Author for correspondence:
P. Bijma
e-mail: piter.bijma@wur.nl

The Price equation as a bridge between animal breeding and evolutionary biology

P. Bijma

Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700AH Wageningen, The Netherlands

PB, 0000-0002-9005-9131

The genetic response to selection is central to both evolutionary biology and animal and plant breeding. While Price's theorem (PT) is well-known in evolutionary biology, most breeders are unaware of it. Rather than using PT, breeders express response to selection as the product of the intensity of selection (i), the accuracy of selection (ρ) and the additive genetic standard deviation (σ_A); $R = i\rho\sigma_A$. In contrast to the univariate 'breeder's equation', this expression holds for multivariate selection on Gaussian traits. Here, I relate $R = i\rho\sigma_A$ to PT, and present a generalized version, $R = i_w\rho_{A,w}\sigma_A$, valid irrespective of the trait distribution. Next, I consider genotype–environment covariance in relation to the breeder's equation and PT, showing that the breeder's equation may remain valid depending on whether the genotype–environment covariance works across generations. Finally, I consider the response to selection in the prevalence of an endemic infectious disease, as an example of an emergent trait. The result shows that disease prevalence has much greater heritable variation than currently believed. The example also illustrates that the indirect genetic effect approach moves elements of response to selection from the second to the first term of PT, so that changes acting via the social environment come within the reach of quantitative genetics.

This article is part of the theme issue 'Fifty years of the Price equation'.

1. Introduction

Price's theorem (PT, [1,2]) has received very little attention in animal breeding, and many breeders appear to be unaware of it. This seems surprising for two reasons. First, PT provides a general expression for genetic change, and optimizing such change is the main objective of animal breeders. Second, the first term of PT,

$$R = \text{cov}(w,A), \tag{1.1}$$

also known as the Robertson–Price identity (RPI) or the 'secondary theorem of natural selection', which states that response to selection (R) in a trait equals the covariance of relative fitness (w) and the breeding value (A) for the trait, was first developed by an animal breeder in a study on the culling process of cows at dairy farms [3,4]. The response to selection is the cross-generation change in mean breeding value of the trait. The relative unfamiliarity of breeders with the RPI is illustrated by the citations of the Robertson [3] publication. Based on an inventory using Google Scholar, only two out of the first 50 publications citing this work originate from the field of animal breeding, while the vast majority relate to evolutionary biology. A similar observation applies to the 'breeder's equation' [5],

$$R = h^2 S, \tag{1.2}$$

with h^2 the narrow-sense heritability and S the selection differential, and its multivariate version in particular [6],

$$\mathbf{r} = \mathbf{GP}^{-1}\mathbf{s}, \tag{1.3}$$

which are primarily referred to in evolutionary genetic studies. In equation (1.3), $\mathbf{r}$ is a vector of responses to selection for each trait, $\mathbf{G}$ the matrix of additive genetic

(co)variances of all traits, **P** the matrix of phenotypic (co)variances of all traits and **s** a vector of (phenotypic) selection differentials for each trait. The breeder's equation follows directly from the regression of additive genetic values (breeding values) on trait values. While breeders are very familiar with regression theory for the prediction of response to selection, they mainly use selection-index theory [5,7,8], rather than the breeder's equation, the RPI or PT.

Two important reasons for which breeders are relatively unfamiliar with PT is that they select on a known 'index', and that they usually take a predefined proportion of top-ranking individuals to become parents of the next generation. The index is a linear combination of phenotypic values or estimates of breeding values for each selection candidate for all traits of interest, each trait being weighed according to its known (or perceived) importance [8]. Hence, breeders do not need to estimate the contribution of each trait to individual fitness (the 'selection gradient') because the index weights are known. Moreover, they also do not need to infer the overall strength of selection because the selected proportion is a breeder's decision. A third reason is the relatively limited role of the environment in livestock populations, where the environment is largely under human control and environmental confounders are often known. Hence, simpler approaches such as regression of breeding values on the selection index often suffice for breeders, but not always.

Here, I show how an approach based on PT can also be enlightening with more complex cases in animal breeding, for example in the presence of genotype–environment covariance or when interest is in emergent traits of populations, such as the prevalence of an infectious disease. I first connect the common animal breeding expression for response to selection to the RPI, and subsequently investigate response in the presence of genotype–environment covariance and in emergent traits, taking an approach building on the RPI. Because most traits relevant for animal breeding are highly polygenic, the per generation change in allele frequency due to selection is typically small. As a consequence, the per generation change in the average effects of alleles owing to non-additive effects is also small. For this reason, I will focus on the first term of PT (equation (1.1)) and not consider incomplete fidelity of transmission of the breeding value in the mathematical derivations. In the interpretation of results, however, I will illustrate how components of response can be shifted between the first and second terms of PT.

2. Price's theorem and expressions for response in animal breeding

The limited use of PT in animal breeding, as compared to the field of evolutionary biology, relates to the selection criterion and the measurement of selection in the two fields. The selection criterion in animal breeding is typically an 'index', a linear combination of phenotypes or estimates of breeding values for multiple traits (e.g. [8]),

$$I = b_1 z_1 + b_2 z_2 + \ldots, \tag{2.1}$$

where the b's are the index weights, and the z's the information sources. The information sources may include trait values of the selection candidate or of its relatives, but also estimates of breeding values for multiple traits. The index weights (b) are chosen by the breeder [8]. Selection in animal breeding is

measured by two parameters, the intensity of selection and the accuracy of selection, while no reference is made to fitness or selection gradients. Because of this difference in the parameterization of response to selection between animal breeding and evolutionary biology, the relationship between the RPI and the common expression used by breeders is not immediately clear.

In animal breeding, response to selection in any single trait, as a result of selection for an index (that potentially contains information sources on multiple traits), is commonly expressed as

$$R = i\rho_{I,A}\sigma_A, \tag{2.2}$$

where i is the intensity of selection, which is the selection differential in the index (S_I) expressed in standard deviation units (σ_I), $i = S_I/\sigma_I$, $\rho_{I,A}$ is the accuracy of selection, which is the correlation between the breeding value for the trait of interest and the index value I in the candidates for selection, and σ_A is the additive genetic standard deviation in the trait [7,9]. Equation (2.2) is equivalent to the multivariate breeder's equation (equation (1.3)) when the index includes only phenotypic trait values measured on the selection candidate itself. In most cases, however, the index also includes phenotypes measured on relatives of the selection candidate.

Equation (2.2) expresses response as a function of three clearly distinct parameters: the overall strength of selection (i), the precision of the relationship between the selection criterion (I) and the breeding value for the trait of interest ($\rho_{I,A}$), and the amount of heritable variation in the trait of interest (σ_A). In animal breeding practice, it is useful to distinguish these three parameters because they relate to rather different components of a breeding program; the heritable variation is a biological given, the selection intensity reflects the number of selection candidates evaluated per selected parent and depends primarily on the reproductive potential of the species, while the accuracy reflects the amount and quality of the phenotypic information collected to estimate the breeding values and calculate the index value for each selection candidate. Equation (2.2) follows directly from regression of the breeding values on the index (e.g. [10, p. 490]). Under the condition that the index and the breeding values are linearly related, for example when breeding values and phenotypic values are multivariate normal, equation (2.2) represents the expected change in mean breeding value given the selection differential applied to the index.

The relationship between equation (2.2) and the RPI is clarified in the following. From the RPI and the law of full covariance,

$$R = \text{cov}(A,w) = \text{E}[\text{cov}(A,w|I) + \text{cov}[\text{E}(A|I),\text{E}(w|I)],$$

where E denotes the expectation. Since selection is based solely on the index, there is no covariance between the breeding value and fitness of an individual conditional on its index value. So $\text{cov}(A,w|I) = 0$, and only the second term remains,

$$R = \text{cov}[\text{E}(A|I),\text{E}(w|I)].$$

Breeding values and index values are typically defined relative to a mean of zero. With means of zero and bivariate normality of A and I, conditional expectations follow from linear regression without the need for an intercept. Thus $\text{E}(A|I) = b_{A,I}I$, where $b_{A,I}$ is the regression coefficient of A on I. Thus

$$R = b_{A,I}\, \text{cov}[I,\text{E}(w|I)].$$

Next, $\text{cov}[I,\text{E}(w\,|\,I)] = \text{cov}(I,w)$, since any deviation of w from $\text{E}(w\,|\,I)$ must be independent from I. From the first term of PT, $\text{cov}(I,w)$ is the (phenotypic) change in the mean index value owing to selection, which by definition is the selection differential in the index, S_I. Thus

$$R = b_{A,I}\, S_I.$$

Replacing $b_{A,I}$ by its definition, expressing the selection differential as the product of the selection intensity and the standard deviation in index value, $S_I = i\sigma_I$, and multiplying the result by σ_A/σ_A yields

$$R = \frac{\text{cov}(A,I)}{\sigma_I^2}\, i\, \sigma_I\, \frac{\sigma_A}{\sigma_A} = \quad i\, \rho_{A,I}\, \sigma_A. \tag{2.3}$$

This result shows that the common expression for response to selection used by breeders (equation (2.2)) follows from the RPI under the condition that the selection index and the breeding value for the trait of interest are linearly related (e.g. bivariate normal), while the distribution of fitness is irrelevant (see, e.g. also Chapter 6 of [11]). In other words, the relationship between the selection index and fitness can take any shape, as long as there is no covariance between fitness and breeding value other than via the index; $\text{cov}(A,w\,|\,I) = 0$. A common practice in animal breeding is to select the individuals with an index value above a certain threshold as the parents of the next generation, which is known as 'truncation selection' [12].

While equation (2.2) is generally used in animal breeding, it is exact only when breeding values and the index are linearly related. However, an analogous equation can be found in which elements of equation (2.2) relate to breeding values and relative fitness, and which is exactly valid irrespective of the multivariate distribution of the breeding values and the selection criterion. (Note that the focus here is on the first term of PT, the RPI. We do not consider the second term, which relates to incomplete fidelity of transmission of the breeding value.) Starting from the RPI,

$$R = \text{cov}(A,w) = \rho_{A,w}\sigma_A\sigma_w,$$

where $\rho_{A,w}$ is the correlation between the breeding value for the trait of interest and individual fitness in the candidates for selection, σ_A is the additive genetic standard deviation for the trait of interest and σ_w is the (phenotypic) standard deviation in individual relative fitness. Using $\sigma_w = \sigma_w^2/\sigma_w$, where, from the first term of PT, the variance in relative fitness is equal to the selection differential in relative fitness ($\sigma_w^2 = S_w$, which is a phenotypic analogy of Fisher's [13] theorem of natural selection), it follows that

$$\sigma_w = \frac{S_w}{\sigma_w} = i_w,$$

which, by definition, is the intensity of selection for relative fitness (i_w). Hence, the selection intensity for relative fitness is equal to the standard deviation in relative fitness, always. On substitution in the expression for response to selection, it follows that response to selection in trait value is given by

$$R = i_w\rho_{A,w}\sigma_A. \tag{2.4}$$

This expression is valid for any type of selection and any distribution of breeding values, selection criterion and fitness. Hence, equations (2.2) and (2.4) are analogous, but equation (2.2) is equivalent to the breeder's equation (with the associated
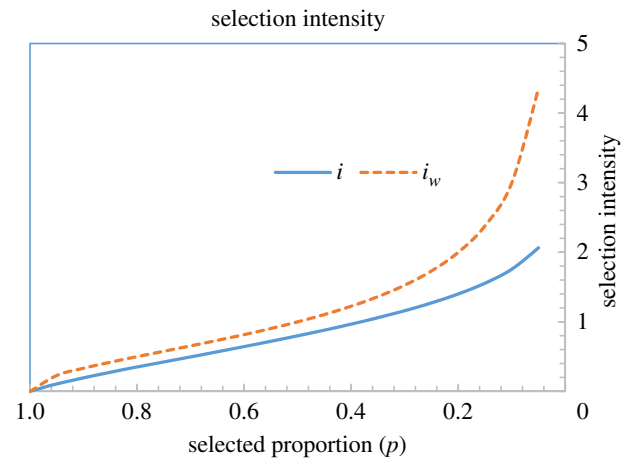

Figure 1. Selection intensity with truncation selection, for the trait (i), and for relative fitness (i_w), as a function of the selected proportion (p). (Online version in colour.)

linearity assumption), whereas equation (2.4) is the general RPI version that does not require this assumption.

Equation (2.4) is mathematically equivalent to the RPI, and thus valid for any type of (multivariate) selection and trait distribution. It is a reparameterization of the RPI, but it emphasizes different aspects of the response to selection from the RPI. The i_w measures the overall strength of selection in the population as the standard deviation in relative fitness values among individuals. The correlation between breeding value and relative fitness, $\rho_{A,w}$, measures the precision of the genetic relationship between fitness and the trait of interest, while the σ_A measures the heritable variation in the trait. Note that i_w is the selection intensity for fitness, $i_w = S_w/\sigma_w$, not for trait value. Hence, even when selection is for a single trait, so that individual fitness is fully determined by individual trait value, the value of i_w may differ from the classical selection intensity for trait value, $i = S_z/\sigma_z$ [12].

To illustrate the magnitude of the selection intensity for fitness, i_w, consider truncation selection as an example. With truncation selection, a proportion p of the available selection candidates becomes parents of the next generation, each leaving an equal number of descendants, while a proportion $1-p$ leaves no descendants. Thus individuals below the selection threshold have $w = 0$, and individuals above the threshold have $w = 1/p$, so that the mean relative fitness equals $\bar{w} = (1-p)\times 0 + p\times 1/p = 1$, as it should. Selection intensity in relative fitness equals $i_w = \sigma_w = \sqrt{(\overline{w^2} - \bar{w}^2)} = \sqrt{[p\times(1/p)^2 - 1^2]}$, giving

$$i_w = \sqrt{p^{-1} - 1}\,. \tag{2.5}$$

To illustrate the difference between the selection intensity for fitness and the classical intensity of selection, figure 1 shows both i_w and i, as a function of the selected proportion, for a normally distributed selection criterion (see, e.g. [12] for a table of selection intensities). Both measures of selection intensity take rather similar values, unless the selection is very strong (small p), in which case the selection intensity for fitness is clearly larger than the classical intensity of selection.

3. Genotype–environment covariance

One reason the RPI has received considerably more attention in evolutionary biology than in animal breeding is because

simpler expressions, such as the breeder's equation, may break down in natural populations, for example owing to a genotype–environment covariance [14,15]. Note that genotype–environment covariance refers to a non-zero covariance between the (additive) genetic component and the environmental component of individual phenotypes, $cov(A,E) \neq 0$, and should not be confused with genotype by environment interaction. A genotype–environment covariance may occur, for example, when genetically larger individuals are able to acquire better environments [10, p. 678]; giraffes with genetically longer necks, for example, may be able to access more food because longer necks allow them to reach leaves higher up in the trees. Genotype–environment covariance may occur also in livestock populations, for example when genetically better individuals are predominantly present in herds with better animal management. In most cases, such a covariance between genotype and herd environment is not a problem in animal breeding, since breeders can separate genetic effects from effects of the herd environment with the use of linear mixed models (particularly the so-called 'Animal Model'; [16]) because of considerable genetic connections between herds.

However, genetically better animals may also experience better environments on an individual level. In dairy cattle herds in European countries, for example, cows that produce more milk often also receive better nutrition, as a greater proportion of their feed consists of so-called 'concentrates'. Such a feeding strategy generates a positive genotype–environment covariance. Because data on individual nutrition are not available in most cases, breeders cannot control for the individual nutrition environment in models for breeding value and genetic parameter estimation. In this section, I investigate whether such a genotype–environment covariance indeed results in errors in genetic inference and in the prediction of response to selection in livestock populations, borrowing a model from the field of evolutionary biology ([14]; see also [10, p. 678]). I will assume a bivariate normal distribution of breeding values and phenotypic values, so that conditional expectations follow from linear regression.

Following van Tienderen and de Jong [14], individual phenotypic values are modelled as

$$z = A + E + e, \tag{3.1}$$

where A is the breeding value, E the effect of the macro environment on z and e the micro-environmental effect or residual. Thus $cov(A,e)$ is zero by definition. The issue is whether a non-zero $cov(A,E)$ indeed leads to errors in estimates of genetic parameters, such as additive genetic variance or heritability, and in predictions of response to selection.

In general, individual fitness may be a function of z, A and E [14],

$$w = \alpha + \beta_z z + \beta_A A + \beta_E E + \varepsilon, \tag{3.2}$$

where the betas represent selection gradients, being the partial regression coefficients of relative fitness (w) on the elements of the trait model (equation (3.1)). Because of multiple regression, β_z measures the causal effect of an individual's trait value on its fitness, β_A measures the causal effect of an individual's breeding value on its fitness (i.e. the component not acting via z or E), while β_E measures the causal effect of the macro-environmental value an individual experiences on its fitness.

Using the RPI, van Tienderen and de Jong [14] show that genetic response to selection equals (see also [10, p. 680])

$$R_A = \sigma_A^2 \beta_z + \sigma_A^2 \beta_A + cov(A,E)(\beta_z + \beta_E). \tag{3.3}$$

The first term of equation (3.3) is the breeder's equation. The middle term originates from the potential selection on genetically correlated traits, and is not of interest for our current objective. The last term of equation (3.3) represents the contribution of a genotype–environment covariance to the response to selection, and is the term of main interest here. Hence, even when selection is solely for individual phenotype, so that $\beta_z \neq 0$ while $\beta_A = \beta_E = 0$, presence of a genotype–environment covariance causes the response to selection to deviate from the breeder's equation,

$$R_A = \sigma_A^2 \beta_z + cov(A,E)\beta_z. \tag{3.4}$$

Thus equation (3.4) suggests a failure of the breeder's equation in the presence of a genotype–environment covariance. The issue is whether this result indeed implies an error in genetic inference and in the prediction of response to selection when ignoring a genotype–environment covariance, created for example by better feed for genetically better cows or more food for giraffes with genetically longer necks.

We use the following simple model to investigate this issue. Suppose cows with higher milk yield receive better nutrition, so that they experience E-terms that further increase their phenotypic value

$$E = \gamma + bz, \tag{3.5}$$

where γ is a fixed intercept and b a regression coefficient taking positive values smaller than one; $0 < b < 1$. A $b > 0$ indicates better nutrition for cows with higher yield, while $b < 1$ is required to obtain realistic phenotypic values.

Substitution into equation (3.1) and dropping the constant γ for brevity yields

$$z = A + bz + e. \tag{3.6}$$

Since A is a component of z, while z affects E, this situation generates a non-zero genotype–environment covariance, $cov(E,A) = cov(bz,A) = bcov(A + E + e,A) = b[\sigma_A^2 + cov(A,E)]$. Solving for $cov(A,E)$ yields

$$cov(A,E) = \frac{b}{1 - b} \sigma_A^2 . \tag{3.7}$$

Hence, better feed for better cows ($b > 0$) creates a positive genotype–environment covariance, as expected. For example if a unit increase in z causes E to increase by 0.1, then $cov(A,E)$ is equal to 11.1% of the additive genetic variance ($b = 0.1 \rightarrow cov(A,E) = 0.111 \sigma_A^2$).

Together, equations (3.4) and (3.7) suggest that a simple prediction of response based on the breeder's equation, $R_A = \sigma_A^2 \beta_z$, is incorrect when $b \neq 0$. However, the problem can be formulated in an alternative way, as follows. Solving equation (3.6) for z yields

$$z = \frac{A}{1 - b} + \frac{e}{1 - b} , $$

where the E-term has now disappeared. This result suggests the following reparameterization:

$$z = A' + e', \tag{3.8a}$$

with

$$A' = \frac{A}{1-b} \quad \text{and} \quad e' = \frac{e}{1-b}, \tag{3.8b}$$

$$\sigma_{A'}^2 = \frac{\sigma_A^2}{(1-b)^2}, \sigma_{e'}^2 = \frac{\sigma_e^2}{(1-b)^2} \quad \text{and} \quad \text{cov}(A', e') = 0 \tag{3.8c}$$

and

$$R_{A'} = \sigma_{A'}^2 \beta_z. \tag{3.8d}$$

In this reparametrized model, the original cov(A,E) has disappeared, and is absorbed into A and e, leading to A' and e'. What essentially happens here is that the expected contribution of E to z *that is caused by* cov(A,E) is added to A, leading to A'. In other words, given the genotype–environment covariance, the expected value of z given A is no longer equal to A, but equals E(z|A) = A/(1 − b) = A'. Similarly, the expected contribution of E to z that is caused by cov(e,E) is added to e, leading to e'. So we have essentially removed cov(A,E) by redefining the breeding value and the residual. See Appendix A for a short numerical illustration.

The expressions for $\sigma_{A'}^2$ and $\sigma_{e'}^2$ show that this redefinition leads to an increase of both the additive genetic variance and the residual variance when the original cov(A,E) is greater than zero (i.e. b > 0; equation 3.7). Moreover, since cov(A', e') = 0 for the reparametrized model, the response follows immediately from the breeder's equation. Thus the apparent failure of the breeder's equation has been remedied.

The above may seem mathematical trickery to conceal the failure of the breeder's equation owing to a genotype–environment covariance. However, it perfectly agrees with the fundamental statistical nature of the additive genetic effect (i.e. the breeding value) according to Fisher [13,17,18]. By definition, the breeding value is the expected trait value of an individual (apart from an intercept) given its allele counts at all loci, as obtained by least-squares regression (see also [12,19]). The A' indeed represents the expected phenotypic value of the individual given its alleles, as illustrated in Appendix A. This can also be seen from the substitution of equation (3.1) into equation (3.5),

$$E(z|A) = A + E(E|A) = A + bE(z|A),$$

and solving for E(z | A), which yields

$$E(z|A) = \frac{A}{1-b} = A'. \tag{3.9}$$

Hence, from a quantitative genetic perspective, the A' is the 'true', or the relevant, breeding value. Moreover, the A' is actually easier to use than A (see below).

The distinction between A and A' is of a conceptual, almost philosophical, nature. In general, the breeding value is necessarily a statistical construction defined for a specific population in a specific environment. The necessity of this construction arises because we cannot partition phenotypic values into functional genetic and environmental components; we can only partition the variance statistically. Hence, also in the absence of a genotype–environment covariance, the breeding value remains a statistical construction. In other words, there seems to be no justification to interpret the breeding value in the original model, A, as the true biological quantity.

The reparametrized breeding value, A', is not just of conceptual interest, but also relevant in animal breeding practice. When fitting an Animal Model [16,20],

$$\mathbf{z} = \boldsymbol{\mu} + \mathbf{Za} + \mathbf{e}, \tag{3.10}$$

to phenotypic data (**z**) from a population in which individuals experience an environment (E) that depends on their trait value, such as nutrition depending on milk yield (z), the resulting estimates of breeding values and additive genetic variance will refer to A' and $\sigma_{A'}^2$, not to A and σ_A^2. This occurs because the cov(A,E) contributes to the covariance between phenotypes of relatives, and therefore also surfaces in the estimate of additive genetic variance. Thus A' and $\sigma_{A'}^2$ are easy to obtain, because they are the estimates that result from the default model, whereas estimation of A and σ_A^2 would require correction for the dependency of E on z in the Animal Model (i.e. would require knowledge of b). Moreover, if animal management is the same in the next generation (i.e. animals are fed according to their *absolute* milk yield), then also response to selection follows from $\Delta\overline{A'}$, not from $\Delta\overline{A}$. In other words, if cows are fed according to their absolute milk yield, then this truly increases the additive genetic variance, which in turn translates into the greater response to selection than would have been observed in a population where such nutrition strategy is not applied. This makes sense, because with this feeding strategy an increase in yield over generations will be accompanied by an increase in E over generations. Hence, in that case, the response follows from equation (3.8d), not from equation (3.4). Note that equation (3.4) refers to the change in A, while equation (3.8d) refers to the change in A'; so these equations are not equivalent, as illustrated below.

A comparison of response to selection derived from each model illustrates the conceptual difference. For the original model, the response in breeding value follows from equations (3.4) and (3.7),

$$R_A = \sigma_A^2 \beta_z + \text{cov}(A,E)\beta_z$$

$$= \sigma_A^2 \beta_z + \frac{b}{1-b}\sigma_A^2 \beta_z$$

$$= \frac{1}{1-b}\sigma_A^2 \beta_z. \tag{3.11a}$$

Note this is response in A, not A'. If the feeding strategy stays the same, the genetic change in A increases the mean z, which in turn translates in an increase in E (owing to the feeding strategy), and this increase in E also adds to the response. Hence, given the nutrition environment, the total response in phenotypic value owing to genetic selection consists of two parts, R = R_A + R_E. The response in E owing to a change in mean breeding value follows from the regression of E on A,

$$R_E = b_{E,A} R_A = \frac{\text{cov}(A,E)}{\sigma_A^2} R_A = \frac{b}{1-b}\frac{1}{1-b}\sigma_A^2 \beta_z$$

$$= \frac{b}{(1-b)^2}\sigma_A^2 \beta_z . \tag{3.11b}$$

Total response therefore equals

$$R = R_A + R_E = \left(\frac{1}{1-b} + \frac{b}{(1-b)^2}\right)\sigma_A^2 \beta_z$$

$$= \frac{1}{(1-b)^2}\sigma_A^2 \beta_z. \tag{3.11c}$$

Hence, in this perspective, the total response is the sum of the change in mean breeding value plus a response in the macro environment that, given the nutrition strategy, is caused by the genetic change. In the context of PT, the R_E would be part of the second term of PT.

For the reparametrized model, the response follows directly from the breeder's equation (equation 3.8*d*),

$$R_{A'} = \sigma_{A'}^2 \beta_z = \frac{1}{(1-b)^2} \, \sigma_A^2 \beta_z. \tag{3.12}$$

This result is identical to equation (3.11*c*). Thus $R = R_A + R_E = R_{A'}$. Hence, response predictions from both models agree, but follow from a different perspective. In contrast to the original parameterization, R_E is part of the first term of PT in the reparametrized model.

This example illustrates that a genotype–environment covariance does not necessarily render the breeder's equation invalid. In other words, given that cows are fed according to their *absolute* milk yield, equation (3.8d) yields a valid prediction of response to selection, whereas equation (3.4) would result in under-prediction because it lacks the R_E component. The key issue here is whether the genotype–environment covariance is present across generations, or whether it is solely a within-generation phenomenon.

If the genotype–environment covariance is present across generations, then a change in mean breeding value over time will be accompanied by a change in the macro environment. For this situation, response to selection is given by $R_{A'}$. This is for example the case when cows are fed according to their absolute milk yield, so that an increase in population average milk yield over generations will be accompanied by an improvement of nutrition. A corresponding (hypothetical) example for a natural population would be the case where giraffes with longer necks can acquire more food because they can reach higher up in the trees. In that case, an increase in mean neck length would be accompanied by an increase in the 'nutrition environment', as an increasingly large proportion of the trees can be used for food.

By contrast, if the genotype–environment covariance is solely present within a generation, then a change in mean breeding value from one generation to the next does not go together with a corresponding change in the macro environment. For this situation, response to selection is given by R_A. This occurs, for example, when genetically larger individuals are able to occupy *relatively* better environments in natural populations, but there is no improvement of the environment over generations. An example for animal breeding would be the situation where better cows get better nutrition within a generation, but the overall average nutrition is not adjusted to the increasing milk yield over generations.

Response to selection for the situation where the genotype–environment covariance is present solely within generation (i.e. R_A) can be modelled in at least three different ways. First, one could predict R_A using equation (3.4). The practical challenge, in that case, is to estimate cov(A,E) (i.e. the b in this example). Alternatively, R_A could be found by considering A' and using the full version of PT. The first term of PT would be equal to equation (3.12), while the second term of PT would be equal to $-R_E$ (so the negative of equation (3.11b)), so that the total response from both terms of PT equals $R_{A'} - R_E = R_A$. The $-R_E$ term represents infidelity of transmission, because the cov(A,E) that was present within a generation is not present across generations. In this approach, the practical challenge is to quantify the second term of PT. A third approach would be to use a model with A' that includes both direct and indirect genetic effects (IGE). For example, when genetically larger individuals are able to occupy *relatively* better environments, then such individuals have a negative

impact on the environment acquired by their population members. In this perspective, the lack of an increase in the macroenvironment over generations relates to competition between individuals within a generation, which surfaces as a negative direct–indirect genetic correlation in IGE models. Such a negative direct–indirect genetic correlation reduces the heritable variation in the trait [21–23]. The negative direct–indirect genetic correlation will result in an unfavourable correlated response in the IGE, analogous to the 'deterioration of the environment' in the context of Fisher's fundamental theorem of natural selection [13,24,25]. This negative response in IGE corresponds to the $-R_E$ term mentioned above.

4. Emergent traits and heritable variation in infectious disease prevalence

The partitioning of individual phenotypic values into individual breeding values and remaining non-additive genetic terms, $z = A + e$, is foundational to quantitative genetics, and to animal breeding in particular. Some traits, however, cannot be attributed to single individuals, but are a property of a group of individuals or of a 'society'. Obvious examples in natural populations are the size of a termite colony or the number of prey caught by a hunting pack. Since single individuals do not have phenotypes for such 'emergent traits', one cannot simply partition phenotypic values into individual breeding values and residuals. Nevertheless, emergent traits must have evolved and may still show heritable variation. Moreover, the breeding values of single individuals are relevant for response to selection in the emergent trait, because the transmission of genes to the next generation is from single parents to their offspring, not from groups to groups. The RPI provides a natural starting point to find the genetic variation and response to selection for such emergent traits [23].

Emergent traits may seem absent in livestock populations because the environment is largely controlled by the farmer or the breeder. Indeed, emergent traits have received little attention in animal breeding. The prevalence of an endemic infectious disease in a population, however, provides a typical example of an emergent trait, because it arises from a population process of disease transmission between individuals. Thus the infectious disease status of an individual (0/1, indicating non-infected versus infected) does not represent the sum of the breeding value of the individual and a residual, but it is the result of a disease transmission process in the population, and thus also depends on other individuals and on the genes of those individuals.

Reducing the prevalence of endemic infectious diseases in livestock is highly relevant for animal health and welfare, and also for human health in the case of zoonosis [26]. The classical picture, however, is that infectious disease status (0/1) shows low heritability, which suggests that the potential for genetic selection against infectious diseases is limited. (Note that disease prevalence, i.e. the fraction of the population that is infected at any point in time, is simply the mean of individual disease status; so the response in prevalence is the same as response in disease status.) The classical approach, however, ignores that the prevalence of an endemic disease is an emergent trait of the population [27,28].

In the following, the trait of interest is the equilibrium prevalence of an endemic infectious disease in a population, and the focus is on the heritable variation that determines the potential
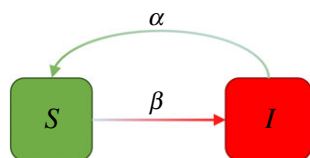
Figure 2. The susceptible–infected–susceptible (SIS) compartmental model for an endemic disease. S is the number of non-infected and susceptible individuals, I is the number of infected and infectious individuals, and $S + I = N$. β is the transmission rate parameter and α the recovery rate parameter. The expected transmission rate equals $\beta SI/N$, and the expected recovery rate equals αI. At equilibrium, $\beta SI/N = \alpha I$, so that $I/N = 1 - \alpha/\beta$, where $\alpha/\beta = 1/R_0$ so that $P = 1 - 1/R_0$ (figure 3). (Online version in colour.)

Figure 3. Prevalence as a function of R_0 for an endemic disease following a SIS-model (figure 2), $P = 1 - 1/R_0$. (Online version in colour.)

response to selection in the prevalence. I show how the quantity relevant for response to selection can be found, and how this quantity differs from the classical breeding value for individual disease status. In contrast to the convention in quantitative genetics, I will not express breeding values as a deviation from their mean, because this would introduce unnecessary mathematical detail. Moreover, to minimize mathematical detail, I will consider genetic variation among individuals only in their susceptibility to become infected. Thus I will ignore potential genetic variation in the propensity of individuals to infect others (infectivity), and in the rate at which individuals recover from infection. In this section I will follow the notational convention common in epidemiology, so symbols' meanings may differ from their usage in the above text.

Consider a simple model of an endemic disease, where individuals can either be non-infected and susceptible (S), or both infected (I) and infectious, the so-called compartmental SIS-model ([29]; figure 2). Thus there are two classes of individuals (compartments): susceptible individuals and infected individuals. Moreover, there are two processes: disease transmission, $S \to I$, where a susceptible individual becomes infected, and recovery, $I \to S$, where an infected individual recovers from the disease and becomes susceptible again. When individuals become infected they are also immediately infectious, and when infected individuals recover, they become immediately susceptible again. The total number of individuals in the population equals $N = S + I$, S denoting the number of susceptible individuals and I the number of infected individuals. Thus the symbols S and I are used both to indicate individual disease status and the number of individuals with this disease status, as common in epidemiology.

Quantitative epidemiology shows that the prevalence of an infectious disease, which is the fraction of the population that is infected, $P = I/N$, depends on the so-called basic reproduction number (R_0) of the disease. R_0 is the number of individuals who become infected by a typical infected individual in an otherwise fully susceptible population, and is a property of the population [30–32]. When $R_0 > 1$, an infectious individual can infect more than one new individual, so that the disease can persist in the population.

The prevalence reaches an equilibrium value when an infected individual on average infects a single other individual. This occurs when the product of R_0 and the fraction of contact individuals who are susceptible is equal to 1; $R_0 (1 - P) = 1$. For example, when $R_0 = 3$ an infectious individual could in principle infect three other individuals. Howe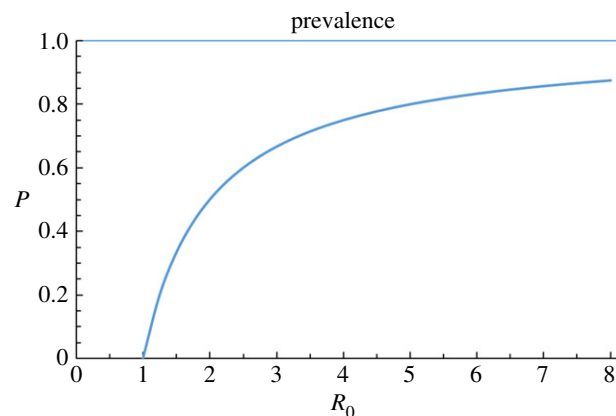ver, when only one-third of their contact individuals are susceptible, meaning $1 - P = 1/3$, then the effective reproduction number equals $3 \times 1/3 = 1$. Hence, at $1 - P = 1/3$ the number of infected individuals stays constant because an infected individual is on average replaced by a single new infected individual, so that an equilibrium occurs at $P_{eq} = 1 - 1/3 = 2/3$. The equilibrium prevalence, therefore, equals [32]

$$P_{eq} = 1 - \frac{1}{R_0}. \tag{4.1}$$

Hence, the equilibrium prevalence is determined by R_0 (figure 3). The equilibrium is reached quickly (i.e. on much shorter time scales than a generation) and actual prevalence tends to fluctuate around the equilibrium because of random perturbations.

Because the equilibrium prevalence is determined by R_0 (equation (4.1)), response to selection, i.e. the genetic change in prevalence from one generation to the next, follows from the genetic change in R_0. Thus, to measure the value of a selection candidate with respect to response to selection, we should base this measure on its genetic impact on R_0. Hence, based on equation (4.1), we may define an individual breeding value for prevalence based on an individual's additive genetic impact on R_0,

$$A_{P,i} = -\frac{1}{A_{R_0,i}}, \tag{4.2}$$

where $A_{R_0,i}$ is the additive genetic contribution of individual i to R_0, or the 'breeding value for R_0'. (See [28] for an example for an epidemic disease. While there are issues here with nonlinearity and interactions, these may be addressed by considering breeding values as sums of marginal effects of alleles, similar to breeding values for any trait [17,18]. This is not of interest for the current objective.) Thus A_p is the quantity relevant for response to selection, i.e. for the change in equilibrium prevalence from one generation to the next,

$$\Delta P_{eq} = \Delta \overline{A_P} = \text{cov}(A_p, w). \tag{4.3}$$

In equation (4.2), I have defined the breeding value as the value of an individual with respect to response to selection. In other words, since $\Delta P_{eq} = \Delta \overline{A_P}$, A_p is the measure of breeding value that is relevant for response to selection. The key question now is whether the ordinary breeding value for individual disease status, meaning the value that we estimate when analysing individual disease status data with, e.g. an

animal model, is the same as the A_P defined in equation (4.2), or whether it is something completely different. To answer this question, I now derive the expected value of the ordinary breeding value for individual disease status, $y_i = 0,1$ indicating non-infected versus infected. (But note that prevalence is simply the mean of individual disease status, $\bar{y} = P$.) Because I consider genetic variation in susceptibility to infection only, in the following the transmission rate parameter between a pair of individuals ij (β_{ij}) depends only on the susceptibility of the recipient individual i, not on the infectivity of the donor individual j; so I will use $\beta_{ij} = \beta_i$, i denoting the susceptible recipient individual.

The genetic component of an individual's disease status follows from the expectation of y_i given its genes. The expectation of y_i is equal to the expected fraction of the time that i is infected (i.e. that $y_i = 1$, since $y_i = 0$ otherwise). This fraction is equal to the mean duration (Δt) of the infected status divided by the mean duration of a full SIS cycle (figure 2),

$$A_{y,i} = E(y_i | \text{genes of } i) = \frac{E(\Delta t_{I \to S,i})}{E(\Delta t_{I \to S,i}) + E(\Delta t_{S \to I,i})} \ .$$

Assuming approximately constant transmission and recovery rates in the endemic equilibrium, the mean duration of a status follows from the exponential distribution, and equals the reciprocal of the corresponding rate [33]. Thus the mean duration of the infected status is the reciprocal of the recovery rate, α,

$$E(\Delta t_{I \to S,i}) = \frac{1}{\alpha},$$

(note, we assume the absence of genetic variation in recovery rate, so $\alpha_i = \alpha$). For example, if the recovery rate is 0.5/day, then the mean recovery time is 2 days. Similarly, the mean duration of the susceptible status is the reciprocal of the rate with which the individual becomes infected,

$$E(\Delta t_{S \to I,i}) = \frac{1}{\beta_i I / N},$$

where β_i is the transmission rate parameter for individual i, which reflects its susceptibility, and I/N is the fraction of its contact individuals who are infectious. The mean duration of a full SIS cycle, therefore, equals $1/\alpha + 1/(\beta_i I/N)$. Hence, the ordinary breeding value for individual disease status, i.e. the expected phenotypic value (y) of an individual given the average effects of its genes, equals

$$A_{y,i} = \frac{1/\alpha}{1/\alpha + 1/(\beta_i I/N)} \tag{4.4}$$

where β_i is to be interpreted as the additive genetic value for individual i, including its mean.

To investigate whether equations (4.2) and (4.4) are identical, equation (4.4) can be expressed in terms of R_0, by relating α and β to the breeding value for R_0. Recall that R_0 is the number of individuals infected by a single infected individual over their entire infectious lifetime in a fully susceptible population. Hence, R_0 is the product of the mean duration of the infectious period ($1/\alpha$) and the mean transmission rate parameter (β; $R_0 = \beta/\alpha$, [33]). Thus the individual breeding value for R_0 may be expressed as

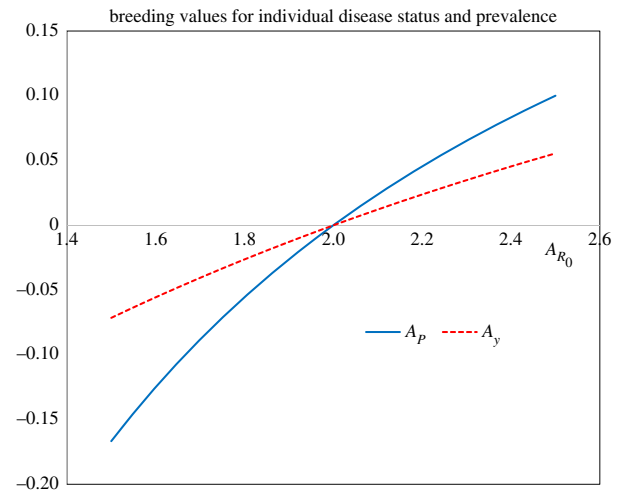$$A_{R_0,i} = \frac{\beta_i}{\alpha}. \tag{4.5}$$

Figure 4. Breeding value relevant for response to selection (A_P) and breeding value for individual disease status (A_y), as a function of the breeding value for R_0 (A_{R_0}), for a population with $R_0 = 2$ so that disease prevalence is 50%. The A_P and A_y are shown as a deviation from their mean, while the breeding value for R_0 includes the mean. (Online version in colour.)

Multiplying numerator and denominator of equation (4.4) by β_i and substituting $\beta_i/\alpha = A_{R0,i}$ and $I/N = P$, yields

$$A_{y,i} = \frac{A_{R_0,i} P}{A_{R_0,i} P + 1}. \tag{4.6}$$

Equation (4.6) is the expected value of an individual's disease status given the average effects of its genes, and represents the breeding value that is (implicitly) estimated when fitting a mixed model to data on individual disease status, which is the common approach.

A comparison of equations (4.2) and (4.6) shows that the breeding value for individual disease status ($A_{y,i}$) is markedly different from the breeding value that determines response to selection ($A_{P,i}$); $A_{y,i}$ is approximately only a fraction P of $A_{P,i}$. This is illustrated in figure 4 for a disease with an equilibrium prevalence of 50% (i.e. $R_0 = 2$). Hence, $A_{y,i} \cong P A_{P,i} = 0.5 A_{P,i}$ in this example. Thus the breeding value for response in prevalence takes a considerably wider range of values. This means that the ordinary breeding values for disease status strongly underestimate the prospects for genetic change by artificial or natural selection, particularly when the prevalence is small.

The difference between A_P and A_y originates from positive feedback via the 'social environment'. Less susceptible individuals, for example, are not only less likely to become infected themselves, but they are also less likely to infect other individuals in the population simply because they are less often infected themselves. This mechanism leads to a reduction in exposure to the pathogen for the entire population, resulting in lower prevalence. The opposite holds for individuals who are more susceptible than average. This positive feedback via the social environment is included in A_P, but not in the ordinary breeding value for individual disease status A_y.

In the above example, the response in prevalence followed from the RPI (equation (4.3)), without a contribution of the second term of PT. By starting from the expression for the equilibrium prevalence as a property of the population (equation (4.1)), the effect of a change in the environment (i.e. changing exposure to the pathogen via infectious population members) on individual disease status is included implicitly.

In other words, I started from an expression for prevalence that follows from the disease transmission process in the population, rather than treating prevalence as the average value of the disease status of individual animals. For this reason, the breeding value for response to selection (A_P) represents a so-called total breeding value [21,23], which includes both the direct genetic effect of disease susceptibility on the disease status of the focal individual itself as well as the IGE of the focal individual's disease susceptibility on the disease statuses of its population members.

In an alternative approach, one could have treated the prevalence as the mean of individual disease status values, and derived response to selection using both terms of PT. The first term of PT (RPI) would then reflect the mean change in individual disease status owing to individual genes in a constant environment (i.e. the same exposure to infected population members as before selection) as given by $\Delta \overline{A_y}$, while the second term of PT would reflect the change in prevalence owing to a change in the environment (i.e. a reduction of the fraction of infected conspecifics leading to lower exposure to the pathogen after selection). For the present case, however, the first approach is considerably simpler.

In general, the IGE approach [34–36] may be interpreted as moving components of the response to selection from the second term of PT to the first term of PT. By extending the definition of breeding value to include the social genetic effects, they are shifted from the 'environment' to the breeding value so that their contribution to response becomes part of $\mathrm{cov}(A,w)$. In many cases this is convenient, because IGE can be estimated together with direct genetic effects in an Animal Model framework [37–39], and this approach also reveals the total heritable variation that is available for response to selection (the σ_A to be used in equations (2.2) and (2.4); [23]).

5. Conclusion

In this article, I have connected PT to animal breeding. First, I have connected the common expression for response to selection used by breeders to PT, and presented an exact version of that expression, formulated in terms of fitness (equation (2.4)). Next, I explored the problem of genotype–environment covariance, showing that response to selection can be partitioned into two terms (R_A and R_E) that relate to the definition of the breeding value and to the partitioning of response between the first and second terms of PT. The results show that a genotype–environment covariance leads to failure of the breeder's equation when the genotype–environment covariance is present only within generations but not across generations. Finally, I investigated the breeding value and response to selection for the prevalence of an endemic infectious disease, as an example of an emergent trait. By starting from an expression for the equilibrium prevalence of the disease as a consequence of the transmission process in the population, instead of focussing on individual trait values, I implicitly moved elements of response to selection from the second to the first term of PT. This is valuable, because it removes the 'incomplete fidelity of transmission', and thus captures more of the response in terms of classical quantitative genetic parameters, such as the breeding value and heritable variation. These examples illustrate the power of PT as a starting point to understand the response to selection.

Appendix A – Numerical illustration of genotype–environment covariance owing to the feeding regime

Suppose the average cow produces 20 kg of milk per day. We put the trait average in E, so that the mean breeding value is zero. Hence, $\bar{E} = \gamma = 20$. Suppose farmers feed cows according to their milk yield with a constant of proportionality of $b = 0.1$. Now consider a cow with a breeding value of $A = +1$ kg of milk. The initially expected milk yield of the cow equals $20 + 1 = 21$ kg. The farmer observes that the cow is better than average, and adjusts nutrition according to $E = \gamma + b(z - \bar{z}) = 20 + 0.1 \times (21 - 20) = 20.1$. Thus the expected milk yield of the cow increases to $20.1 + 1 = 21.1$. The farmer again adjusts nutrition, giving $E = 20 + 0.1 \times (21.1 - 20) = 20.11$, so that expected milk yield of the cow increases to $20.11 + 1 = 21.11$, etc. This series converges to an expected milk yield of 21.111… Hence, the redefined breeding value, being the expected value of z given A, measured as a deviation from the average, equals 1.111…, which is equal to $A' = 1/(1 - b) = 1/0.9$.

References

1. Price GR. 1970 Selection and covariance. *Nature* **227**, 520–521. (doi:10.1038/227520a0)
2. Price GR. 1972 Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485–490. (doi:10.1111/j.1469-1809.1957.tb01874.x)
3. Robertson A. 1966 A mathematical model of the culling process in dairy cattle. *Anim. Sci.* **8**, 95–108. (doi:10.1017/S0003356100037752)
4. Robertson A. 1968 The spectrum of genetic variation. In *Population biology and evolution* (ed. RC Lewontin), pp. 5–16. Syracuse, NY: Syracuse University Press.
5. Lush J. 1937 *Animal breeding plans*. Ames, IA: Iowa State College Press.
6. Lande R, Arnold SJ. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226. (doi:j.1558-5646.1983.tb00236.x)
7. Hazel LN. 1943 The genetic basis for constructing selection indexes. *Genetics* **28**, 476–490.
8. Hazel LN, Dickerson GE, Freeman AE. 1994 The selection index—then, now, and for the future. *J. Dairy Sci.* **77**, 3236–3251. (doi:10.3168/jds.S0022-0302(94)77265-9)
9. Lush J. 1994 *The genetics of populations. Iowa agriculture and home economics experiment station, college of agriculture*. Ames, IA: Iowa State University.
10. Walsh B, Lynch M. 2018 *Evolution and selection of quantitative traits*. London, UK: Oxford University Press.
11. Rice SH. 2004 *Evolutionary theory: mathematical and conceptual foundations*. Sunderland, MA: Sinauer Associates.

12. Falconer DS, Mackay TFC. 1996 *Introduction to quantitative genetics*. Essex, UK: Longman Group.

13. Fisher RA. 1930 *The genetical theory of natural selection*. Oxford, UK: Clarendon.

14. van Tienderen PH, de Jong G. 1994 A general model of the relation between phenotypic selection and genetic response. *J. Evol. Biol.* **7**, 1–12. (doi:10.1046/j.1420-9101.1994.7010001.x)

15. Hadfield JD. 2008 Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. B* **275**, 723–734. (doi:10.1098/rspb.2007.1013)

16. Quaas RL, Pollak EJ. 1980 Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* **51**, 1277–1287. (doi:10.2527/jas1981.5161277x)

17. Fisher RA. 1918 XV.—the correlation between relatives on the supposition of Mendelian inheritance. *Tran. R. Soc. Edinburgh* **52**, 399–433. (doi:10.1017/S0080456800012163)

18. Fisher RA. 1941 Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63. (doi:10.1111/j.1469-1809.1941.tb02272.x)

19. Lynch M, Walsh B. 1998 *Genetics and analysis of quantitative traits,* vol. 1, pp. 535–557. Sunderland, MA: Sinauer.

20. Mrode RA. 2014 *Linear models for the prediction of animal breeding values*. Wallingford, UK: Cabi.

21. Bijma P, Muir WM, Van Arendonk JA. 2007 Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics* **175**, 277–288. (doi:10.1534/genetics.106.062711)

22. Wilson AJ, Morrissey MB, Adams MJ, Walling CA, Guinness FE, Pemberton JM, Clutton-Brock TH, Kruuk LEB. 2011 Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, *Cervus elaphus*. *J. Evol. Biol.* **24**, 772–783. (doi:10.1111/j.1420-9101.2010.02212.x)

23. Bijma P. 2011 A general definition of the heritable variation that determines the potential of a population to respond to selection. *Genetics* **189**, 1347–1359. (doi:10.1534/genetics.111.130617)

24. Price GR. 1972 Fisher's fundamental theorem made clear. *Ann. Hum. Genet.* **36**, 129–140. (doi:10.1111/j.1469-1809.1972.tb00764.x)

25. Bijma P. 2010 Fisher's fundamental theorem of inclusive fitness and the change in fitness due to natural selection when conspecifics interact. *J. Evol. Biol.* **23**, 194–206. (doi:10.1111/j.1420-9101.2009.01895.x)

26. EFSA Panel on Animal Health and Welfare (AHAW). 2012 Scientific Opinion on Review of the European Union Summary Report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2009 and 2010 specifically for the data related to bovine tuberculosis, Echinococcus, Q fever, brucellosis and non-food borne diseases. *EFSA J.* **10**, 2765. (doi:10.2903/j.efsa.2012.2765)

27. Bishop S, Doeschl-Wilson AB, Woolliams JA. 2012 Uses and implications of field disease data for livestock genomic and genetics studies. *Front. Genet.* **3**, 114. (doi:10.3389/fgene.2012.00114)

28. Anche MT, De Jong MCM, Bijma P. 2014 On the definition and utilization of heritable variation among hosts in reproduction ratio R_0 for infectious diseases. *Heredity* **113**, 364. (doi:10.1038/hdy.2014.38)

29. Hethcote HW. 1976 Qualitative analyses of communicable disease models. *Math. Biosci.* **28**, 335–356. (doi:10.1016/0025-5564(76)90132-2)

30. Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)

31. Diekmann O, Heesterbeek JAP, Metz JA. 1990 On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382. (doi:10.1007/BF00178324)

32. Anderson RM, Anderson B, May RM. 1992 *Infectious diseases of humans: dynamics and control*. London, UK: Oxford University Press.

33. Diekmann O, Heesterbeek JAP. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, Vol. 5. Chichester, UK: John Wiley & Sons.

34. Griffing B. 1967 Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.* **20**, 127–140. (doi:10.1071/BI9670127)

35. Moore AJ, Brodie III ED, Wolf JB. 1997 Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* **51**, 1352–1362. (doi:10.1111/j.1558-5646.1997.tb01458.x)

36. Wolf JB, Brodie III ED, Cheverud JM, Moore AJ, Wade MJ. 1998 Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**, 64–69. (doi:10.1016/S0169-5347(97)01233-0)

37. Muir WM. 2005 Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**, 1247–1259. (doi:10.1534/genetics.104.035956)

38. Bijma P, Muir WM, Ellen ED, Wolf JB, Van Arendonk JA. 2007 Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**, 289–299. (doi:10.1534/genetics.106.062729)

39. Bijma P. 2014 The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity* **112**, 61. (doi:10.1038/hdy.2013.15)