive variance for AFR, random breeding values (RBVs) were generated over the known pedigree, and the regression of the RBVs over time verses that for PBVs was compared for each run of the

sampler. The proportion of times where the absolute regression slope of the RBVs exceeded the slope based on PBVs was taken as the posterior probability the response is due to drift, and was found to be less that 0.01. The average slope for RBVs is given in the figure by the dotted line.

An issue debated in the literature is whether one should also include a year (or generation) effect in models for trends. When PBVs have low reliability, they are strongly influenced by the environment, with the PBV trend partly reflecting any underlying environmental trend (Postma 2006). Incorporation of a year effect (a fixed value for each year/generation) can account for such an environmental trend, but it will also partly absorb any true genetic trend, reducing power. Postma and Charmantier (2007) offer some guidelines as to how to proceed (Figure 20.6). Given that PBVs are biased toward an environmental trend when their reliability is low, if the genetic and environmental trends are in opposite directions, this is evidence of a genetic trend, and a year effect need not be incorporated. If the change is in the same direction, and still persists after a year effect has been incorporated, this is also supportive of a genetic trend. Figure 20.6 presents their flow chart for the other combinations.

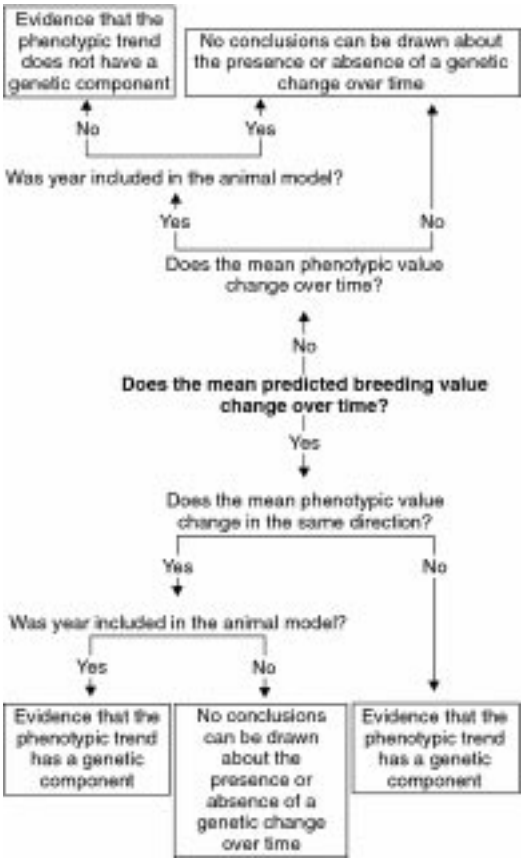


Figure 20.6. Postma and Charmantier's (2007) recommendations for the interpretation of genetic trends.

Table 20.3. Examples of natural populations of mammals and birds in which apparent strong directional selection on a heritable trait fails to show response. Length is the length of the study (in years). After Merilä et al. (2001c) and Gienapp et al. (2008).

Species/Trait	h^2	$\lvert\bar{i}\rvert$	Response	Lenght	
Mammals					
Cervus elaphus (Red deer)					
Antler mass	0.33	0.43	Opposite	29	Kruuk et al. (2000, 2002)
Birth mass (Male)	0.11	0.40	No change		
Birth mass (Female)	0.25	0.22	No change		
Ovis aries (Soay sheep)					
Body mass (Male)	0.12	0.11	No change	12	Milner et al. (1999, 2000)
Body mass (Female)	0.24	0.07	No change		
Ovis canadensis (Big Horn sheep)					
Body weight	0.23	-0.295	As expected	29	Coltman et al. (2003, 2005)
Horn length	0.39	-0.331	As expected	26	
Tamiascurus hudsonicus (Red Squirrel)					
Parturition date	0.16	-0.17	As expected	10	Réale et al. (2003)
Birds					
Branta leucopsis (Barnacle Goose)					
Tarsus lenght (M)	0.53	0.03	Opposite	13	Larsson et al. (1998)
Tarsus lenght (F)		0.09	Opposite		
Anser caerulescens (Snow Goose)					
Clutch size	0.20	0.3	Opposite	20	Cooke et al. (1990)
Cygnus olor (Mute Swan)					
Clutch size	0.20	0.66	As expected	25	Charmantier et al. (2006)
Ficedula albicollis (Collared Flycatcher)					
Relative mass	0.30	0.23	Opposite	17	Merilä et al. (2001a, b)
Tarsus lenght	0.52	0.12	No change	4	Alatalo et al. (1990)
	0.35	0.18	No change	17	Kruuk et al. (2001)
Breeding time	0.19	-0.22	No change	19	Sheldon et al. (2003)
Parus caeruleus (Blue Tit)					
Body mass	0.27	0.31	No change	14	Charmantier et al. (2004)
	0.35	0.42	No change	12	
Tarus length	0.47	0.27	No change	13	
	0.48	0.21	No change	12	
Parus major (Great Tit)					
Breeding time	0.17	-0.21	No change	30	Perrins and Jones (1974) Gienapp etal. (2006)
Egg size	0.80	0.38	No change	7	Hõrak et al. (1997)
Fledging Mass	0.24	0.21	Opposite	36	Garant et al. (2004)
	0.20	0.14	Opposite	36	Garant et al. (2005)
	0.29	0.18	No change	36	

CAUSES OF APPARENT FAILURES OF RESPONSE IN NATURAL POPULATIONS

Given the above concerns on the suitability of the breeder's equation in natural populations, what do the data say? A detailed review of well-studied mammal and bird populations by Merilä et al. (2001c) noted a number of cases where (i) there was a consistent selection

differential on a particular trait, (ii) that trait was heritable, and yet (iii) no response (or worse, response in the opposite direction) was observed over a lengthy period (many generations) of study (Table 20.3). While there are several classic examples of natural populations responding to either imposed artificial selection (Example 18.1, Semlitsch and Wilbur 1989), a new environmental challenge (such as a habit shift, major weather event, or introduction of a novel selective agent, e.g., Example 20.2; Losos et al. 1997; Reznick et al. 1997), or even the introduction of a new competitor species (Grant and Grant 2006), Merilä et al. lamented the apparent lack of response outside of these situations, raising the central question as to the basis for this apparent **statsis** in response in the face of apparent selection.

Table 20.4 summarizes possible (not mutuality exclusive) explanations, most of which have been discussed previously. The most obvious is that the phenotype of the focal trait is not the sole target of selection. However, it could be as trivial as lack of sufficient power to detect a small expected response (e.g., Gienapp et al. 2006, Postma et al. 2007). A related design issue is that most studies only sample a small part of an open population, so that immigration and/or differential dispersal can mask, or enhance, any local selection response (Garant et al. 2005). Changing environments can result in multiple effects that could cause apparent stasis. The selection differential could be changing sign over generations, resulting in a net long-term differential of close to zero. Deterioration in the environment can mask underlying genetic change. Finally, when G x E is present, heritabilities change with the environment, offering the possibility of low heritabilities when selection is most intense. The tools developed in this chapter help an investigator to sort through these possible explanations, as the following case studies illustrate.

Table 20.4. Possible causes for an observed stasis in response despite heritable variation and a significant selection differential.

Genetic response has occured, but not is detected
Low power for detecting a genetic trend
Genetic gain countered by environmental deterioration
Focal trait not target of selection
Trait and fitness correlated through an environmental variable
Selection on a phenotypically, but not geneticallly, correlated trait
Consequence of open population structure
Immigration from populations outside of study areaa
Consequence of fluctuating environmental conditions
Fluctuating selection differential, with little net selection
Fluctuating h^2 that is smallest when selection is strongest
Constraints and tradeoffs
Direct response on a trait countered by correlated responses from other traits
Incomplete measurement of Fitness

Cryptic Evolution: Genetic Change Masked by Environmental Change

One explanation for stasis is that change in the environment can dilute, and indeed even swamp, any underlying genetic change. In the extreme, one can have **cryptic evolution** – significant genetic change that does not show up as phenotypic change because it is countered by environmental change. Levins (1968) and Conover and Schultz (1995) coin the phrase **countergradient variation** for situations in which the environmental trend is opposite to the direction of selection. Such situations can increase the strength of a selection on a trait,

as the population struggles to keep pace with the declining environment. In the extreme, a population faces the risk of extinction if the environment is deteriorating at a faster rate than compensating traits can evolve. Obviously, this is an important issue for populations attempting to track climate change.

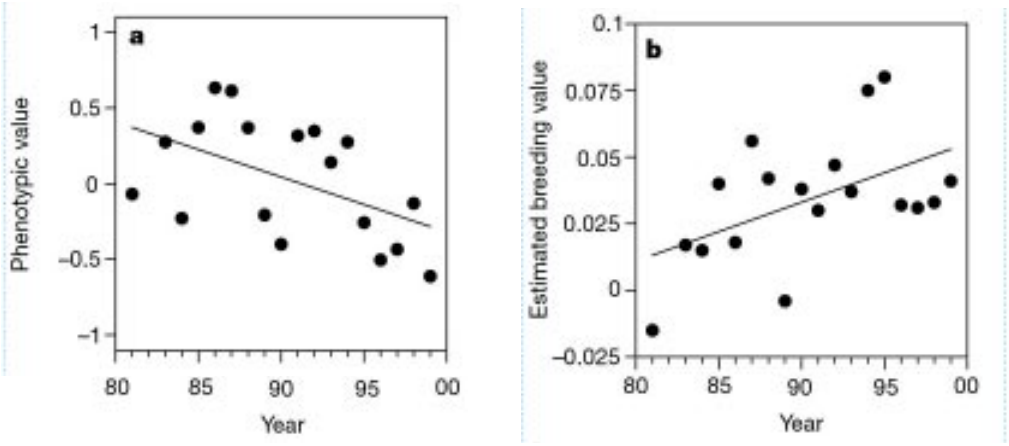


Figure 20.6. Body condition at fledging for a Gotland population of collared flycatchers between 1980 and 1999. **Left:** Mean phenotype. **Right:** Mean breeding value (the average of the PBVs). After Merilä et al. 2001b.

One striking example of apparent cryptic evolution is the study by Merilä et al. (2001b) on the Gotland population of collared flycatchers. These authors examined body condition (a measure of relative body weight) at fledging. They defined condition as the residual of the regression of body mass on tarsus (leg) length, and found that this trait has substantial heritable genetic variation (estimated h^2 = 0.30). Further, it appears to be under constant positive selection, with an average selection intensity of 0.23 (survivors are, on average, 0.23 standard deviations above the mean before selection). Despite the heritable nature of this trait coupled with strong positive selection, condition *declined* over time (Figure 20.6), with the regression of mean condition from 1981 to 1999 having a significant negative slope ($b = -0.036$ /year). Merilä et al. showed the covariance beween condition breeding value and fitness was nonzero, so that selection for condition occurs directly on the breeding values. Why then the apparent lack of response? As shown in Figure 20.6, the regression of predicted breeding values on time has a positive slope (b = 0.0022), with the population showing genetic improvement, despite the mean phenotype declining over time. The environmental component of condition has been declining over time, and at a rate faster than the genetic improvement, resulting in a net phenotypic decline. Merilä et al. (2001b) suggest this is likely attributable to reductions in the caterpillar food supply due to large-scale climatic trends.

Example 20.8. Another example of a negative environmental trend was offered by Larsson et al. (1998), who examined body size in the Barnacle Goose (*Branta leucopsis*). The natural colonization of the Baltic area of Sweden of this normally Arctic species started with a single breeding pair in 1971, followed by subsequent rapid increase in the population size. The authors studied the two largest Baltic colonies from 1984 to 1996. Head size and tarsus length were measured to extract a structural body size index, and larger females were found to have larger, and earlier, clutches (with larger eggs, resulting in more and heavier young than smaller females). Further, juvenile body mass was significantly correlated with post-

fledging survival. Both size measures are highly heritable, but average body size declined over the 13 year study period (by 0.7 and 0.5 standard deviations for head and tarsus length, respectively). The authors concluded (from a variety of evidence) that the environment had declined due to the growth of the colony. For example, birds from the smaller (and younger) colony studied were initially around a standard deviation larger than birds from the larger (and older) colony. In this case, the declining environment seems to be density-dependent effects on individual growth brought on by the overall success of the colony itself.

As Example 20.8 highlights, as organisms evolve they necessarily change at least part of their environment. Indeed, van Valen's (1973) **Red Queen hypothesis** states that organisms have to evolve just to stay where they are relative to the evolving biosphere around them. (The Red Queen, introduced in Lewis Carroll 's *Through the Looking Glass*, had to run just to stay in place.) Cooke et al. (1990) suggest that the lack of response to selection on clutch size in birds may have a red-queen style of explanation. While Price and Liou (1989) suggested that selection for clutch size was largely on nutritional state (Example 20.3), Cooke et al. counter that birds with the better quality territories have larger clutches and higher fitness, so that a component of selection for clutch size is selection to compete. Although competitive ability may increase over time, average territory quality, and hence average clutch size, remains relatively constant, and hence no response is seen. Under their model, if one could hypothetically compete ancestral and current populations for territories, current individuals, possessing higher breeding values for competitiveness, would win.

Antler Size in Red Deer: The Focal Trait is Not the Target of Selection

Free-living red deer (*Cervus elaphus*) on the Isle of Rum in Scotland are another well-studied natural population with a largely complete pedigree spanning several decades. Males fight to compete for mates, suggesting antler size as a potential trait under selection. Males shed antlers in the early spring, and given that antler shape is very individual-specific, cast antlers found in the field can easily be assigned to a specific stag. Kruuk et al. (2002) found that males with larger antlers had increased lifetime breeding success (total number of offspring), with a selection differential of $S = 0.445 \pm 0.094$ (scaled in phenotypic standard deviations). While body size (measured by leg length) also had an effect on lifetime breeding success, antler size still had a significant effect on fitness even after accounting for body size, with a (standardized, i.e., trait scaled to unit variance) selection gradient of $\beta = 0.44 \pm 0.18$. Antler size (measured as the mass of the annually-shed antlers) was heritable ($h^2 = 0.329 \pm 0.12$), and the breeder's equation would suggest a response of $R = h^2 S = 0.329 \cdot 0.445 = 0.146$ standard deviations per generation. Given a generation time of roughly 8 years and a standard deviation of 163 grams for antler mass, this suggests an expected change of roughly 2.3 grams/year. However, the average mass of antlers *declined* by 6.7 grams per year. One apparent reason for the decline was an environmental change due to increased population density over the study period, with antler size decreasing with increasing density.

Was this also a case of genetic change being masked by this environmental change? Apparently not. The REML estimate the genetic correlation between lifetime breeding success and antler size was not significant, -0.254 ± 0.289. The significant selection differential appears to be generated through selection on some environmental component of the trait. (More formally, on the residual value of phenotype after removal of the breeding value, which can include a genetic component, but not one transmissible from a parent to their offspring). The authors suggest that male fighting ability is, at least in part, a function of the nutritional condition of a male, and males with better nutritional value may be both better fighters and also grow larger antlers. Being better fighters, they have a greater lifetime

reproductive success as well as having higher antler mass.

Lower Heritabilities in Environments with Stronger Selection?

A more subtle implication of environmental change arises when genotype-environment interactions are present. As the environment changes, so can heritabilities, either due to changes in the environmental and/or the genetic variances (Hoffmann and Parsons 1991, 1997a,b; Hoffmann and Merilä 1999; Merilä and Sheldon 2001; Sgró and Hoffmann 2004; Charmantier and Garant 2005). Further, there are hints of (weak) trends in the direction of change. Data from wild vertebrate populations show increased heritabilities for morphological traits in more favorable environments, while traits more closely associated with fitness show no pattern (Merilä and Sheldon 2001, Charmantier and Garant 2005). Charmantier and Garant examined 46 traits, 38 of which showed no significant difference in heritabilities in good versus poor environments, but of the remaining 8 that were significant, all were higher in the more favorable environment. Roughly 65% of the traits showed decreased additive variation in less favorable environment, but most differences were not significant. Environmental variation also tended to increase under poor conditions. If there is a weak trend for lower h^2 in more unfavorable environments, this suggests less response in more stressed environments, exactly those likely to be under more selective pressures (Example 20.10).

These more recent surveys of wild populations are at odds with older laboratory experiments in *Drosophila*, which find higher additive genetic variance, and heritabilities, in stressed environments. While this observation may simply suggest that there are no general trends, it may also be a reflection of conditions in the lab versus the wild. They are also consistent with **Holloway's conjecture**, which states that adaptive traits should show higher additive variance, and reduced genetic correlations, in novel environments (Holloway et al. 1990). One could view the laboratory *Drosophila* experiments in this light, in that many of the artificially-imposed stresses created novel environments, especially for laboratory-adapted strains. Holloway's conjecture was supported by Robinson et al. (2009), working with Soay sheep, who found smaller genetic correlations in more favorable environment (measured by first year population-wide survival), and stronger correlations in poorer environments.

Example 20.9. Charmantier et al. (2004) examined chick tarsus length and body mass at fledging in a population of blue tits (*Parus caeruleus*) in three French populations, two on the island of Corsica and a third on the mainland. Their study followed roughly 8000 banded chicks from roughly 1200 individual broods representing three different habitats, which the authors were able to rank in quality. They found that poorer habitats showed weak selection to increase tarsus length and strong selection to increase body mass, while in good habitats there was no significant selection on either trait. Interestingly, heritability for body mass increased with habitat quality, with the lowest heritability occurring in the habitats experiencing the strongest selection for increased body weight. In such low quality environments strong selection would be at least partly countered by lower heritabilities, leading to a weaker response.

A similar situation was observed by Wilson et al. (2006) for birthweight Soay Sheep (*Ovis aries*). These authors used a random regression animal model (Volume 3), which allowed the estimation of maternal performance over a continuous environmental variable (here the population-wide neonatal mortality for a given year). Harsh environmental conditions generated strong selection for higher birthweights but also result in a lower genetic variance in this trait. More benign environments resulted in weaker selection but higher birthweight heritability.

Fitness Tradeoffs and Multivariate Constraints

Finally, our discussion of response in natural populations has neglected two extremely important issues that will be addressed later. The first is the estimation of fitness, examined in detail in Chapters 28 and 29. Operationally, lifetime (or total) fitness can be very difficult to measure, so often a component of fitness (such as viability or fecundity) is measured and assume to be a faithful indicator of total fitness. The concern is fitness **tradeoffs**, wherein a trait has a positive effect on one fitness component but a negative effect on another. As a result, its *net* effect on fitness is far less than would be suggested by either component separately. Likewise, one might imagine sex-specific tradeoffs, wherein a trait has a positive fitness effect in one sex and a negative effect in the other. Since selection tends to remove additive variation in fitness (Chapters 5, 6, 27), there is a widespread view that such tradeoffs likely occur. However, despite an obsession among ecologists and evolutionary biologists with such tradeoffs, they have proven rather elusive to find (Chapters 28, 29; Volume 3). A full discussion of this topic touches on the very vexing question of what maintains quantitative genetic variation (Chapter 27), so we return to it in later chapters.

The second issue, **multivariate constraints**, is a potential concern because all response is inherently multivariate. We focused in this chapter on the consequences of ignoring a correlated character, which is only one aspect of multivariate response. A critical observation is that each component trait under selection can have genetic variation ($h^2 \neq 0$), but the multivariate direction favored by response (the selection gradient vector $\boldsymbol{\beta}$) may contain little, or no, usable additive variation. To see this, suppose that only two traits are under selection, giving the response in trait one (from Equation 20.1b) as $\sigma^2(A_1)\beta_1 + \sigma(A_1, A_2)\beta_2$. Note that direct selection on the phenotype of trait one $(\beta_1 \neq 0)$ coupled with heritable variation ($\sigma^2(A_1) \neq 0$) is *not* sufficient to ensure that $R_1 \neq 0$. If trait one is sufficiently genetically correlated with trait two, then for certain combinations of selection (β_1, β_2), $\sigma^2(A_1)\beta_1 \simeq -\sigma(A_1, A_2)\beta_2$, and response in trait one is effectively zero. Geometrically, this means that the $\mathbf{G}$ matrix has eigenvectors (axes of variation) with eigenvalues close to (or at) zero, implying essentially no variation in that direction. If the angle between these nearly-null eigenvectors of $\mathbf{G}$ and the direction $\boldsymbol{\beta}$ favored by selection is very small, there is essentially no response in *any* of the component traits. In this setting, even if the phenotype is the focal target of selection, and the trait is heritable, the would still be essentially no response. We examine these issues in some detail in Volume 3. It is worth noting in such case that Robertson's theorem would predict that response is small to absent, with $\sigma(A_z, w)$ being very close to zero. Although this is a univariate treatment, the covariance of the breeding value of our focal trait with fitness would be nearly zero, as Robertson's theorem fully accounts for all of the genetically correlated traits that impact the focal trait-fitness covariance.

Literature Cited

Alatalo, R. V., L. Gustafsson, and A. Lunberg. 1990. Phenotypic selection on heritable size traits: environmental variance and genetic response. *Amer. Natl.* 135: 464–471. [20]

Blouin, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol. Evol.* 18: 503–511. [20]

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulson, M. H. H. Stevens, and J. S. S. White. 2009. Generalized linear mixed models: a practical guide for ecolgy and evolution. *Trends Ecol. Evol.* 24: 127–135. [20]

Charmantier, A., and D. Garant. 2005. Environmental quality and evolutionary potential: lessons from wild populations. *Proc. Roy. Soc. B* 272: 1415–1425. [20]

Charmantier, A., L. E. B. Kruuk, J. Blondel, and M. M. Lambrechts. 2004. Testing for microevolution in body size in three blue tit populations. *J. Evol. Biol* 17: 732–743. [20]

Charmantier, A., C. Perrins, R. H. McCleery, and B. C. Sheldon. 2006. Evolutionary response to selection in clutch size in a long-term study of the mute swan. *Amer. Natl.* 167: 453-465. [20]

Charmantier, A., and D. Réale. 2005. How do missassigned paternities affect the estimation of heritability in the wild? *Mol. Ecology* 14: 2839–2850. [20]

Cheverud, J. M., and W. P. J. Dittus. 1992. Primate population studies at Polonnaruwa. 2. Heritability of body measurements in a natural population of Toque macaques (*Macaca sinica*). *Amer. J. Primatology* 27: 145–156. [20]

Clément, V., B. Bibé, E. Verrier, J.-M. Elsen, E. Manfredi, J. Bouix, and E. Hanocq. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Genet. Sel. Evol.* 33: 369–395. [20]

Coltman, D. W. 2005. Testing marker-based estimates of heritability in the wild. *Mol. Ecol.* 14: 2593–2599. [20]

Coltman, D.W., P. O'Donoghue, J. T. Hogg, and M. Festa-Bianchet. 2005. Selection and genetic (co)variance in bighorn sheep. *Evol.* 59: 1372–1382. [20]

Coltman, D.W., P. O'Donoghue, J. T. Jorgenson, J. T. Hogg, C. Strobeck, and M. Festa-Bianchet. 2003. Undesirable evolutionary consequences of trophy hunting. *Nature* 426: 655–658. [20]

Conover, D. O. and E. T. Schultz. 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends Ecol. Evol.* 10: 248–252. [20]

Cooke, F., P. D. Taylor, C. M. Francis, and R. F. Rockwell. 1990. Directional selection and clutch size in birds. *Amer. Natl.* 136: 261–267. [20]

Csilléry, K., T. Johnson, D. Beraldi, R. Clutton-Brock, D. Coltman, B. Hansson, G. Spong, and J. M. Pemberton. 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* 173: 2091–2101. [20]

Darwin, C. 1871. *The descent of man and selection in relation to sex*. Murray, London. [20]

DiBattista, J. D., K. A. Feldheim, D. Garant, S. H. Gruber, and A. P. Hendry. 2009. Evolutionary potential of a large marine vertebrate: quantitative genetic parameters in a wild population. *Evol.* 63: 1051–1067. [20]

Etterson, J. R., and R. G. Shaw. 2001. Constraint to adaptive evolution in response to global warming. *Science* 294: 151–154. [20]

Fisher, R. A. 1958. *The genetical theory of natural selection*, 2nd Ed. Dover, New York. [20]

Frentiu, F. D., S. M. Clegg, J. Chittock, T. Burke, M. W. Blows, and I. P. F. Owens. 2008. Pedigree-free animal models: the relatedness matrix reloaded. *Proc. Roy. Soc. B* 275: 639–648. [20]

Garant, D., and L. E. B. Kruuk. 2005. How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol. Ecol.* 14: 1843–1859. [20]

Garant, D., L. E. B. Kruuk, R. H. McCleery, and B. C. Sheldon. 2004. Evolution in a changing environment: a case study with great tit fledging mass. *Am. Natl.* 164: E115–EE129. [20]

Garant, D., L. E. B. Kruuk, T. A. Wilkin, R. H. McCleery, and B. C. Sheldon. 2005. Evolution driven by differential dispersal within a wild bird population. *Nature* 433: 60–64. [20]

Garant, D., B. C. Sheldon, and L. Gustaffson. 2004. Climatic and temporal effects on the expression of secondary sexual characters: genetic and environmental components. *Evol.* 58: 634–644. [20]

Geldermann, H., U. Pieper, and W. E. Weber. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. *J. Anim. Sci.* 63: 1759-1768. [20]

Gienapp, P., E. Postma, and M. E. Visser. 2006. Why breeding time has not responsed to selection for earlier breeding in a songbird population. *Evolution* 60: 2381–2388. [20]

Gienapp, P., C. Teplitsky, J. S. Alho, J. A. Mills, and J. Merilä. 2008. Climate change and evolution: disentangling environmental and genetic responses. *Mol. Ecol.* 17: 167–178. [20]

Gosler, A. G., and D. G. C. Harper. 2000. Assessing the heritability of body condition in birds: a challenge exemplified by the great tit *Parus major*. L. (Aves). *Biol. J. Linn. Soc.* 71: 103–117. [20]

Grant, P. R., and B. R. Grant. 1995. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* 49: 241–251. [20]

Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 296: 707–711. [20]

Grant, P. R., and B. R. Grant. 2006. Evolution of character displacement in Darwin's finches. *Science* 313: 224–226. [20]

Green, A. J. 2001. Mass/length residuals: Measures of body condition or generators of spurious results? *Ecology* 82: 1473–1483. [20]

Guo, S.W. 1996. Variation in genetic identity among relatives. *Hum Hered.* 46: 61–70. [20]

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. Roy. Soc. B.* 275: 723–734. [20]

Hadfield, J. D., D. S. Richardson, and T. Burke. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* 15: 3715–3730. [20]

Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldo, and L. E. B. Kruuk. 2010. The misuse of BLUP in ecology and evolution. *Amer. Natl.* 175: 116–125. [20]

Haldane, J. B. S. 1954. The measurement of natural selection. *Proc. IX Internal. Cong. Genet.* 1: 480–487. [20]

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 9: 47–60. [20]

Henderson, C. R. 1988. Use of an average numerator relationship matrix for multiple-sire joinings. *J. Anim. Sci.* 66: 1614–1621. [20]

Heywood, J. S. 2005. An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. *Evolution* 59: 2287–2298. [20]

Hoffmann, A. A., and J. Merilä. 1999. Heritable variation and evolution under favorable and unfavourable conditions. *Trends Ecol. Evol.* 14: 96–101. [20]

Hoffmann, A. A., and P. A. Parsons. 1991. *Evolutionary genetics and environmental stress*. Oxford University Press, Oxford. [20]

Hoffmann, A. A., and P. A. Parsons. 1997a. *Extreme environmental change and evolution*. Cambridge University Press [20]

Hoffmann, A. A., and P. A. Parsons. 1997b. Consistent heritability changes under poor growth conditions. *Trends Ecol. Evol.* 12: 460 – 461. [20]

Holloway, G. J., S. R. Povey, and R. M. Sibly. 1990. The effect of new environment on adapted genetic architecture. *Heredity* 64: 323–330. [20]

Hõrak, P., R. Mänd, and I. Ots. 1997. Identifying targets of selection: a multivariate analysis of reproductive traits in the great tit. *Oikos* 78: 592–600. [20]

Israel C., and J. I. Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci.* 83: 181-187. [20]

Jakob, E. M., S. D. Marshall, and G. W. Uetz. 1996. Estimating fitness: A comparison of body condition indices. *Oikos* 77: 61–67. [20]

Jones, A. G., C. M. Small, K. A. Paczolt, and N. L. Ratterman. 2010. A practical guide to methods of parentage analysis. *Mol. Ecol. Resources* 10: 6–30. [30]

Jones, A. G., and W. R. Ardren. 2003. Methods for parentage analysis in natural populations. *Mol. Ecol.* 12: 2511–2523. [20]

Keller, L. F., P. R. Grant, B. R. Grant, and K. Petren. 2001. Heritability of morphological traits in Drawin's finches: misidentified paternity and maternal effects. *Heredity* 87: 325–336. [20]

Konigsberg, L. K., and J. M. Cheverud. 1992. Uncertain paternity in primate quantitative genetic studies. *Amer. J. Primatology* 27: 133–143. [20]

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the 'animal models'. *Phil. Trans. R. Soc. Lond B* 358: 873–890. [20]

Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E. Guniness. 2000. Heritablility of fitness in a wild mammal population. *Proc. Natl. Acad. Sci. USA* 97: 698–703. [20]

Kruuk, L. E. B., and J. D. Hadfield. 2007. How to separate genetic and environmental causes of similarty between relatives. *J. Evol. Biol.* 20: 1890–1903. [20]

Kruuk, L. E. B., and W. G. Hill. 2008. Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proc. R. Soc. B* 275: 593–596. [20]

Kruuk, L. E. B., J. Merilä, B. C. Sheldon. 2001. Phenotypic selection on a heritable size trait revisited. *Amer. Natl.* 158: 557–571. [20]

Kruuk, L. E. B., J. Slate, J. M. Pemberton, S. Brotherstone, F. Guinness, and T. Clutton-Brock. 2002. Antler size in red deer: heritability and selection but no evolution. *Evolution* 56: 1683–1695. [20]

Kruuk, L. E. B., J. Slate, and A. J. Wilson. 2008. New answers from old questions: The evolutionary quantitative genetics of wild animal populations. *Ann. Rev. Ecol. Evol. Syst*, 39: 525–548. [20]

Larsson, K., H. P. van der Jeugd, I. T. van der Veen, P. Forslund. 1998. Body size declines despite positive selection on heritable size traits in a Barnacle Goose population. *Evolution* 52: 1169–1184. [20]

Le Cren, E. 1951. The length-weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca fluviatilis*). *J. Anim. Ecol.* 66: 1504–1512. [20]

Lee, C., and E. J. Pollak. 1997. Influence of sire misidentification on sire x year interaction variance and direct-maternal genetic covariance for weaning weight in beef cattle. *J. Anm. Sci.* 75: 2858–2863. [20]

Levins, R. 1968. *Evolution in changing environments*. Princeton Univeristy Press, Princeton. [20]

Losos, J. B., T. R. Jackman, A. Larson, K. de Querioz, and L. Rodríguez-Schettino. 1998. Adaptive differentation following experimental island colonization in *Anolis* lizards. *Nature* 387: 70–73. [20]

Lynch, M. 1988. Estimation of relatedness by DNA fingerprinting. *Mol. Biol. Evol.* 5: 584–599. [20]

Mauricio, R., and L. E. Mojonnier. 1997. Reducing bias in the measurement of selection. *Trends Ecol. Evol.* 12: 433–436. [20]

Merilä, J., L. E. B. Kruuk, and B. C. Sheldon. 2001a. Natural selection on the genetic component of variance in body condition in a wild bird population. *J. Evol. Biol.* 14: 98–929. [20]

Merilä, J., L. E. B. Kruuk, and B. C. Sheldon. 2001b. Cryptic evolution in a wild bird population. *Nature* 412: 76–79. [20]

Merilä, J., and B. C. Sheldon. 2001. Avian quantitative genetics. *Curr. Ornothol.* 16: 179–255. [20]

Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001c. Explaining stasis: microevolutionary studies in natural populations. *Genetica* 112-113: 199–222. [20]

Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal models. *Livest. Prod. Sci.* 21: 87–100. [20]

Milner, J. M., S. D. Albon, A. W. Illius, J. M. Pemberton, and T. H. Clutton-Brock. 1999. Repeated selection on morphometric traits in the Soay sheep on St. Kilda. *J. Anim. Ecol.* 68: 472–488. [20]

Milner, J. M., S. Brotherson, J. M. Pemberton, and S. D. Albon. 2000. Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach. *J. Evol. Biol.* 13: 804–813. [20]

Milot, E., F. M. Mayer, D. H. Nussey, M. Boisvert, F. Pelletier, and D. Réale. 2011. Evidence for evolution in response to natural selection in a contemporary human population, *Proc. Natl. Acad. Sci. USA* 108: 17040–17045. [20]

Moore, A. J., and P. F. Kukuk. 2002. Quantitative genetic analysis of natural populations. *Nat. Rev. Gen.* 3: 971–978. [20]

Morrissey, M. B., and J. D. Hadfield. 2011. Directional selection in temporally replicated studies is remarkably consistent. *Evol.* 66: 435–442. [20]

Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder's equation in observational studies in natural populations. *J. Evol. Biol.* 23: 2277–2288. [20]

Morrissey, M. B., D. J. Parker, P. Korsten, K. M. Pemberton, L. E. B. Kruuk, and A. J. Wilson. 2012. The prediction of adaptive evolution: Empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution* 66: 2399–2410. [20]

Morrissey, M. B., A. J. Wilson, J. M. Pemberton, and M. M. Ferguson. 2007. A framework for power and sensitivity analysis of natural populations, and case studies in Soay sheep (*Ovis aries*). *J. Evol. Biol.* 20: 2309–2321. [20]

Morrissey, M. B., and A. J. Wilson. 2010. PEDANTICS: An R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol. Ecol. Res.* 10: 711–719. [20]

O'Hara, R. B., J. M. Cano, O. Ovaskainen, C. Teplitsky, and J. S. Alho. 2008. Bayesian approaches in evolutionary quantitative genetics. *J. Evol. Biol.* 21: 949–957. [20]

Oliehoek, P. A., J. J. Windig, J. A. N. van Arendonk, and P. Bijma. 2006. Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics* 173: 483–496. [20]

Ovaskainen, O., J. M. Cano, and J. Merilä. 2008. A Bayesian framework for comparative quantitative genetics. *Proc. Roy. Soc. B* 275: 669–678. [20]

Pemberton, J. M. 2008. Wild pedigrees: the way forward. *Proc. R. Soc. B* 275: 613–621. [20]

Pemberton, J. M. 2010. Evolution of quantitative traits in the wild: mind the ecology. *Phil. Trans. R. Soc. B* 365: 2431–2438. [20]

Perrins, C. M., and P. J. Jones. 1974. The inheritance of clutch size in the great tit (*Parus major*). *Condor* 76: 225–228. [20]

Postma, E. 2006. Implications of the differences between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol.* 19: 309–320. [20]

Postma, E., and A. Charmantier. 2007. What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. *J. Ornith.* 148.: 633–642. [20]

Postma, E., J. Visser, and A. J. van Noordwijk. 2007. Strong artifical selection in the wild results in predicted small evolutionary change. *J. Evol. Biol.* 20: 1823–1832. [20]

Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. *Science* 240: 798–799. [20]

Price, T., and L. Liou. 1989. Selection on clutch size in birds. *Amer. Natl.* 134: 950–959. [20]

Queller, D. C. 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139: 540–558. [20]

Quinn, J. L., A. Charmantier, D. Garant and B. C. Sheldon. 2006. Data depth, data completeness, and their influence on quantitative genetic estimation in two contrasting bird populations. *J. Evol. Biol.* 19: 994–1002. [20]

Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46: 616–626. [20]

Rauscher, M. D., and E. L. Simms. 1989. The evolution of resistance to heribivory in *Ipomoea purpurea.* I. Attempts to detect selection. *Evolution* 43: 563–572. [20]

Réale, D,. D. Berteaux, A. G. McAdam , and S. Boutin. 2003. Lifetime selection on heritable life-history traits in a natural population of red squirrels. *Evol.* 57: 2416–2423. [20]

Réale, D., M. Festa-Bianchet, and J. T. Jorgenson. 1999. Heritability of body mass varies with age and season in wild bighorn sheep. *Heredity* 83: 526–532. [20]

Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia retuculata*). *Science* 275: 1934–1937. [20]

Risch, N., and K. Lange. 1979. Application of a recombination model in calculating the variance of sib pair genetic identity. *Ann. Hum. Genet.* 43: 177–186. [20]

Ritland, K. 2000. Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* 9: 1195–1204. [20]

Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* 8: 95–108. [20]

Robertson, A. 1968. The spectrum of genetic variation. *In* R. C. Lewontin (ed.), *Population biology and evolution*, pp. 5–16. Syracuse University Press, Syracuse, NY [20]

Robinson, M. R., A. J. Wilson, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and L. E. B. Kruuk. 2009. The impact of environmental heterogeneity on genetic architecture in a wild population of soay sheep. *Genetics* 181: 1639–1648. [20]

Scheiner, S. M., K. Donohue, L. A. Dorn, S. J. Mazer, and L. M. Wolfe. 2002. Reducing environmental bias when measuring natural selection. *Evolution* 56: 2156–2167. [20]

Schulte-Hostedde, A. I., B. Zinner, J. S. Millar, and G. J. Hickling. 2005. Restitution of mass-size residuals: Validating body condition indice. *Ecology* 86: 155–163. [20]

Semlitsch, R. D., and H. M. Wilbur. 1989. Artificial selection for paedomorphosis in the salamander *Ambystoma talpoideum*. *Evolution* 43: 105–112. [20]

Sgró, C. M., and A. A. Hoffmann. 2004. Genetic correlations, tradeoffs and environmental variation. *Heredity* 93: 241–248. [20]

Sheldon, B. C., and H. Ellegren. 1999. Sexual selection resulting from extra-pair paternity in collared flycatchers. *Anim. Behav.* 57: 285–298. [20]

Sheldon, B. C., L. E. B. Kruuk, and J. Merilä. 2003. Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution* 57: 406–420. [20]

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecol. Let.* 12: 1261–1276. [20]

Stam, P. 1980. The distribution of the fraction of the genome identical-by-descent in finite random mating populations. *Genet. Res.* 35: 131–155. [20]

Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Ruahser, and R. Mauricio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Amer. Natl.* 160: 511–523. [20]

Suarez, B.K., T. Reich, and P. M. Fishman. 1979. Variability in sib pair genetic identity. *Hum. Hered.* 29: 37–41. [20]

Thériault, V., D. Garant, L. Bernatchez, and J. J. Dodson. 2007. Heritability of life-history tactics and genetic correlation with body size in a natural population of brook charr (*Salvelinus fontinalis*). *J. Evol. Biol.* 20: 2266–2277. [20]

Thomas, S. C. 2005. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural popualtions. *Phil. Trans. Roy. Soc. B* 360: 1457–1467. [20]

Thomas, S. C., D. W. Coltman, and J. M. Pemberton. 2002. The use of marker-based relationship information to estmete the heritability of body weight in a natural population: a cautionary tale. *J. Evol. Biol.* 15: 92–99. [20]

Thomas, S. C., and W. G. Hill. 2000. Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* 155: 1961–1972. [20]

Toro, M., C. Barragán, C. Ovilo, J. Rodrigañez, C. Rodriguez, and L. Silió. 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Cons. Gen.* 3: 309–320. [20]

Tosh, J. J., and J. W. Wilton. 1994. Effects of data structure on variance of prediction error and accuracy of genetic evaluation. *J. Anim. Sci.* 72: 2568–2577. [20]

van de Casteele, T., P. Galbusera, and E. Matthysen. 2001. A comparison of microsatellite-based pairwise relatedness estimators. *Mol. Ecol.* 10: 1539–1549. [20]

van Noordwijk, A. J. 1988. Two-stage selection in which the first stage only reduces the environmental variation in body size in the great tit. *XIX International Ornithological Congress*, 1408–1415. [20]

van Tienderen, P. H., and G. de Jong. 1994. A general model of the relation between phenotypic selection and genetic response. *J. Evol. Biol.* 7: 1–12. [20]

van Valen, L. 1973. A new evolutionary law. *Evol. Theory* 1: 1–30. [20]

Visscher, P. M., S. E. Medland, M. A. Ferreira, K. I. Morley, G. Zhu, B. K. Cornes, G. W. Montgomery, and N. G. Martin. 2006. Assumption free estimation of heritability from genome-wide identity-by descent sharing between full siblings. *PLoS Genet* 2: e41 [20]

Visscher, P., M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* 85: 2368–2375. [20]

Walling, C. A., J. M. Pemberton, J. D. Hadfield, and L. E. B. Kruuk. 2010. Comparing parentage inference software: reanalysis of a red deer pedigree. *Mol. Ecol.* 19: 1914–1928. [20]

Weldon, W. F. R. 1895. Attempt to measure the death-rate due to the selective destruction of *Carcinus moenas* with respect to a particular dimension. *Proc. Roy. Soc.* 57: 360–379. [20]

Wilson, A. 2008. Why h^2 does not always equal VA/VP. *J. Evol. Bio.* 21: 647–650. [20]

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. V. Mifsud, T. H. Clutton-Brock, and L. E. B. Kruuk. 2006. Environmental coupling of selection and heritability limits evolution. *PLoS Biology* 4: 1270–1275. [20]

Wilson, A. J., D. Réale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, and D. H. Nussey. 2010. An ecologist's guide to the animal model. *J. Animl Ecol.* 79: 13–26. [20]