REVIEW

Natural selection. IV. The Price equation*

S. A. FRANK

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

Abstract

The Price equation partitions total evolutionary change into two components. The first component provides an abstract expression of natural selection. The second component subsumes all other evolutionary processes, including changes during transmission. The natural selection component is often used in applications. Those applications attract widespread interest for their simplicity of expression and ease of interpretation. Those same applications attract widespread criticism by dropping the second component of evolutionary change and by leaving unspecified the detailed assumptions needed for a complete study of dynamics. Controversies over approximation and dynamics have nothing to do with the Price equation itself, which is simply a mathematical equivalence relation for total evolutionary change expressed in an alternative form. Disagreements about approach have to do with the tension between the relative valuation of abstract versus concrete analyses. The Price equation's greatest value has been on the abstract side, particularly the invariance relations that illuminate the understanding of natural selection. Those abstract insights lay the foundation for applications in terms of kin selection, information theory interpretations of natural selection and partitions of causes by path analysis. I discuss recent critiques of the Price equation by Nowak and van Veelen.

The heart and soul of much mathematics consists of the fact that the 'same' object can be presented to us in different ways. Even if we are faced with the simple-seeming task of 'giving' a large number, there is no way of doing this without also, at the same time, 'giving' a hefty amount of extra structure that comes as a result of the way we pin down—or the way we present—our large number. If we write our number as 1729, we are, sotto voce, ordering a preferred way of 'computing it' (add one thousand to seven hundreds to two tens to nine). If we present it as $1 + 12^3$, we are recommending another mode of computation, and if we pin it down—as Ramanujuan did—as the first number expressible as a sum of two cubes in two different ways, we are being less specific about how to compute our number, but have underscored a characterizing property of it within a subtle diophantine arena.…

This issue has been with us, of course, forever: the general question of *abstraction*, as separating what we want from what we are presented with. It is neatly packaged in the Greek verb *aphairein*, as interpreted by Aristotle in the

later books of the *Metaphysics* to mean simply *separation*: if it is *whiteness* we want to think about, we must somehow separate it from *white horse*, *white house*, *white hose* and all the other white things that it invariably *must* come along with, in order for us to experience it at all

(Mazur, 2008, pp. 222–223).

Somewhere … between the specific that has no meaning and the general that has no content there must be, for each purpose and at each level of abstraction, an optimum degree of generality

(Boulding, 1956, pp. 197–198).

Introduction

Evolutionary theory analyses the change in phenotype over time. We may interpret *phenotype* broadly to include organismal characters, variances of characters, correlations between characters, gene frequency, DNA sequence – essentially anything we can measure.

How does a phenotype influence its own change in frequency or the change in the frequencies of correlated phenotypes? Can we separate that phenotypic influence

Correspondence: S. A. Frank, Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697–2525, USA.
Tel.: +1 949 824 2244; fax: +1 949 824 2181; e-mail: safrank@uci.edu
*Part of the Topics in Natural Selection series. See Box 1.

Box 1: Topics in the theory of natural selection

This article is part of a series on natural selection. Although the theory of natural selection is simple, it remains endlessly contentious and difficult to apply. My goal is to make more accessible the concepts that are so important, yet either mostly unknown or widely misunderstood. I write in a nontechnical style, showing the key equations and results rather than providing full derivations or discussions of mathematical problems. Boxes list technical issues and brief summaries of the literature.

Box 2: Price equation literature

A large literature introduces and reviews the Price equation. I list some key references that can be used to get started (Hamilton, 1975; Frank, 1995, 1997; Grafen, 2002; Page & Nowak, 2002; Andersen, 2004; Rice, 2004; Okasha, 2006; Gardner, 2008).

Diverse applications have been developed with the Price equation. I list a few examples (Hamilton, 1970; Wade, 1985; Frank & Slatkin, 1990; Queller, 1992a,b; Michod, 1997a,b; Frank, 1998; Day & Gandon, 2006; Fox, 2006; Grafen, 2007; Alizon, 2009).

Quantitative genetics theory often derives from the covariance expression given by Robertson (1966), which is a form of the covariance term of the Price equation. The basic theory can be found in textbooks (Falconer & Mackay, 1996; Charlesworth & Charlesworth, 2010). Much of the modern work can be traced through the widely cited article by Lande & Arnold (1983).

Harman (2010) provides an interesting overview of Price's life and evokes an Olympian sense of the power and magic of the Price equation. See the study of Schwartz (2000) for an alternative biographical sketch.

---

from other evolutionary forces that also cause change? The association of a phenotype with change in frequency, separated from other forces that change phenotype, is one abstract way to describe natural selection. The Price equation is that kind of abstract separation.

Do we really need such abstraction, which may seem rather distant and vague? Instead of wasting time on such things as the abstract essence of natural selection, why not get down to business and analyse real problems? For example, we may wish to know how the evolutionary forces of mutation and selection interact to determine biological pattern. We could make a model with genes that have phenotypic effects, selection that acts on those phenotypes to change gene frequency, and mutation that changes one gene into another. We could do some calculations, make some predictions about, for example, the frequency of deleterious mutations that cause disease, and compare those predictions to observations. All clear and concrete, without need of any discussion of the essence of things.

However, we may ask the following. Is there some reorientation for the expression of natural selection that may provide subtle perspective, from which we can understand our subject more deeply and analyse our problems with greater ease and greater insight? My answer is, as I have mentioned, that the Price equation provides that sort of reorientation. To argue the point, I will have to keep at the distinction between the concrete and the abstract, and the relative roles of those two endpoints in mature theoretical understanding.

Several decades have passed since Price's (1970, 1972a) original articles. During that span, published claims, counter-claims and misunderstandings have accumulated to the point that it seems worthwhile to revisit the subject. On the one hand, the Price equation has been applied to numerous practical problems and has also been elevated by some to almost mythical status, as if it were the ultimate path to enlightenment for those devoted to evolutionary study (Box 2).

On the other hand, the opposition has been gaining adherents who boast the sort of disparaging anecdotes and slogans that accompany battle. In a recent book, Nowak & Highfield (2011) counter

The Price equation did not, however, prove as useful as [Price and Hamilton] had hoped. It turned out to be the mathematical equivalent of a tautology. … If the Price equation is used instead of an actual model, then the arguments hang in the air like a tantalizing mirage. The meaning will always lie just out of the reach of the inquisitive biologist. This mirage can be seductive and misleading. The Price equation can fool people into believing that they have built a mathematical model of whatever system they are studying. But this is often not the case. Although answers do indeed seem to pop out of the equation, like rabbits from a magician's hat, nothing is achieved in reality.

Nowak & Highfield (2011) approvingly quote van Veelen *et al.* (2012) with regard to calling the Price equation a *mathematical tautology*. van Veelen *et al.* (2012) emphasize the point by saying that the Price equation is like soccer/football star Johan Cruyff's quip about the secret of success: 'You always have to make sure that you score one goal more than your opponent'. The statement is always true, but provides no insight. Nowak & Highfield (2011) and van Veelen *et al.* (2012) believe their arguments demonstrate that the Price equation is true in the same trivial sense, and they call that trivial type of truth a *mathematical tautology*. Interestingly, magazines, online articles and the scientific literature have for several years been using the phrase *mathematical tautology* for the Price equation, although Nowak & Highfield (2011) and van Veelen *et al.* (2012) do not provide citations to previous literature.

As far as I know, the first description of the Price equation as a mathematical tautology was in the study of Frank (1995). I used the phrase in the sense of the epigraph from Mazur, a formal equivalence between

different expressions of the same object. Mathematics and much of statistics are about formal equivalences between different expressions of the same object. For example, the Laplace transform changes a mathematical expression into an alternative form with the same information, and analysis of variance decomposes the total variance into a sum of component variances. For any mathematical or statistical equivalence, value depends on enhanced analytical power that eases further derivations and calculations and on the ways in which previously hidden relations are revealed.

In the light of the contradictory points of view, the main goal of this article is to sort out exactly what the Price equation is, how we should think about it, and its value and limitations in reasoning about evolution. Subsequent articles will show the Price equation in action, applied to kin selection, causal analysis in evolutionary models and an information perspective of natural selection and Fisher's fundamental theorem.

Overview

The first section derives the Price equation in its full and most abstract form. That derivation allows us to evaluate the logical status of the equation in relation to various claims of fundamental flaw. The equation survives scrutiny. It is a mathematical relation that expresses the total amount of evolutionary change in an alternative and mathematically equivalent way. That equivalence provides insight into aspects of natural selection and also provides a guide that, in particular applications, often leads to good approaches for analysis.

The second section contrasts two perspectives of evolutionary analysis. In standard models of evolutionary change, one begins with the initial population state and the rules of change. The rules of change include the fitness of each phenotype and the change in phenotype between ancestor and descendant. Given the initial state and rules of change, one deduces the state of the changed population. Alternatively, one may have data on the initial population state, the changed population state and the ancestor–descendant relations that map entities from one population to the other. Those data may be reduced to the evolutionary distance between two populations, providing inductive information about the underlying rules of change. Natural populations have no intrinsic notion of fitness or rules of change. Instead, they inductively accumulate information. The Price equation includes both the standard deductive model of evolutionary change and the inductive model by which information accumulates in relation to the evolutionary distance between populations.

The third and fourth sections discuss the Price equation's abstract properties of invariance and recursion. The invariance properties include the information theory interpretation of natural selection. Recursion provides the basis for analysing group selection and other models of multilevel selection.

The fifth section relates the Price equation to various expressions that have been used throughout the history of evolutionary theory to analyse natural selection. The most common form describes natural selection by the covariance between phenotype and fitness or by the covariance between genetic breeding value and fitness. The covariance expression is one part of the Price equation that, when used alone, describes the natural selection component of total evolutionary change. The essence of those covariance forms arose in the early studies of population and quantitative genetics, had been used extensively during much of the modern history of animal breeding, and began to receive more mathematical development in the 1960s and 1970s. Recent critiques of the Price equation focus on the same covariance expression that has been widely used throughout the history of population and quantitative genetics to analyse natural selection and to approximate total evolutionary change.

The sixth section returns to the full abstract form of the equation. I compare a few variant expressions that have been promoted as improvements on the original Price equation. Variant forms are indeed helpful with regard to particular abstract problems or particular applications. However, most variants are simply minor rearrangements of the mathematical equivalence for total evolutionary change given by the original Price equation. The recent extension by Kerr & Godfrey-Smith (2009) does provide a slightly more general formulation by expanding the fundamental set mapping that defines Price's approach. The set mapping basis for the Price equation deserves more careful study and further mathematical work.

The seventh section analyses various flaws that have been ascribed to the Price equation. For example, the Price equation in its most abstract form does not contain enough information to follow evolutionary dynamics through multiple rounds of natural selection. By contrast, classical dynamic models of population genetics are sufficient to follow change through time. Much has been made of this distinction with regard to dynamic sufficiency. The distinction arises from the fact that classical dynamics in population genetics makes more initial assumptions than the abstract Price equation. It must be true that all mathematical equivalences for total evolutionary change have the same dynamic status given the same initial assumptions. Each additional well-chosen assumption typically enhances the specificity and reduces the scope and generality of the analysis. The epigraph from Boulding emphasizes that the degree of specificity versus generality is an explicit choice of the analyst with respect to initial assumptions.

The Discussion considers the value and limitations of the Price equation in relation to recent criticisms by Nowak and van Veelen. The critics confuse the distinct roles of general abstract theory and concrete dynamical models for particular cases. The enduring power of the Price equation arises from the discovery of essential invariances in natural selection. For example, kin selection theory expresses biological problems in terms of relatedness coefficients. Relatedness measures the association between social partners. The proper measure of relatedness identifies distinct biological scenarios with the same (invariant) evolutionary outcome. Invariance relations provide the deepest insights of scientific thought.

The Price equation

> The mathematics given here applies not only to genetical selection but to selection in general. It is intended mainly for use in deriving general relations and constructing theories, and to clarify understanding of selection phenomena, rather than for numerical calculation (Price, 1972a, p. 485).

I have emphasized that the Price equation is a mathematical equivalence. The equation focuses on separation of total evolutionary change into a part attributed to selection and a remainder term. That separation provides an abstraction of the nature of selection. As Price wrote sometime around 1970 but published posthumously in Price (1995), 'Despite the pervading importance of selection in science and life, there has been no abstraction and generalization from genetical selection to obtain a general selection theory and general selection mathematics'.

It is useful first to consider the Price equation in this most abstract form. I follow my earlier derivations (Frank, 1995, 1997, 1998, 2009), which differ little from the derivation given by Price (1972a) when interpreted in the light of the study of Price (1995).

The abstract expression can best be thought of in terms of mapping items between two sets (Frank, 1995; Price, 1995). In biology, we usually think of an ancestral population at some time and a descendant population at a later time. Although there is no need to have an ancestor–descendant relation, I will for convenience refer to the two sets as ancestor and descendant. What does matter is the relations between the two sets, as follows.

Definitions

The full abstract power of the Price equation requires adhering strictly to particular definitions. The definitions arise from the general expression of the relations between two sets.

Let q_i be the frequency of the ith type in the ancestral population. The index i may be used as a label for any sort of property of things in the set, such as allele, genotype, phenotype, group of individuals and so on. Let q_i' be the frequencies in the descendant population, defined as the fraction of the descendant population that is derived from members of the ancestral population that have the label i. Thus, if $i = 2$ specifies a particular phenotype, then q_2' is not the frequency of the phenotype $i = 2$ among the descendants. Rather, it is the fraction of the descendants derived from entities with the phenotype $i = 2$ in the ancestors. One can have partial assignments, such that a descendant entity derives from more than one ancestor, in which case each ancestor gets a fractional assignment of the descendant. The key is that the i indexing is always with respect to the properties of the ancestors, and descendant frequencies have to do with the fraction of descendants derived from particular ancestors.

Given this particular mapping between sets, we can specify a particular definition for fitness. Let $q_i' = q_i(w_i/\bar{w})$, where w_i is the fitness of the ith type and $\bar{w} = \sum q_i w_i$ is average fitness. Here, $w_i/\bar{w}$ is proportional to the fraction of the descendant population that derives from type i entities in the ancestors.

Usually, we are interested in how some measurement changes or evolves between sets or over time. Let the measurement for each i be z_i. The value z may be the frequency of a gene, the squared deviation of some phenotypic value in relation to the mean, the value obtained by multiplying measurements of two different phenotypes of the same entity and so on. In other words, z_i can be a measurement of any property of an entity with label, i. The average property value is $\bar{z} = \sum q_i z_i$, where this is a population average.

The value z_i' has a peculiar definition that parallels the definition for q_i'. In particular, z_i' is the average measurement of the property associated with z among the descendants derived from ancestors with index i. The population average among descendants is $\bar{z}' = \sum q_i' z_i'$.

The Price equation expresses the total change in the average property value, $\Delta\bar{z} = \bar{z}' - \bar{z}$, in terms of these special definitions of set relations. This way of expressing total evolutionary change and the part of total change that can be separated out as selection is very different from the usual ways of thinking about populations and evolutionary change. The derivation itself is very easy, but grasping the meaning and becoming adept at using the equation is not so easy.

I will present the derivation in two stages. The first stage makes the separation into a part ascribed to selection and a part ascribed to property change that covers everything beyond selection. The second stage retains this separation, changing the notation into standard statistical expressions that provide the form of the Price equation commonly found in the literature. I follow with some examples to illustrate how particular set relations are separated into selection and property change components. The next section considers two distinct interpretations of the Price equation in relation to dynamics.

Derivation: separation into selection and property value change

We use $\Delta q_i = q_i' - q_i$ for frequency change associated with selection and $\Delta z_i = z_i' - z_i$ for property value change. Both expressions for change depend on the special set relation definitions given above.

We are after an alternative expression for total change, $\Delta \bar{z}$. Thus,

$$
\begin{aligned}
\Delta \bar{z} &= \bar{z}' - \bar{z} \\
&= \sum q_i' z_i' - \sum q_i z_i \\
&= \sum q_i'(z_i' - z_i) + \sum q_i' z_i - \sum q_i z_i \\
&= \sum q_i'(\Delta z_i) + \sum (\Delta q_i) z_i.
\end{aligned}
$$

Switching the order of the terms on the right side of the last line yields

$$
\Delta \bar{z} = \sum (\Delta q_i) z_i + \sum q_i'(\Delta z_i), \tag{1}
$$

a form emphasized by Frank (1997, eqn 1). The first term separates the part of total change caused by changes in frequency. We call this the part caused by selection, because this is the part that arises directly from differential contribution by ancestors to the descendant population (Price, 1995). Because the set mappings define all of the direct attributions of success for each i with respect to the associated properties z_i, it is reasonable to separate out this direct component as the abstraction of selection. It is of course possible to define other separations. I discuss one particular alternative later. However, it is hard to think of other separations that would describe selection in a better way at the most abstract and general level of the mappings between two sets. This first term has also been called the partial evolutionary change caused by natural selection (eqn 7).

The second term describes the part of total change caused by changes in property values. Recall that $\Delta z_i = z_i' - z_i$ and that z_i' is the property value among entities that descend from i. Many different processes may cause descendant property values to differ from ancestral values. In fact, the assignment of a descendant to an ancestor can be entirely arbitrary, so that there is no reason to assume that descendants should be like ancestors. Usually, we will work with systems in which descendants do resemble ancestors, but the degree of such associations can be arranged arbitrarily. This term for change in property value encompasses everything beyond selection. The idea is that selection affects the relative contribution of ancestors and thus the changes in frequencies of representation, but what actually gets represented among the descendants will be subject to a variety of processes that may alter the value expressed by descendants.

The equation is exact and must apply to every evolutionary system that can be expressed as two sets with certain ancestor–descendant or mapping relations. It is in that sense that I first used the phrase *mathematical tautology* (Frank, 1995). The nature of separation and abstraction is well described by the epigraph from Mazur at the start of this article.

Derivation: statistical notation

Price (1972a) used statistical notation to write eqn 1. For the first term, by following prior definitions, we have

$$
\begin{aligned}
\Delta q_i &= q_i' - q_i \\
&= q_i \frac{w_i}{\bar{w}} - q_i \\
&= q_i \left(\frac{w_i}{\bar{w}} - 1 \right),
\end{aligned}
$$

so that

$$
\sum (\Delta q_i) z_i = \sum q_i \left(\frac{w_i}{\bar{w}} - 1 \right) z_i = \mathrm{Cov}(w, z)/\bar{w},
$$

using the standard definition for population covariance. For the second term, we have

$$
\sum q_i'(\Delta z_i) = \sum q_i \frac{w_i}{\bar{w}}(\Delta z_i) = \mathrm{E}(w \Delta z)/\bar{w},
$$

where E means expectation, or average over the full population. Putting these statistical forms into eqn 1 and moving $\bar{w}$ to the left side for notational convenience yields a commonly published form of the Price equation

$$
\bar{w} \Delta \bar{z} = \mathrm{Cov}(w, z) + \mathrm{E}(w \Delta z). \tag{2}
$$

Frank (1995) and Price (1995) present examples of set mappings expressed in relation to the Price equation.

Dynamics: inductive and deductive perspectives

The Price equation describes evolutionary change between two populations. Three factors express one iteration of dynamical change: initial state, rules of change and next state. In the Price equation, the phenotypes, z_i, and their frequencies, q_i, describe the initial population state. Fitnesses, w_i, and property changes, Δz_i, set the rules of change. Derived phenotypes, z_i', and their frequencies, q_i', express the next population state.

Models of evolutionary change essentially always analyse forward or deductive dynamics. In that case, one starts with initial conditions and rules of change and calculates the next state. Most applications of the Price equation use this traditional deductive analysis. Such applications lead to predictions of evolutionary outcome given assumptions about evolutionary process, expressed by the fitness parameters and property changes.

Alternatively, one can take the state of the initial population and the state of the changed population as given. If one also has the mappings between initial and changed populations that connect each entity, i, in the initial population to entities in the changed population, then one can calculate (induce) the underlying rules of

change. At first glance, this inductive view of dynamics may seem rather odd and not particularly useful. Why start with knowledge of the evolutionary sequence of population states and ancestor–descendant relations as given, and inductively calculate fitnesses and property changes? The inductive view takes the fitnesses, w_i, to be derived from the data rather than an intrinsic property of each type.

The Price equation itself does not distinguish between the deductive and inductive interpretations. One can specify initial state and rules of change and then deduce outcome. Or one can specify initial state and outcome along with ancestor–descendant mappings and then induce the underlying rules of change. It is useful to understand the Price equation in its full mathematical generality and to understand that any specific interpretation arises from additional assumptions that one brings to a particular problem. Much of the abstract power of the Price equation comes from understanding that, by itself, the equation is a minimal description of change between populations.

The deductive interpretation of the Price equation is clear. What value derives from the inductive perspective? In observational studies of evolutionary change, we only have data on population states. From those data, we use the inductive perspective to make inferences about the underlying rules of change. Note that inductive estimates for evolutionary process derive from the amount of change, or distance, between ancestor and descendant populations. The Price equation includes that inductive, or retrospective, view by expressing the distance between populations in terms of $\Delta \bar{z}$. I develop that distance interpretation in the following sections.

Perhaps more importantly, natural selection itself is inherently an inductive process by which information accumulates in populations. Nature does not intrinsically 'know' of fitness parameters. Instead, frequency changes and the mappings between ancestor and descendant are inherent in a population's response to the environment, leading to a sequence of population states, each separated by an evolutionary distance. That evolutionary distance provides information that populations accumulate inductively about the fitnesses of each phenotype (Frank, 2009). The Price equation includes both the deductive and inductive perspectives. We may choose to interpret the equation in either way depending on our goals of analysis.

Abstract properties: invariance

The Price equation describes selection by the term $\sum(\Delta q_i)z_i = \mathrm{Cov}(w,z)/\bar{w}$. Any instance of evolutionary change that has the same value for this sum has the same amount of total selection. Put another way, for any particular value for total selection, there is an infinite number of different combinations of frequency changes and character measurements that will add up to the same

total value for selection. All of those different combinations lead to the same value with respect to the amount of selection. We may say that all of those different combinations are *invariant* with respect to the total quantity of selection. The deepest insights of science come from understanding what does not matter, so that one can also say exactly what does matter – what is invariant (Feynman, 1967; Weyl, 1983).

The invariance of selection with respect to transformations of the fitnesses, w, and the phenotypes, z, that have the same $\mathrm{Cov}(w,z)$ means that, to evaluate selection, it is sufficient to analyse this covariance. At first glance, it may seem contradictory that the covariance, commonly thought of as a linear measure of association, can be a complete description for selection, including nonlinear processes. Let us step through this issue, first looking at why the covariance is a sufficient expression of selection and then at the limitations of this covariance expression in evolutionary analysis.

Covariance as a measure of distance: definitions

Much of the confusion with respect to covariance and variance terms in selection equations arises from thinking only of the traditional statistical usage. In statistics, covariance typically measures the linear association between pairs of observations, and variance is a measure of the squared spread of observations. Alternatively, covariances and variances provide measures of distance, which ultimately can be understood as measures of information (Frank, 2009). This section introduces the notation for the geometric interpretation of distance. The next section gives the main geometric result, and the following section presents some examples.

The identity $\sum(\Delta q_i)z_i = \mathrm{Cov}(w,z)/\bar{w}$ provides the key insight. It helps to write this identity in an alternative form. Note from the prior definition $q_i' = q_i w_i/\bar{w}$ that

$$\Delta q_i = q_i' - q_i = q_i(w_i/\bar{w} - 1) = q_i a_i, \qquad (3)$$

where $a_i = w_i/\bar{w} - 1$ is Fisher's average excess in fitness, a commonly used expression in population and quantitative genetics (Fisher, 1930, 1941; Crow & Kimura, 1970). A value of zero means that an entity has average fitness, and therefore, fitness effects and selection do not change the frequency of that entity. Using the average excess in fitness, we can write the invariant expression for selection as

$$\sum(\Delta q_i)z_i = \sum q_i a_i z_i = \mathrm{Cov}(w,z)/\bar{w}. \qquad (4)$$

We can think of the state of the population as the listing of character states, z_i. Thus, we write the population state as $\mathbf{z} = (z_1, z_2, \ldots)$. The subscripts run over every different entity in the population, so the vector $\mathbf{z}$ is a complete description of the entire population. Similarly, for the frequency fluctuations, $\Delta q_i = q_i a_i$, we can write the listing of all fluctuations as a vector, $\Delta \mathbf{q} = (\Delta q_1, \Delta q_2, \ldots)$.

It is often convenient to use the dot product notation

$$\Delta \mathbf{q} \cdot \mathbf{z} = \sum (\Delta q_i) z_i = \mathrm{Cov}(w, z)/\bar{w}$$

in which the dot specifies the sum obtained by multiplying each pair of items from two vectors. Before turning to some geometric examples in the following section, we need a definition for the length of a vector. Traditionally, one uses the definition

$$\|\mathbf{z}\| = \sqrt{\sum z_i^2},$$

in which the length is the square root of the sum of squares, which is the standard measure of length in Euclidean geometry.

Covariance as a measure of distance: examples

A simple identity relates a dot product to a measure of distance and to covariance selection

$$\Delta \mathbf{q} \cdot \mathbf{z} = \|\Delta \mathbf{q}\| \, \|\mathbf{z}\| \cos \phi = \mathrm{Cov}(w, z)/\bar{w}, \quad (5)$$

where ϕ is the angle between the vectors $\Delta \mathbf{q}$ and $\mathbf{z}$ (Fig. 1). If we standardize the character vector $\mathbf{\underline{z}} = \mathbf{z}/\|\mathbf{z}\|$, then the standardized vector has a length of one, $\|\mathbf{\underline{z}}\| = 1$, which simplifies the dot product expression of selection to

$$\Delta \mathbf{q} \cdot \mathbf{\underline{z}} = \|\Delta \mathbf{q}\| \cos \phi,$$

providing the geometric representation illustrated in Fig. 1.

The covariance can be expressed as the product of a regression coefficient and a variance term

$$\mathrm{Cov}(w, z)/\bar{w} = \beta_{zw} \mathrm{Var}(w)/\bar{w} = \beta_{wz} \mathrm{Var}(z)/\bar{w}, \quad (6)$$

where the notation β_{xy} describes the regression coefficient of x on y (Price, 1970). This identity shows that the expression of selection in terms of a regression coefficient and a variance term is equivalent to the geometric expression of selection in terms of distance.

I emphasize these identities for two reasons. First, as Mazur stated in the epigraph, 'The heart and soul of much mathematics consists of the fact that the ''same'' object can be presented to us in different ways'. If an object is important, such as natural selection surely is, then it pays to study that object from different perspectives to gain deeper insight.

Second, the appearance of statistical functions, such as the covariance and variance, in selection equations sometimes leads to mistaken conclusions. In the selection equations, it is better to think of the covariance and variance terms arising because they are identities with geometric or other interpretations of selection, rather than thinking of those terms as summary statistics of probability distributions. The problem with thinking of those terms as statistics of probability distributions is that the variance and covariance are
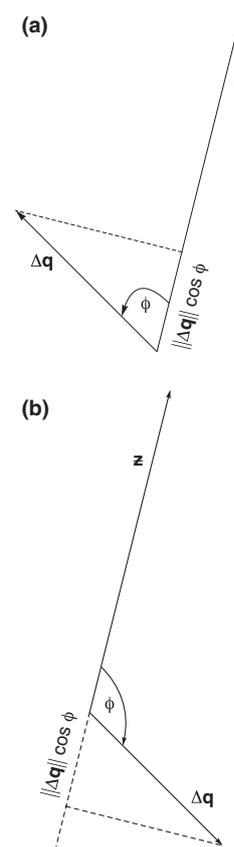

Fig. 1 Geometric expression of selection. The plots show the equivalence of the dot product, the geometric expression and the covariance, as given in eqn 5. For both plots, $\mathbf{z} = (1,4)$ and $\mathbf{\underline{z}} = \mathbf{z}/\|\mathbf{z}\| = (0.24, 0.97)$. The dashed line shows the perpendicular between the pattern of frequency changes derived from fitnesses, $\Delta \mathbf{q}$, and the phenotypic pattern, $\mathbf{\underline{z}}$. The vertex of the two vectors is at the origin (0,0). The distance from the origin to the intersection of the perpendicular along $\mathbf{\underline{z}}$ is the total amount of selection, $\|\Delta \mathbf{q}\| \cos \phi$. (a) The vector of frequency changes that summarize fitness is $\Delta \mathbf{q} = (-0.4, 0.4)$. The angle between the vector of frequency changes and the phenotypes is $\phi = \arccos[(\Delta \mathbf{q} \cdot \mathbf{\underline{z}})/\|\Delta \mathbf{q}\|]$ which, in this example, is 1.03 radians or 59°. In this case, the total selection is $\|\Delta \mathbf{q}\| \cos \phi = 0.29$. (b) In this plot, $\Delta \mathbf{q} = (0.4, -0.4)$, yielding an angle ϕ of 121°. The perpendicular intersects the negative projection of the phenotype vector, shown as a dashed line, associated with the negative change by selection of $\|\Delta \mathbf{q}\| \cos \phi = -0.29$.

not in general sufficient descriptions for probability distributions. That lack of sufficiency for probability may lead one to conclude that those terms are not sufficient for a general expression of selection. However, those covariance and variance terms are sufficient. That sufficiency can be understood by thinking of those terms as identities for distance or measures of information (Frank, 2009).

It is true that in certain particular applications of quantitative genetics or stochastic sampling processes,

one does interpret the variances and covariances as summary statistics of probability distributions, usually the normal or Gaussian distribution. However, it is important to distinguish those special applications from the general selection equations.

Invariance and information

For the general selection expression in eqn 5, any transformations that do not affect the net values are invariant with respect to selection. For example, transformations of the fitnesses and associated frequency changes, $\Delta \mathbf{q}$, are invariant if they leave unchanged the distance expressed by $\Delta \mathbf{q} \cdot \mathbf{z} = \mathrm{Cov}(w, z)/\bar{w}$. Similarly, changes in the pattern of phenotypes are invariant to the extent that they leave $\Delta \mathbf{q} \cdot \mathbf{z}$ unchanged. These invariance properties of selection, measured as distance, may not appear very interesting at first glance. They seem to be saying that the outcome is the outcome. However, the history of science suggests that studying the invariant properties of key expressions can lead to insight.

Few authors have developed an interest in the invariant qualities of selection. Fisher (1930) initiated discussion with his fundamental theorem of natural selection, a special case of eqn 5 (Frank, 1997). Although many authors commented on the fundamental theorem, most articles did not analyse the theorem with respect to its essential mathematical insights about selection. Ewens (1992) reviewed the few attempts to understand the mathematical basis of the theorem and its invariant quantities. Frank (2009) tied the theorem to Fisher information (Frieden *et al.*, 2001; Frieden, 2004), hinting at an information theory interpretation that arises from the fundamental selection equation of eqn 5.

In spite of the importance of selection in many fields of science, the potential interpretation of eqn 5 with respect to invariants of information theory has hardly been developed. I briefly outline the potential connections here (Frank, 2009). I develop this information perspective of selection in a later article, along with Fisher's fundamental theorem.

To start, define the partial change in phenotype caused by natural selection as

$$\Delta_s \bar{z} = \Delta \mathbf{q} \cdot \mathbf{z} = \mathrm{Cov}(w, z)/\bar{w}. \tag{7}$$

The concept of a partial change caused by natural selection arises from Fisher's fundamental theorem (Fisher, 1930; Price, 1972b; Ewens, 1989; Frank & Slatkin, 1992). With this definition, we can use eqns 5 and 6 to write

$$\Delta_s \bar{z} = \beta_{zw} \mathrm{Var}(w)/\bar{w} = \bar{w} \beta_{zw} \mathrm{Var}(w/\bar{w}). \tag{8}$$

From eqn 3, we have the definition for the average excess in fitness $a_i = w_i/\bar{w} - 1$. Thus, we can expand the expression for the variance in fitness as

$$\mathrm{Var}(w/\bar{w}) = \sum q_i \left(\frac{w_i}{\bar{w}} - 1 \right)^2 = \sum q_i a_i^2.$$

From eqn 3, we also have the change in frequency in terms of the average excess, $\Delta q_i = q_i a_i$, and equivalently, $a_i = \Delta q_i / q_i$, thus

$$\mathrm{Var}(w/\bar{w}) = \sum q_i \left(\frac{\Delta q_i}{q_i} \right)^2$$
$$= \sum \left(\frac{\Delta q_i}{\sqrt{q_i}} \right)^2$$
$$= \Delta \hat{\mathbf{q}} \cdot \Delta \hat{\mathbf{q}},$$

where $\Delta \hat{q}_i = \Delta q_i / \sqrt{q_i}$ is a standardized fluctuation in frequency and $\Delta \hat{\mathbf{q}}$ is the vector of standardized fluctuations. These alternative forms simply express the variance in fitness in different ways. The interesting result follows from the fact that

$$\mathrm{Var}(w/\bar{w}) = \Delta \hat{\mathbf{q}} \cdot \Delta \hat{\mathbf{q}} = F(\Delta \hat{\mathbf{q}})$$

is the Fisher information, F, in the frequency fluctuations, $\Delta \hat{\mathbf{q}}$. Fisher information is a fundamental quantity in information theory, Bayesian analysis, likelihood theory and the informational foundations of statistical inference. Fisher information is a variant form of the more familiar Shannon and Kullback–Leibler information measures, in which the Fisherian form expresses changes in information.

Once again, we have a simple identity. Although it is true that Fisher information is just an algebraic rearrangement of the variance in fitness, some insight may be gained by relating selection to information. The variance form calls to mind a statistical description of selection or a partial description of a probability distribution. The Fisher information form suggests a relation between natural selection and the way in which populations accumulate information (Frank, 2009).

We may now write our fundamental expression for selection as

$$\Delta_s \bar{z} = \bar{w} \beta_{zw} F(\Delta \hat{\mathbf{q}}).$$

We may read this expression for selection as follows: the change in mean character value caused by natural selection, $\Delta_s \bar{z}$, is equal to the total Fisher information in the frequency fluctuations, F, multiplied the scaling β that describes the amount of the potential information that the population captures when expressed in units of phenotypic change. In other words, the distance $\Delta_s \bar{z}$ measures the informational gain by the population caused by natural selection.

The invariances set by this expression may be viewed in different ways. For example, the distance of evolutionary change by selection, $\Delta_s \bar{z}$, is invariant with respect to many different combinations of frequency fluctuations, $\Delta \hat{\mathbf{q}}$, and scalings between phenotype and fitness. Similarly, any transformations of frequency fluctuations that leave the measure of information, $F(\Delta \hat{\mathbf{q}})$, invariant

do not alter the scaled change in phenotype caused by natural selection. The full implications remain to be explored.

Summary of selection identities

The various identities for the part of total evolutionary change caused by selection include

$$
\begin{aligned}
\Delta_s \bar{z} &= \mathrm{Cov}(w, z)/\bar{w} \\
&= \bar{w}\beta_{zw}\mathrm{Var}(w/\bar{w}) \\
&= \Delta\mathbf{q} \cdot \mathbf{z} \\
&= \|\Delta\mathbf{q}\|\|\mathbf{z}\| \cos\phi \\
&= \bar{w}\beta_{zw}(\Delta\hat{\mathbf{q}} \cdot \Delta\hat{\mathbf{q}}) \\
&= \bar{w}\beta_{zw}F(\Delta\hat{\mathbf{q}}).
\end{aligned}
\tag{9}
$$

These forms show the equivalence of the statistical, geometrical and informational expressions for natural selection. These general abstract forms make no assumptions about the nature of phenotypes and the patterns of frequency fluctuations caused by differential fitness. The phenotypes may be squared deviations so that the average is actually a variance, or the product of measurements on different characters leading to measures of association, or any other nonlinear combination of measurements. Thus, there is nothing inherently linear or restrictive about these expressions.

Selection versus evolution

The previous sections discussed the part of evolutionary change caused by selection. The full Price equation (eqn 2) gives a complete and exact expression of total change, repeated here as

$$
\Delta\bar{z} = \mathrm{Cov}(w, z)/\bar{w} + \mathrm{E}(w\Delta z)/\bar{w}
\tag{10}
$$

or in terms of the dot product notation as

$$
\Delta\bar{z} = \Delta\mathbf{q} \cdot \mathbf{z} + \mathbf{q}' \cdot \Delta\mathbf{z}.
\tag{11}
$$

The full change in the phenotype is the sum of the two terms, which we may express in symbols as

$$
\Delta\bar{z} = \Delta_s\bar{z} + \Delta_E\bar{z}.
$$

Fisher (1930) called the term $\Delta_E\bar{z}$ the change caused by the environment (Frank & Slatkin, 1992). However, the word *environment* often leads to confusion. The proper interpretation is that $\Delta_E\bar{z}$ encompasses everything not included in the expression for selection. The term is *environmental* only in the sense that it includes all those forces external to the particular definition of the selective forces for a particular problem.

The Δ_E term is sometimes associated with changes in transmission (Frank, 1995, 1997, 2012a; Okasha, 2006). This interpretation arises because $\mathrm{E}(w\Delta z)$ is the fitness-weighted changes in character value between ancestor and descendant. One may think of changes in character values as changes during transmission.

It is important to realize that *everything* truly means every possible force that might arise and that is not accounted for by the particular expression for selection. Lightning may strike. New food sources may appear. The Price equation in its general and abstract form is a mathematical identity–what I previously called a *mathematical tautology* (Frank, 1995).

In applications, one considers how to express $\Delta_E\bar{z}$, or one searches for ways to formulate the problem so that $\Delta_E\bar{z}$ is zero or approximately zero. This article is not about particular applications. Here, I simply note that when one works with Fisher's breeding value as z, near equilibria (fixed points), one typically obtains $\Delta z \to 0$ and thus $\mathrm{E}(w\Delta z) \to 0$. In other cases, the search for a good way to express a problem means finding a form of character measurement that defines z such that characters tend to remain stable over time, so that $\Delta z \to 0$ and thus $\mathrm{E}(w\Delta z) \to 0$. For applications that emphasize calculation of complex dynamics rather than a more abstract conceptual analysis of a problem, methods other than the Price equation often work better.

Abstract properties: recursion and group selection

To iterate is human, to recurse, divine (Coplien, 1998).

Essentially, all modern discussions of multilevel selection and group selection derive from Price (1972a), as developed by Hamilton (1975). Price and Hamilton noted that the Price equation can be expanded recursively to represent nested levels of analysis, for example individuals living in groups.

Start with the basic Price equation as given in eqn 10. The left side is the total change in average phenotype, $\bar{z}$. The second term on the right side includes the terms Δz_i in $\mathrm{E}(w\Delta z) = \sum q_i w_i \Delta z_i$.

Recall that in defining z_i, we specified the meaning of the index i to be any sort of labelling of set members, subject to minimal consistency requirements. We may, for example, label all members of a group by i and measure z_i as some property of the group. If the index i itself represents a set, then we may consider the members of that set. For example, z_{ij} may be the jth member of the ith set, or we may say, the ith group. In the abstract mathematical expression, there is no need to think of the ith group as having any spatial or biological meaning. However, we may consider i as a label for spatially defined groups if we wish to do so.

With i defining a group, we may analyse the selection and evolution of that ith group. The term Δz_i becomes the average change in the z measure for the ith group, composed of members with values z_{ij}. The terms z'_{ij} are the average property values of the descendants of the jth entity in the ith group. The descendant entities that derive from the ith group do not have to form any sort of group or other meaningful structuring, just as the original i labelling does not have to refer to group

structuring in the ancestors. However, we may if we wish consider descendants of i as retaining some sense of the ancestral grouping.

Because z_i represents an averaging over the entities j in the ith group, we are assuming the notational equivalence $\Delta z_i = \Delta \bar{z}_i$. From that point of view, for each group i we may from eqn 10 express the change in the group mean by thinking of each group as a separate set or population, yielding for each i the expression

$$\Delta z_i = \Delta \bar{z}_i = \text{Cov}(w_i, z_i)/\bar{w}_i + \text{E}(w_i \Delta z_i)/\bar{w}_i.$$

We may substitute this expression for each i into the $\text{E}(w\Delta z) = \sum q_i w_i \Delta z_i$ term on the right side of eqn 10. That substitution recursively expands each change in property value, Δz_i, to itself be composed of a selection term and property value change term. For each group, i, we now have expressions for selection within the group, $\text{Cov}(w_i, z_i)/\bar{w}_i$, and average property value change within the group, $\text{E}(w_i \Delta z_i)/\bar{w}_i$. If we write out the full expression for this last term, we obtain

$$\text{E}(w_i \Delta z_i)/\bar{w}_i = \sum_j w_{ij} \Delta z_{ij}/\bar{w}_i.$$

In the term Δz_{ij}, each labelling, j, may itself be a subgroup within the larger grouping represented by i. The recursive nature of the Price equation allows another expansion to the characters z_{ijk} for the kth entity in the jth grouping that is nested in the ith group and so on. Once again, the indexing for levels i, j and k do not have to correspond to any particular structuring, but we may choose to use a structuring if we wish.

One could analyse biological problems of group selection without using the Price equation. Because the Price equation is a mathematical identity, there are always other ways of expressing the same thing. However, in the 1970s, when group selection was a very confused subject, the Price equation's recursive nature and Hamilton's development provided the foundation for subsequent understanding of the topic. All modern conceptual insights about group selection derive from Price's recursive expansion of his abstract expression of selection.

History and alternative expressions of selection

I have emphasized the general and abstract form of the Price equation. That abstract form was first presented rather cryptically by Price (1972a). In that article, Price described the recursive expansion to analyse group selection. Apart from the recursive aspect, the more general abstract properties were hardly mentioned in the study of Price (1972a) and not developed by others until 1995.

While I was writing my history of Price's contributions to evolutionary genetics (Frank, 1995), I found Price's unpublished manuscript *The nature of selection* among W. D. Hamilton's papers. Price's unpublished manuscript gave a very general and abstract scheme for analysing

selection in terms of set relations. However, Price did not explicitly connect the abstract set relation scheme to the Price equation or to his earlier publications (Price, 1970, 1972a).

I had *The nature of selection* published posthumously as Price (1995). In my own article, I explicitly developed the general interpretation of the Price equation as the formal abstract expression of the relation between two sets (Frank, 1995).

Price (1970) wrote an earlier article in which he presented a covariance selection equation that emphasized the connection to classical models of population genetics and gene frequency change. That earlier covariance form lacks the abstract set interpretation and generally has narrower scope. Preceding Price, Robertson (1966) and Li (1967) also presented selection equations that are similar to Price's (1970) covariance expression. Robertson's covariance form itself arises from classical quantitative genetics and the breeder's equation, ultimately deriving from the foundations of quantitative genetics established by Fisher (1918). Li's form presents a covariance type of expression for classical population genetic models of gene frequency change.

One cannot understand the current literature without a clear sense of this history. Almost all applications of the Price equation to kin and group selection, and to other problems of evolutionary analysis, derive from either the classical expressions of quantitative genetics (Robertson, 1966) or classical expressions of population genetics (Li, 1967).

In the light of this history, criticisms can be confusing with regard to the ways in which the Price equation is commonly used. For example, in applications to kin or group selection, the Price equation mainly serves to package the notation for the Robertson form of quantitative genetic analysis or the Li form of population genetic analysis. The Price equation packaging brings no extra assumptions. In some applications, critics may believe that the particular analysis lacks enough assumptions to attain a desired level of specificity. One can, of course, easily add more assumptions, at the expense of reduced generality.

The following sections briefly describe some alternative forms of the Price equation and the associated history. That history helps to place criticisms of the Price equation and its applications into clearer light.

Quantitative genetics and the breeder's equation

Fisher (1918) established the modern theory of quantitative genetics, following the early work of Galton, Pearson, Weldon, Yule and others. The equations of selection in quantitative genetics and animal breeding arose from that foundation. Many modern applications of the Price equation to particular problems follow this tradition of quantitative genetics. A criticism of these Price equation applications is a criticism of the central approach of evolutionary quantitative genetics. Such

criticisms may be valid for certain applications, but they must be evaluated in the broader context of quantitative genetics theory. This section shows the relation between quantitative genetics and a commonly applied form of the Price equation (Rice, 2004).

Evolutionary aspects of quantitative genetics developed from the breeder's equation

$$R = Sh^2,$$

in which the response to selection, R, equals the selection differential, S, multiplied by the heritability, h^2. The separation of selection and transmission is the key to the breeder's equation and to quantitative genetics theory.

The covariance term of the Price equation is equivalent to the selection differential, S, when one interprets the meaning of *fitness* and *descendants* in a particular way. Suppose that we label each potential parent in the ancestral population of size N with the index, i. The initial weighting of each parent in the ancestral population is $q_i = 1/N$. Assign to each potential parent a weighting with respect to breeding contribution, $q_i' = q_i w_i$, with fitnesses standardized so that $\bar{w} = 1$ and the w_i are relative fitnesses.

With this set-up, ancestors are the initial population of potential parents, each weighted equally, and descendants are the same population of parents, weighted by their breeding contribution. The character value for each individual remains unchanged between the ancestor and descendant labellings. These assumptions lead to $\Delta \bar{z}^* = \text{Cov}(w, z)$, the change in the average character value between the breeding population and the initial population. That difference is defined as S, the selection differential.

To analyse the fraction of the selection differential transmitted to offspring, classical quantitative genetics follows Fisher (1918) to separate the character value as $z = g + \epsilon$, with a transmissible genetic component, g, and a component that is not transmitted, which we may call the environmental or unexplained component, ϵ. Following standard regression theory for this sort of expression, $\bar{\epsilon} = 0$.

For a parent with $z = g + \epsilon$, the average character value contribution ascribed to the parent among its descendants is $z' = g$, following the idea that g represents the component of the parental character that is transmitted to offspring. If we assume that the only fluctuations of average character value in offspring are caused by the transmissible component that comes from parents, then the genetic component measured by g is sufficient to explain expected offspring character values. Thus, $\Delta z = z' - z = -\epsilon$, and $\text{E}(w \Delta z) = -\text{Cov}(w, \epsilon)$.

Substituting into the full Price equation from eqn 2 and assuming $\bar{w} = 1$ so that all fitnesses are normalized

$$\begin{aligned} \Delta \bar{z} &= \text{Cov}(w, z) + \text{E}(w \Delta z) \\ &= \text{Cov}(w, g) + \text{Cov}(w, \epsilon) - \text{Cov}(w, \epsilon) \\ &= \text{Cov}(w, g). \end{aligned} \quad (12)$$

The expression $\Delta \bar{z} = \text{Cov}(w, g)$ was first emphasized by Robertson (1966) and is sometimes called Robertson's

secondary theorem of natural selection. Robertson's expression summarizes the foundational principles of quantitative genetics, as conceived by Fisher (1918) and developed over the past century (Falconer & Mackay, 1996; Lynch & Walsh, 1998; Hartl, 2006).

It is commonly noted that Robertson's theorem is related to the classic breeder's equation. In particular,

$$R = \Delta \bar{z} = \text{Cov}(w, g) = \text{Cov}(w, z)h^2 = Sh^2,$$

where R is the response to selection, $S = \text{Cov}(w, z)$ is the selection differential and $h^2 = \text{Var}(g)/\text{Var}(z)$ is a form of heritability, a measure of the transmissible genetic component. Additional details and assumptions can be found in several articles and texts (Crow & Nagylaki, 1976; Frank, 1997; Rice, 2004).

Population genetics and the covariance expression

Price (1970) expressed his original formulation in terms of gene frequency change and classical population genetics, rather than the abstract set relations that I have emphasized. At that time, it seems likely that Price already had the broader, more abstract theory in hand and was presenting the population genetics form because of its potential applications. The article begins

> This is a preliminary communication describing applications to genetical selection of a new mathematical treatment of selection in general.
> Gene frequency change is the basic event in biological evolution. The following equation… which gives frequency change under selection from one generation to the next for a single gene or for any linear function of any number of genes at any number of loci, holds for any sort of dominance or epistasis, for sexual or asexual reproduction, for random or nonrandom mating, for diploid, haploid or polyploid species, and even for imaginary species with more than two sexes…

Using my notation, Price writes the basic covariance form

$$\Delta P = \text{Cov}(w, p)/\bar{w} = \beta_{wp} \text{Var}(p)/\bar{w}. \quad (13)$$

In a simple application, p could be interpreted as gene frequency at a single diploid locus with two alleles. Then, $P = \bar{p}$ is the gene frequency in the population, and β_{wp} is the regression of individual fitness on individual gene frequency, in which the individual gene frequency is either 0, 1/2 or 1 for an individual with 0, 1 or 2 copies of the allele of interest. Li (1967) gave an identical gene frequency expression in his eqn 4.

In more general applications, one can study a p-score that summarizes the number of copies of various alleles present in an individual or in whatever entities are being tracked. In classical population genetics, the p-score would be, in Price's words above, 'any linear function of any number of genes at any number of loci'. Here, *linearity* means that p is essentially a counting of presence versus absence of various things within the ith entity.

Such counting does not preclude nonlinear interactions between alleles or those things being counted with respect to phenotype, which is why Price said that the expression holds for any form of dominance or epistasis.

Hamilton (1970) used Price's gene frequency form in his first clear derivations of the direct and the inclusive fitness models of kin selection theory. Most early applications of the Price equation used this gene frequency interpretation.

Price (1970) emphasized that the value of eqn 13 arises from its benefits for qualitative reasoning rather than calculation. The necessary assumptions can be seen from the form given by Price, which is always exact, here written in my notation

$$\Delta P = \mathrm{Cov}(w, p)/\bar{w} + \mathrm{E}(w\Delta p)/\bar{w},$$

where Δp is interpreted as the change in state between parental gene frequency for the ith entity and the average gene frequency for the part of descendants derived from the ith entity.

In practice, $\Delta p = 0$ usually means Mendelian segregation, no biased mutation and no sampling biases associated with drift. Most population genetics theory of traits such as social behaviour typically make those assumptions, so that eqn 13 is sufficient with respect to analysing change in gene frequency or in p-scores (Grafen, 1984). However, the direction of change in gene frequency or p-score is not sufficient to predict the direction of change in phenotype. To associate the direction of change in p-score with the direction of change in phenotype, one must make the assumption that phenotype changes monotonically with p-score. Such monotonicity is a strong assumption, which is not always met. For that reason, p-score models sometimes buy simplicity at a rather high cost. In other applications, monotonicity is a reasonable assumption, and the p-score models provide a very simple and powerful approach to understanding the direction of evolutionary change.

The costs and benefits of the p-score model are not particular to the Price equation. Any analysis based on the same assumptions has the same limitations. The Price equation provides a concise and elegant way to explore the consequences when certain simplifying assumptions can reasonably be applied to a particular problem.

Alternative forms or interpretations of the full equation

The full Price equation partitions total evolutionary change into components. Many alternative partitions exist. A partition provides value if it improves conceptual clarity or eases calculation.

Which partitions are better than others? *Better* is always partly subjective. What may seem hard for me may appear easy to you. Nonetheless, it would be a mistake to suggest that all differences are purely subjective. Some forms are surely better than others for

particular problems, even if *better* remains hard to quantify. As Russell (1958, p. 14) said in another context, 'All such conventions are equally legitimate, though not all are equally convenient'.

Many partitions of evolutionary change include some aspect of selection and some aspect of property or transmission change. Most of those variants arise by minor rearrangements or extensions of the basic Price expression. A few examples follow.

Contextual analysis

Heisler & Damuth (1987) introduced the phrase *contextual analysis* to the evolutionary literature. Contextual analysis is a form of path analysis, which partitions causes by statistical regression models. Path analysis has been used throughout the history of genetics (Li, 1975). It is a useful approach whenever one wishes to partition variation with respect to candidate causes. The widely used method of Lande & Arnold (1983) to analyse selection is a particular form of path analysis.

Okasha (2006) argued that contextual analysis is an alternative to the Price equation. To develop a simple example, let us work with just the selection part of the Price equation

$$\bar{w}\Delta\bar{z} = \mathrm{Cov}(w, z).$$

A path (contextual) analysis refines this expression by partitioning the causes of fitness with a regression equation. Suppose we express fitness as depending on two predictors: the focal character that we are studying, z, and another character, y. Then we can write fitness as

$$w = \beta_{wz}z + \beta_{wy}y + \epsilon$$

in which the β terms are partial regressions of fitness on each character, and ϵ is the unexplained residual of fitness. Substituting into the Price equation, we get the sort of expression made popular by Lande & Arnold (1983)

$$\bar{w}\Delta\bar{z} = \beta_{wz}\mathrm{Var}(z) + \beta_{wy}\mathrm{Cov}(y, z).$$

If the partitioning of fitness into causes is done in a useful way, this type of path analysis can provide significant insight. I based my own studies of natural selection and social evolution on this approach (Frank, 1997, 1998).

Authors such as Okasha (2006) consider the partitioning of fitness into distinct causes as an alternative to the Price equation. If one thinks of the character z in $\mathrm{Cov}(w, z)$ as a complete causal explanation for fitness, then a partition into separate causes y and z does indeed lead to a different causal understanding of fitness. In that regard, the Price equation and path analysis lead to different causal perspectives.

One can find articles that use the Price equation and interpret z as a lone cause of fitness (see Okasha, 2006). Thus, if one equates those specific applications with the

general notion of *the Price equation,* then one can say that path or contextual analysis provides a significantly different perspective from the Price equation. To me, that seems like a socially constructed notion of logic and mathematics. If someone has applied an abstract truth in a specific way, and one can find an alternative method for the same specific application that seems more appealing, then one can say that the alternative method is superior to the general abstract truth.

The abstract Price equation does not compel one to interpret z strictly as a single-cause explanation. Rather, in the general expression, z should always be interpreted as an abstract placeholder. Path (contextual) analysis follows as a natural extension of the Price equation, in which one makes specific models of fitness expressed by regression. It does not make sense to discuss the Price equation and path analysis as alternatives.

Alternative partitions of selection and transmission

In the standard form of the Price equation, the fitness term, w, appears in both components

$$\bar{w}\Delta\bar{z} = \text{Cov}(w, z) + \text{E}(w\Delta z).$$

Frank (1997, 1998) derived an alternative expression

$$
\begin{aligned}
\Delta\bar{z} &= \sum q_i' z_i' - \sum q_i z_i \\
&= \sum q_i (w_i/\bar{w}) z_i' - \sum q_i z_i \\
&= \sum q_i (w_i/\bar{w}) z_i' - \sum q_i z_i' + \sum q_i z_i' - \sum q_i z_i \\
&= \sum q_i (w_i/\bar{w} - 1) z_i' + \sum q_i (z_i' - z_i) \\
&= \text{Cov}(w, z')/\bar{w} + \text{E}(\Delta z). \quad (14)
\end{aligned}
$$

This form sometimes provides an easier method to calculate effects. For example, the second term now expresses the average change in phenotype between parent and offspring without weighting by fitness effects. A biased mutational process would be easy to calculate with this expression – one only needs to know about the mutation process to calculate the outcome. The new covariance term can be partitioned into meaningful components with minor assumptions (Frank, 1997, p. 1721), yielding

$$\text{Cov}(w, z') = \text{Cov}(w, z)\beta_{z'z},$$

where $\beta_{z'z}$ is usually interpreted as the offspring–parent regression, which is a type of heritability. Thus, we may combine selection with the heritability component of transmission into the covariance term, with the second term containing only a fitness-independent measure of change during transmission.

Okasha (2006) strongly favoured the alternative partition for the Price equation in eqn 14, because it separates all fitness effects in the first term from a pure transmission interpretation of the second term. In my view, there are costs and benefits for the standard Price equation expression compared with eqn 14. One gains by having both and using the particular form that fits a particular problem.

For example, the term $\text{E}(\Delta z)$ is useful when one has to calculate the effects of a biased mutational process that operates independently of fitness. Alternatively, suppose most individuals have unbiased transmission, such that $\Delta z = 0$, whereas very sick individuals do not reproduce but, if they were to reproduce, would have a very biased transmission process. Then $\text{E}(\Delta z)$ differs significantly from zero, because the sick, nonreproducing individuals appear in this term equally with the reproducing population. However, the actual transmission bias that occurs in the population would be zero, $\text{E}(w\Delta z) = 0$, because all reproducing individuals have nonbiased transmission.

Both the standard Price form and the alternative in eqn 14 can be useful. Different scenarios favour different ways of expressing problems. I cannot understand why one would adopt an a priori position that unduly limits one's perspective.

Extended set mapping expression

The Price equation's power arises from its abstraction of selection in terms of mapping relations between sets (Frank, 1995; Price, 1995). Although the Price equation is widely cited in the literature, almost no work has developed the set mapping formalism beyond the description given in the initial publications. I know of only one article.

Kerr & Godfrey-Smith (2009) noted that, in the original Price formulation, every descendant must derive from one or more ancestors. There is no natural way for novel entities to appear. In applications, new entities could arise by immigration from outside the system or, in a cultural interpretation, by de novo generation of an idea or behaviour.

Kerr & Godfrey-Smith (2009) present an extended expression to handle unconnected descendants. Their formulation depends on making explicit the connection number between each individual ancestor and each individual descendant, rather than using the fitnesses of types. Some descendants may have zero connections.

With an explicit description of connections, an extended Price equation follows. The two core components of covariance for selection and expected change for transmission occur, plus a new factor to account for novel descendants unconnected to ancestors.

The notation in Kerr & Godfrey-Smith (2009) is complex, so I do not repeat it here. Instead, I show a simplified version. Suppose that a fraction p of the descendants are unconnected to ancestors. Then, we can write the average trait value among descendants as

$$\bar{z}' = p\sum \alpha_j z_j^* + (1 - p)\sum q_i' z_i',$$

where z_j^* is the phenotype for the jth member of the descendant population that is unconnected to ancestors and α_j is the frequency of each unconnected type, with

$\sum \alpha_j = 1$. Given those definitions, we can proceed with the usual Price equation expression

$$\Delta \bar{z} = \bar{z}' - \bar{z}$$
$$= p \sum \alpha_j z_j^* + (1-p) \sum q_i' z_i' - (p+1-p) \sum q_i z_i$$
$$= (1-p) \left(\sum q_i' z_i' - \sum q_i z_i \right) + p \left(\sum \alpha_j z_j^* - \sum q_i z_i \right).$$

Note that the term weighted by $1-p$ leads to the standard form of the Price equation, so we can write

$$\Delta \bar{z} = (1-p)(\mathrm{Cov}(w,z) + \mathrm{E}(w\Delta z))/\bar{w}$$
$$+ p \left(\sum \alpha_j z_j^* - \sum q_i z_i \right)$$
$$= (1-p)(\mathrm{Cov}(w,z) + \mathrm{E}(w\Delta z))/\bar{w} + p(\bar{z}^* - \bar{z}).$$

In the component weighted by p, no connections exist between the descendant z_j^* and a member of the ancestral population. Thus, we have no basis to relate those terms to fitness, transmission or property change. Kerr & Godfrey-Smith (2009) use an alternative notation that associates all entities with their number of connections, including those with zero. The outcome is an extended set mapping theory for evolutionary change. The main concepts and the value of the approach are best explained by the application presented in the next section.

Gains and losses in descendants and ancestors

Fox & Kerr (2012) analyse changes in ecosystem function by modifying the method of Kerr & Godfrey-Smith (2009). They measure ecosystem function by summing the functional contribution of each species present in an ecosystem. To compare ecosystems, they consider an initial site and a second site. When comparing ecosystems, the notion of ancestors and descendants may not make sense. Instead, one appeals to the more general set mapping relations of the Price equation.

Assume that there is an initial site with total function $T = \sum z_i$, where z_i is the function of the ith species. At the initial site, there are s different species; thus, we may also express the total as $T = s\bar{z}$, where $\bar{z}$ is the average function per species. At a second site, total function is $T' = \sum z_j'$, with s' different species in the summation and $T' = s'\bar{z}'$. Let the number of species in common between the sites be s_c. Thus, the initial site has $S = s - s_c$ unique species, and the second site has $S' = s' - s_c$ unique species.

Fox & Kerr (2012) write the change in total ecosystem function as

$$\Delta T = T' - T = s'\bar{z}' - s\bar{z}$$
$$= (s' - s_c)\bar{z}' - (s - s_c)\bar{z} + s_c(\bar{z}' - \bar{z})$$
$$= S'\bar{z}' - S\bar{z} + s_c(\Delta \bar{z}).$$

The term $S'\bar{z}'$ represents the change in function that caused the gain of an average species, in which S' is the number of newly added species, and $\bar{z}'$ is the average function per species. Fox & Kerr (2012) suggest that a

randomly added species would be expected to function as an average species, and so interpret this term as the contribution of random species gain. The term $S\bar{z}$ is interpreted similarly as random species loss with respect to the S unique species in the first ecosystem not present in the second ecosystem.

Fox & Kerr (2012) partition the term $s_c(\Delta \bar{z})$ into three components of species function: deviation from the average for species gained at the second site; deviation from the average for species lost from the first site; and the changes in function for those species in common between sites.

The point here concerns the approach rather than the theory of ecosystem function. To analyse changes between two sets, one often benefits by an explicit decomposition of the relations between the two sets. The original Price equation is one sort of decomposition, based on tracing the ways in which descendants derive from and change with respect to ancestors. Fox & Kerr (2012) extend the decomposition of change by set mapping to include specific components that make sense in the context of changes in ecosystem function.

More work on the mathematics of set mapping and decomposition would be very valuable. The Price equation and the extensions by Kerr, Godfrey-Smith and Fox show the potential for thinking carefully about the abstract components of change between sets and how to apply that abstract understanding to particular problems.

Other examples

No clear guidelines determine what constitutes an extension to the Price equation. From a broad perspective, many different partitions of total change have similarities, because they separate something like selection from other forces that alter the similarity between populations.

For example, the stochastic effects of sampling and drift create a distribution of descendant phenotypes around the ancestral mean. In the classical Price formulation, there is only the single realization of the actual descendants. A stochastic version analyses a collection of possible descendant sets over some probability distribution and a mapping from the ancestor set to each possible realization of the descendant set.

In other cases, partitions will split components more finely or add new components not in Price's formulation. I do not have space to review every partition of total change and consider how each may be related to Price's formulation. I list a few examples here.

Grafen (1999) and Rice (2008) developed stochastic approaches. Grafen (2007) based a long-term project on interpretations and extensions of the Price equation. Page & Nowak (2002) related the Price equation to various other evolutionary analyses, providing some minor extensions. Wolf *et al.* (1998), Bijma & Wade (2008), and many others developed extended partitions by splitting causes with regression or similar methods such as path analysis.

Various forms of the Price equation have been applied in economic theory (Andersen, 2004).

Difficulties with various critiques of the Price equation

> A reliable way to make people believe in falsehoods is frequent repetition, because familiarity is not easily distinguished from truth (Kahneman, 2011, p. 62).

One must distinguish the full, exact Price equation from various derived forms used in applications. The derived forms always make additional assumptions or express approximate relations (Frank, 1997). Each assumption increases specificity and reduces generality in relation to particular goals.

Critiques of the Price equation rarely distinguish the costs and benefits of particular assumptions in relation to particular goals. I use van Veelen's recent series of papers as a proxy for those critiques. That series repeats some of the common misunderstandings and adds some new ones. Nowak recently repeated van Veelen's critique as the basis for his commentary on the Price equation (van Veelen, 2005; Nowak *et al.*, 2010; van Veelen *et al.*, 2010; Nowak & Highfield, 2011; van Veelen, 2011; van Veelen *et al.*, 2012).

Dynamic sufficiency

The Price equation describes the change in some measurement, expressed as $\Delta\bar{z}$. Change is calculated with respect to particular mapping relations between ancestor and descendant populations. We can think of the mappings and the beginning value of $\bar{z}$ as the initial conditions or inputs and $\Delta\bar{z}$ as the output.

The output, $\bar{z}' = \bar{z} + \Delta\bar{z}$, does not provide enough information to iterate the calculation of change to get another value of $\Delta\bar{z}$ starting with $\bar{z}'$. We would also need the mapping relations between the new descendant population and its subsequent descendants. That information is not part of the initial input. Thus, we cannot study the dynamics of change over time without additional information.

This limitation with regard to repeated iteration is called a lack of *dynamic sufficiency* (Lewontin, 1974). Confusion about the nature of dynamic sufficiency in relation to the Price equation has been common in the literature. In Frank (1995, pp. 378–379), I wrote

> It is not true, however, that dynamic sufficiency is a property that can be ascribed to the Price Equation—this equation is simply a mathematical tautology for the relationship among certain quantities of populations. Instead, dynamic sufficiency is a property of the assumptions and information provided in a particular problem, or added by additional assumptions contained within numerical techniques such as diffusion analysis or applied quantitative genetics. … What problems can the Price equation solve that cannot be solved by other methods?

The answer is, of course, none, because the Price Equation is derived from, and is no more than, a set of notational conventions. It is a mathematical tautology.

I showed how the Price equation helps to define the necessary conditions for dynamic sufficiency. Once again, the Price equation proves valuable for clarifying the abstract structure of evolutionary analysis.

Compare my statement with that of van Veelen *et al.* (2012)

> Dynamic insufficiency is regularly mentioned as a drawback of the Price equation (see for example Frank, 1995; Rice, 2004). We think that this is not an entirely accurate description of the problem. We would like to argue that the perception of dynamic insufficiency is a symptom of the fundamental problem with the Price equation, and not just a drawback of an otherwise fine way to describe evolution. To begin with, it is important to realize that the Price equation itself, by its very nature, cannot be dynamically sufficient or insufficient. The Price equation is just an identity. If we are given a list of numbers that represent a transition from one generation to the next, then we can fill in those numbers in both the right and the left hand side of the Price equation. The fact that it is an identity guarantees that the numbers that appear on both sides of the equality sign are the same. There is nothing dynamically sufficient or insufficient about that (this point is also made by Gardner et al., 2007, p. 209). A model, on the other hand, can be dynamically sufficient or insufficient.

This quote from van Veelen *et al.* (2012) demonstrates an interesting approach to scholarship. They first cite Frank as stating that dynamic insufficiency is a drawback of the Price equation. They then disagree with that point of view and present as their own interpretation an argument that is nearly identical in concept and phrasing to my own statement in the very paper that they cited as the foundation for their disagreement.

In this case, I think it is important to clarify the concepts and history, because influential and widely cited authors, such as Nowak, are using van Veelen's articles as the basis for their own critiques of the Price equation and approaches to fundamental issues of evolutionary analysis.

With regard to dynamics, any analysis achieves the same dynamic status given the same underlying assumptions. The Price equation, when used with the same underlying assumptions as population genetics, has the same attributes of dynamic sufficiency as population genetics.

Interpretation of covariance

van Veelen *et al.* (2012) claim that

> Maybe the most unfortunate thing about the Price equation is that the term on the right hand side is denoted as a covariance, even though it is not. The equation thereby turns into something that can easily set us off in the wrong direction, because it now resembles equations as they feature in other sciences, where probabilistic models are used that do use actual covariances.

One can see the covariance expression in the standard form of the Price equation given in eqn 2. In the Price equation, the covariance is measured with respect to the total population; in other words, it expresses the association over all members of the population. In many statistical applications, one only has data on a subset of the full population, that subset forming a sample. It is important to distinguish between population measures and sample measures, because they refer to different things.

Price (1972a, p. 485) made clear that his equation is about total change in entire populations, so the covariance is interpreted as a population measure

> [W]e will be concerned with population functions and make no use of sample functions, hence we will not observe notational conventions for distinguishing population and sample variables and functions.

In addition to population and sample measures, covariance also arises in mathematical models of process. Suppose, for example, that I develop a model in which random processes influence fitness and random processes influence phenotype. If the random fluctuations in fitness and the random fluctuations in phenotype are associated, the random variables of fitness and phenotype would covary. All of these different interpretations of covariance are legitimate; they simply reflect different situations.

Discussion

In Frank (1995), I wrote: 'What problems can the Price equation solve that cannot be solved by other methods? The answer is, of course, none, because the Price Equation is derived from, and is no more than, a set of notational conventions. It is a mathematical tautology'.

Nowak & Highfield (2011) and van Veelen *et al.* (2012) emphasize the same point in their critique of the Price equation, although they present the argument as a novel insight without attribution. Given that the Price equation is a set of notational conventions, it cannot uniquely specify any predictions or insights. A particular set of assumptions leads to the same predictions, no matter what notational conventions one uses. The Price equation is a tool that sometimes helps in analysis or in seeing general connections between apparently disparate ideas. For many problems, the Price equation provides no value, because it is the wrong tool for the job.

If the Price equation is just an equivalence, or tautology, then why am I enthusiastic about it? Mathematics is, in its essence, about equivalences, as expressed beautifully in the epigraph from Mazur. Not all equivalences are interesting or useful, but some are, just as not all mathematical expressions are interesting or useful, but some are.

That leads us to the question of how we might know whether the Price equation is truly useful or a mere identity? It is not always easy to say exactly what makes an abstract mathematical equivalence interesting or useful. However, given the controversy over the Price equation, we should try. Because there is no single answer, or even a truly unique and unambiguous question, the problem remains open. I list a few potential factors.

'[A] good notation has a subtlety and suggestiveness which at times make it seem almost like a live teacher' (Russell, 1922, pp. 17–18). Much of creativity and understanding comes from seeing previously hidden associations. The tools and forms of expression that we use play a strong role in suggesting connections and are inseparable from cognition (Kahneman, 2011). Equivalences and alternative notations are important.

The various forms of the covariance component from the Price equation given in eqn 9 show the equivalence of the statistical, geometrical and informational expressions for natural selection. The recursive form of the full Price equation provides the foundation for all modern studies of group selection and multilevel analysis. The Price equation helped in discovering those various connections, although there are many other ways in which to derive the same relations.

Hardy (1967) also emphasized the importance of seeing new connections between apparently disparate ideas:

> We may say, roughly, that a mathematical idea is 'significant' if it can be connected, in a natural and illuminating way, with a large complex of other mathematical ideas. Thus a serious mathematical theorem, a theorem which connects significant ideas, is likely to lead to important advances in mathematics itself and even in other sciences.

What sort of connections? One type concerns the invariances discovered or illuminated by the Price equation. I discussed some of those invariances in an earlier section, particularly the information theory interpretation of natural selection through the measure of Fisher information (Frank, 2009). Fisher's fundamental theorem of natural selection is a similar sort of invariance (Frank, 2012b). Kin selection theory derives much of its power by identifying an invariant informational quantity sufficient to unify a wide variety of seemingly disparate processes (Frank, 1998, Chapter 6). The interpretation of kin selection as an informational invariance has not been fully developed and remains an open problem.

Invariances provide the foundation of scientific understanding: 'It is only slightly overstating the case to say that physics is the study of symmetry' (Anderson, 1972). *Invariance* and *symmetry* mean the same thing (Weyl, 1983). Feynman (1967) emphasized that invariance is *The Character of Physical Law*. The commonly observed patterns of probability can be unified by the study of invariance and its association with measurement (Frank & Smith, 2010, 2011). There has been little effort in biology to pursue similar understanding of invariance and measurement (Frank, 2011; Houle *et al.*, 2011).

Price argued for the great value of abstraction, in the sense of the epigraph from Mazur. In Price (1995)

[D]espite the pervading importance of selection in science and life, there has been no abstraction and generalization from genetical selection to obtain a general selection theory and general selection mathematics. Instead, particular selection problems are treated in ways appropriate to particular fields of science. Thus one might say that 'selection theory' is a theory waiting to be born—much as communication theory was 50 years ago. Probably the main lack that has been holding back any development of a general selection theory is lack of a clear concept of the general nature or meaning of 'selection'.

This article has been about the Price equation in relation to its abstract properties and its connections to various topics, such as information or fundamental invariances. Some readers may feel that those aspects of abstraction and invariance are nice, but far from daily work in biology. What of the many applications of the Price equation to kin or group selection? Do those applications hold up? How much value has been added?

Because the Price equation is a tool, one can always arrive at the same result by other methods. How well the Price equation works depends partly on the goal and partly on the fit of the tool to the problem. There is inevitably a strongly subjective aspect to deciding about how well a tool works. Nonetheless, hammers truly are good for nails and bad for screws. For valuing tools, there is a certain component that should be open to agreement. For example, the Robertson (1966) form of the Price equation is widely regarded as the foundational method for analysing models of evolutionary quantitative genetics. However, not all problems in quantitative genetics are best studied with the Robertson–Price equation. And not all problems in social evolution benefit from a Price equation approach.

The Price equation or descendant methods have led to many useful models for kin selection (Frank, 1998). The most powerful follow a path analysis decomposition of causes or use a simple maximization method to analyse easily what would otherwise have been difficult. I will return to those applications in subsequent articles.

Acknowledgments

References

Alizon, S. 2009. The Price equation framework to study disease within-host evolution. *J. Evol. Biol.* **22**: 1123–1132.

Andersen, E.S. 2004. Population thinking, Price's equation and the analysis of economic evolution. *Evol. Inst. Econ. Rev.* **1**: 127–148.

Anderson, P. 1972. More is different. *Science* **177**: 393–396.

Bijma, P. & Wade, M.J. 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J. Evol. Biol.* **21**: 1175–1188.

Boulding, K.E. 1956. General systems theory: the skeleton of science. *Manage. Sci.* **2**: 197–208.

Charlesworth, B. & Charlesworth, D. 2010. *Elements of Evolutionary Genetics.* Roberts & Company, Greenwood Village, CO.

Coplien, J.O. 1998. To iterate is human, to recurse, divine. *C++ Rep.* **10**: 43–51.

Crow, J.F. & Kimura, M. 1970. *An Introduction to Population Genetics Theory.* Burgess, Minneapolis, Minnesota.

Crow, J.F. & Nagylaki, T. 1976. The rate of change of a character correlated with fitness. *Am. Nat.* **110**: 207–213.

Day, T. & Gandon, S. 2006. Insights from Price's equation into evolutionary epidemiology. In: *Disease Evolution: Models, Concepts, and Data Analysis* (Z. Feng, U. Dieckmann & S. Levin, eds), pp. 23–44. American Mathematical Society, Washington, DC.

Ewens, W.J. 1989. An interpretation and proof of the fundamental theorem of natural selection. *Theor. Popul. Biol.* **36**: 167–180.

Ewens, W.J. 1992. An optimizing principle of natural selection in evolutionary population genetics. *Theor. Popul. Biol.* **42**: 333–346.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics,* 4th edn. Longman, Essex, UK.

Feynman, R.P. 1967. *The Character of Physical Law.* MIT Press, Cambridge, MA.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**: 399–433.

Fisher, R.A. 1930. *The Genetical Theory of Natural Selection.* Clarendon, Oxford.

Fisher, R.A. 1941. Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**: 53–63.

Fox, J.W. 2006. Using the Price Equation to partition the effects of biodiversity loss on ecosystem function. *Ecology* **87**: 2687–2696.

Fox, J.W. & Kerr, B. 2012. Analyzing the effects of species gain and loss on ecosystem function using the extended Price equation partition. *Oikos* **121**: 290–298.

Frank, S.A. 1995. George Price's contributions to evolutionary genetics. *J. Theor. Biol.* **175**: 373–388.

Frank, S.A. 1997. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**: 1712–1729.

Frank, S.A. 1998. *Foundations of Social Evolution.* Princeton University Press, Princeton, NJ.

Frank, S.A. 2009. Natural selection maximizes Fisher information. *J. Evol. Biol.* **22**: 231–244.

Frank, S.A. 2011. Measurement scale in maximum entropy models of species abundance. *J. Evol. Biol.* **24**: 485–496.

Frank, S.A. 2012a. Natural selection. III. Selection versus transmission and the levels of selection. *J. Evol. Biol.* **25**: 227–243.

Frank, S.A. 2012b. Wright's adaptive landscape versus Fisher's fundamental theorem. In: *The Adaptive Landscape in Evolutionary Biology* (E. Svensson & R. Calsbeek, eds). Oxford University Press, New York (in press). Available from http://arxiv.org/abs/1102.3709v1.

Frank, S.A. & Slatkin, M. 1990. The distribution of allelic effects under mutation and selection. *Genet. Res.* **55**: 111–117.

Frank, S.A. & Slatkin, M. 1992. Fisher's fundamental theorem of natural selection. *Trends Ecol. Evol.* **7**: 92–95.

Frank, S.A. & Smith, E. 2010. Measurement invariance, entropy, and probability. *Entropy* **12**: 289–303.

Frank, S.A. & Smith, E. 2011. A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *J. Evol. Biol.* **24**: 469–484.

Frieden, B.R. 2004. *Science from Fisher Information: A Unification*. Cambridge University Press, Cambridge, UK.

Frieden, B.R., Plastino, A. & Soffer, B.H. 2001. Population genetics from an information perspective. *J. Theor. Biol.* **208**: 49–64.

Gardner, A., West, S.A. & Barton, N.H. 2007. The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**: 207–226.

Gardner, A. 2008. The Price equation. *Curr. Biol.* **18**: R198–R202.

Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural ecology* (J.R. Krebs & N.B. Davies, eds), pp. 62–84. Blackwell Scientific Publications, Oxford.

Grafen, A. 1999. Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proc. R. Soc. Lond. B* **266**: 799–803.

Grafen, A. 2002. A first formal link between the Price equation and an optimization program. *J. Theor. Biol.* **217**: 75–91.

Grafen, A. 2007. The formal Darwinism project: a mid-term report. *J. Evol. Biol.* **20**: 1243–1254.

Hamilton, W.D. 1970. Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**: 1218–1220.

Hamilton, W.D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: *Biosocial Anthropology* (R. Fox, ed.), pp. 133–155. Wiley, New York.

Hardy, G.H. 1967. *A Mathematician's Apology*. Cambridge University Press, Cambridge, UK.

Harman, O.S. 2010. *The Price of Altruism*. Norton, New York.

Hartl, D.L. 2006. *Principles of Population Genetics*, 4th edn. Sinauer, Sunderland, MA.

Heisler, I.L. & Damuth, J. 1987. A method for analyzing selection in hierarchically structured populations. *Am. Nat.* **130**: 582–602.

Houle, D., Pélabon, C., Wagner, G.P. & Hansen, T.F. 2011. Measurement and meaning in biology. *Q. Rev. Biol.* **86**: 3–34.

Kahneman, D. 2011. *Thinking, Fast and Slow*. Farrar Straus & Giroux, New York.

Kerr, B. & Godfrey-Smith, P. 2009. Generalization of the Price equation for evolutionary change. *Evolution* **63**: 531–536.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1212–1226.

Lewontin, R.C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Li, C.C. 1967. Fundamental theorem of natural selection. *Nature* **214**: 505–506.

Li, C.C. 1975. *Path Analysis*. Boxwood, Pacific Grove, CA.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Mazur, B. 2008. When is one thing equal to some other thing? In: *Proof and Other Dilemmas: Mathematics and Philosophy* (B. Gold & R.A. Simons, eds), pp. 221–241. Mathematical Association of America, Washington, DC.

Michod, R.E. 1997a. Cooperation and conflict in the evolution of individuality. I. Multilevel selection of the organism. *Am. Nat.* **149**: 607–645.

Michod, R.E. 1997b. Evolution of the individual. *Am. Nat.* **150**: S5–S21.

Nowak, M.A. & Highfield, R. 2011. *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Free Press, New York.

Nowak, M.A., Tarnita, C.E. & Wilson, E.O. 2010. The evolution of eusociality. *Nature* **466**: 1057–1062.

Okasha, S. 2006. *Evolution and the Levels of Selection*. Oxford University Press, New York.

Page, K.M. & Nowak, M.A. 2002. Unifying evolutionary dynamics. *J. Theor. Biol.* **219**: 93–98.

Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.

Price, G.R. 1972a. Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**: 485–490.

Price, G.R. 1972b. Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet.* **36**: 129–140.

Price, G.R. 1995. The nature of selection. *J. Theor. Biol.* **175**: 389–396.

Queller, D.C. 1992a. A general model for kin selection. *Evolution* **46**: 376–380.

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

Rice, S.H. 2004. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer Associates, Sunderland, MA.

Rice, S.H. 2008. A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evol. Biol.* **8**: 262.

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

Russell, B. 1922. *Introduction to Tractatus Logico-Philosophicus by Ludwig Wittgenstein*. Harcourt, Brace and Company, New York.

Russell, B. 1958. *The ABC of Relativity*. New American Library, New York.

Schwartz, J. 2000. Death of an altruist. *Lingua Franca* **10**: 51–61.

van Veelen, M. 2005. On the use of the Price equation. *J. Theor. Biol.* **237**: 412–426.

van Veelen, M. 2011. A rule is not a rule if it changes from case to case (a reply to Marshall's comment). *J. Theor. Biol.* **270**: 189–195.

van Veelen, M., García, J., Sabelis, M.W. & Egas, M. 2010. Call for a return to rigour in models. *Nature* **466**: 661–661.

van Veelen, M., García, J., Sabelis, M. & Egas, M. 2012. Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics. *J. Theor. Biol.* **299**: 64–80.

Wade, M.J. 1985. Soft selection, hard selection, kin selection, and group selection. *Am. Nat.* **125**: 61–73.

Weyl, H. 1983. *Symmetry*. Princeton University Press, Princeton, NJ.

Wolf, J.B., Brodie, E.D., Cheverud, J.M., Moore, A.J. & Wade, M.J. 1998. Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**: 64–69.