

The Price Equation Since Price: An Accessible Account and a Generalization to Categorical Variables

Stephen Francis Mann*

The Price equation is usually treated as a description of how the population average value of a trait changes due to selection and other factors. Despite the fact that its generality is often emphasised, the Price equation is typically only applied to numeric traits, like weight and height. After a thorough yet accessible introduction to the numeric form, I derive a version of the Price equation for categorical traits, like colour and shape. The new equation describes how the distribution of types, rather than the population average, changes due to selection and other factors. I argue that this is a useful and conceptually sound extension of the traditional Price equation formalism. Although categorical traits can be represented numerically via dummy coding, I argue that the new version of the equation introduces an important perspective that previous versions lack: selection does not just change averages, it changes distributions.

Keywords

price equation • selection • evolutionary biology • cultural evolution

1. Introduction

The Price equation is one of the best-known results in theoretical biology. As usually introduced, it describes change in the average trait value of a population as a consequence of two neatly distinguishable factors: selection, and everything else. By distinguishing change due to selection from other sources of change, the equation suggests a formal definition corresponding to the informal concept of selection at the heart of evolutionary biology. While formalisms of this kind had been presented before (e.g., Fisher 1930; Haldane 1932), one distinguishing aspect of the Price equation is that it is substrate-neutral, indicating an extension of selectionist thinking to domains beyond biology (Strand et al. 2022; Knudsen 2004) including cultural evolution (El Mouden et al. 2014; Jäger 2010).

*Max Planck Institute for Evolutionary Anthropology, Germany, stephenfmann@gmail.com

Received 05 December 2023; Revised 15 March 2024; Accepted 29 August 2024
doi:10.3998/ptpbio.5334



Since George Price's derivation of the equation (Price 1970), much has been written about its interpretation and application, and discussion continues (Okasha and Otsuka 2020; Gardner 2020; van Veelen 2020; Frank 2012a, 2012b; Luque 2017). However, discussions of the Price equation treat the traits whose change is at issue as *numeric*. That is, the traits in question are properties like height and weight: they are represented by numeric variables. The quantity whose change the equation describes in these cases is the average trait value across the population, such as the average height or average weight. If, like Price (1995), we are tempted to seek a general theory of selection, it does not seem plausible that the only traits subject to selection processes are those associated with numbers. Populations can be subject to selection on any trait whatsoever, including *categorical* traits such as colour and shape. Although techniques such as dummy coding can be employed to represent categorical traits in numerical form, these are workarounds that are used because no better method is available. Since the standard Price equation was developed with numeric traits in mind, this poses the question: is there a better way to represent selection on categorical traits?

Here, I derive a version of the Price equation that applies to categorical traits. Categorical variables representing these traits are one-hot vectors. As I explain below, one-hot vectors are mathematical objects well-suited to representing discrete properties that are not numerically measurable. I show that several known features of the numeric Price equation carry over to the categorical form, including its application to multi-level evolutionary dynamics. This shift in perspective motivates us to think about how selection changes the distribution of types in a population, not just the average value of some measurable quantity.

I begin with an overview of the general approach to selection developed by Price and the derivation of the numeric Price equation (section 2). I then derive a categorical Price equation (section 3) and demonstrate some applications (section 4). Section 5 discusses outstanding issues and section 6 concludes.

2. The Numeric Price Equation

This section lays out the standard form of the Price equation. I give a rather detailed introduction to the formalism, so that readers new to the topic will be able to get up to speed. Readers familiar with the Price equation literature can safely skip to section 2.4, where the limitations of the traditional approach are discussed and the new work in the rest of the paper is motivated.

2.1. The Framework

In a posthumously published manuscript, George Price (1995) stated his belief that the concept of selection applies far more generally than evolutionary biology. The putative generality of selection has been discussed by many scholars with varying aims and scope (for a small sample, see D. T. Campbell 1956; Dennett 1995; Hull 2001; Hodgson 2004; Popper 1972; Csányi 1980; J. O. Campbell 2016). Price agreed with the core contention of these theorists that selection should be construed as a process by which variant types in a population are differentially retained. This informal concept is substrate-neutral and, when the term 'population' is read sufficiently broadly, applies to a great deal of phenomena both in everyday life and across scientific disciplines. Consider the following examples, given by Price, of diverse processes satisfying the informal characterisation:

- A grocer has many apples, some good, some bad. A discerning shopper picks only good quality apples, placing them in her basket. Treating the grocer's apples as the initial population and the shopper's chosen apples as the subsequent population, there is higher average quality in the subsequent population. There has therefore been *selection on quality*.
- A population of organisms possesses many alleles, some better at producing descendants that will survive and reproduce, some worse. In the course of environmental interactions, those organisms with better survival prospects produce successful gametes, i.e., gametes that in fact go on to constitute offspring. The initial population is all the alleles and the subsequent population is the collection of alleles possessed by successful gametes. There is higher average reproductive success among successful gametes' alleles, because all have proven to be at least somewhat adept at contributing to successfully reproducing organisms. There has therefore been *selection on reproductive success*.
- A row of flasks contains a solution at different concentrations. A chemist mixes them by taking more solution from flasks with greater concentration, and less from flasks with lower concentration. The resulting mixture will have a greater concentration than the average of the original set of flasks. There has therefore been *selection on concentration*.

To unify these and many other examples Price developed a formal framework within which populational processes from any domain could be represented. He then defined a formal criterion by which it can be determined, for any change in the average trait value of a population over time, whether or not selection has occurred. To do so he introduced formal tools founded on three basic definitions: packages, trait values, and population share.

First, Price conceived of *packages* as the population variants on which selection can operate. Packages are the different types that make up the population whose trait values are changing over time. The packages of which a population is comprised must stand in one-to-one correspondence before and after selection (figure 1). The populations in question can be made up of concrete objects such as apples or chemical solutions, informational sequences such as alleles, or even abstract entities such as pieces of music. Packages are indexed by i .

Second, *trait values* are properties of packages. Apples can be of good or bad quality, chemical solutions can have different concentrations, alleles can have different reproductive success. Traits with respect to which selection takes place are labelled z . The trait value of package i before selection is z_i . For example, suppose we define the trait *apple quality* such that the worst apples have $z = 0$ and the best apples $z = 1$, with others taking intermediate values. Every individual apple $1, \dots, i, \dots, N$ is assigned a trait value $z_1, \dots, z_i, \dots, z_N$. Packages after selection are categorised in terms of which original package they are descended from, and z'_i denotes the average quality of package i 's descendants. For the apples in the grocery example, every apple's descendant is simply itself. Supposing individual apples do not change quality during the time the shopper makes their decision, then $z'_i = z_i$ for all i . On the other hand, supposing handling apples tends to degrade their quality, and the shopper handles every apple in the shop before making a choice, then $z'_i \leq z_i$ for all i .

Finally, *population share* p_i measures how prevalent each package is in the population. In the case of chemical solution, this is just the volume of solution in each flask. For discrete items such as apples, we can model each package as having the same population share. It will be mathematically convenient to define each package's population share such that the sum of population shares is 1. When the population share of every package is equal, each has the value $\frac{1}{N}$, where N is the total number of packages (likewise, the total volume of chemical solution could be normalised to sum to 1). It is sometimes helpful to think of p_i as the probability of

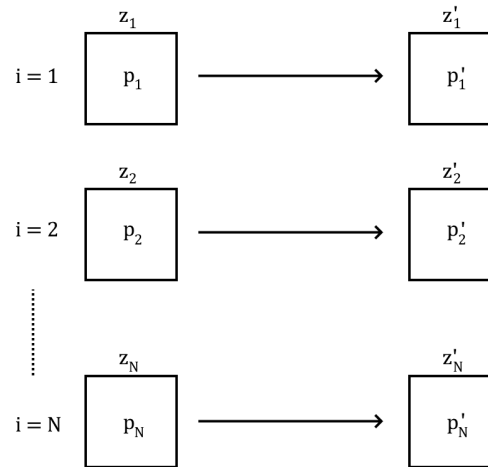


Figure 1: **Price's selection framework.** A population contains N types conceived as distinct 'packages' and indexed by i . Each package has population share p_i at some earlier time and its descendants have population share p'_i at some later time. The trait z is a measurable property of all entities in the population. Trait value z_i is the value of this property for the entities in package i , and the subsequent trait value z'_i is the average value of this property across the descendants of package i . Adapted and relabelled from Price (1995, fig. 4, 392).

observing package i if you sampled randomly from the population. This is the case with the apples: if we sample randomly, every apple has an equal chance of being observed, so each has a population share of $\frac{1}{N}$. The population share of i 's descendant is p'_i . Apples not chosen have $p'_i = 0$, because randomly sampling from the shopper's basket yields zero probability of observing an apple that isn't in it. Again the p'_i s are defined such that they sum to 1. Assuming at least some apples stay in the shop, the total number of selected apples is smaller than the starting set. Therefore, each p'_i is greater than each p_i : apples increase their population share by being selected.

It is important to remember that the terms describing the future population, z'_i and p'_i , relate to descendants of individuals that were type i in the original population, and not to individuals who happen to be type i in the subsequent population (see again figure 1). For example, if a population of organisms is distinguished by their height, package 1 might contain individuals with a height of $z_1 = 160$ cm while package 2 contains individuals with a height of $z_2 = 170$ cm. If an individual from package 1 has an offspring who, due to mutation, grows to be 170 cm, they count towards the average subsequent trait value z'_1 , not z'_2 , and contribute to the subsequent population share p'_1 , not p'_2 .

2.2. Putting the Framework to Work

Further useful quantities can be derived from these basic terms. The average quality of all the apples in the grocer's shop, weighted by their population share, is the total weighted quality divided by the total share (Price 1995, 391; variables changed for consistency):

$$\begin{aligned}\bar{z} &= \frac{\sum p_i z_i}{\sum p_i} \\ &= \frac{\sum p_i z_i}{1} \\ &= \sum p_i z_i\end{aligned}$$

The Greek letter capital sigma, \sum , denotes a sum of all the terms indexed by i (this symbol is used because both ‘sigma’ and ‘sum’ begin with the letter s). So $\sum p_i$ says ‘add together all the values of p_i ,’ while $\sum p_i z_i$ says ‘add together all the products $p_i z_i$.’ After selection, the average quality of apples in the shopper’s basket is calculated the same way:

$$\begin{aligned}\bar{z}' &= \frac{\sum p'_i z'_i}{\sum p'_i} \\ &= \sum p'_i z'_i\end{aligned}$$

If the shopper is discerning, \bar{z}' will be larger than \bar{z} . In this case the *total change* in average quality $\Delta\bar{z} = \bar{z}' - \bar{z}$ will be positive. Here Δ is the Greek letter capital delta, which is often used to signify the change or difference in a value (because both ‘delta’ and ‘difference’ begin with d). If the shopper instead chooses randomly, we would expect the average value of their apples to be the same as the average before selection. Then the total change is zero.

Finally, the *selection coefficient* w_i of package i is defined as the ratio of its population share before and after selection, $w_i = \frac{p'_i}{p_i}$. In biological settings it is usually considered synonymous with *relative fitness*. The selection coefficient of an apple that is not chosen is 0, because its p'_i is defined as 0. The selection coefficient of chosen apples is greater than 1, because each p'_i is greater than the corresponding p_i . From this definition, package i has been selected when its later population share is larger than its earlier population share; that is, when $w_i > 1$. Conversely the package is selected against when $w_i < 1$.

2.3. The Price Equation

When there is a change in the average value of a trait, how do we know whether that change is due to selection? We cannot ignore the fact that processes other than selection affect trait values. Rough handling degrades the quality of apples. Alleles can mutate during transmission, leading them to have diminished reproductive success. A quantity of solvent might evaporate from a solution, leaving it at a slightly higher concentration. All of these processes affect the value of \bar{z}' , and so contribute to the total change $\Delta\bar{z}$.

In order to isolate the effect of selection, it would be helpful to neatly split the total change into two pieces:

$$\Delta\bar{z} = \Delta_s\bar{z} + \Delta_t\bar{z} \quad (1)$$

$\Delta_s \bar{z}$ would be the change in average trait value due to selection, hence the subscript s .¹ $\Delta_t \bar{z}$ would be the change in average trait value due to other factors. Originally these other factors were associated with the concept of transmission bias, hence the t . We'll see later that more than just transmission bias falls under this second term.

Writing down equation (1) is all very well. But there seems to be no guarantee that $\Delta \bar{z}$ separates neatly into two parts and no guarantee that one of these parts corresponds to the informal notion of selection. In deriving his eponymous equation, Price showed that $\Delta \bar{z}$ does indeed separate into two parts and further argued that one of these captures change due to selection. Price (1970) showed that the total change can be written as follows (Okasha (2006, 22) gives a derivation in terms similar to the notation used here):

$$\Delta \bar{z} = \underbrace{\left(\sum p'z - \sum pz \right)}_{\text{change due to selection}} + \underbrace{\left(\sum p'z' - \sum p'z \right)}_{\text{change due to transmission}}$$

Price argued that the first set of brackets should be identified with $\Delta_s \bar{z}$, the change due to selection, while the second set of brackets should be identified with $\Delta_t \bar{z}$, all other sources of change. It's not obvious that those terms capture our pretheoretic notion of selection, so the equation bears some examination. Let's take a closer look at the selection term:

$$\Delta_s \bar{z} = \sum p'z - \sum pz \quad (2)$$

This is a difference between two weighted averages. We are comparing *the trait values of the initial set weighted by the population share of their descendants* with *the trait values of the initial set weighted by their own population share*. If a certain trait value leads to proportionally greater descendants than its current share, its component of this sum will be positive. If a certain trait value leads to proportionally fewer descendants than its current share, its component of this sum will be negative. In short, we are quantifying *the change in population share that can be associated with trait values*. This is precisely the sense in which trait values can be selected.²

The transmission term accounts for a different source of change:

$$\Delta_t \bar{z} = \sum p'z' - \sum p'z \quad (3)$$

This is the difference between the actual subsequent average trait value ($\sum p'z'$), and what that average would have been if trait values didn't change (i.e. if $z' = z$, so $\sum p'z$). The term therefore captures the deviation from perfect retention of package trait values over time. In other words, it describes how trait values within packages change, regardless of changes in population share. This is why the term is said to capture sources of change other than selection.

1. As is well known, the term labelled $\Delta_s \bar{z}$ here turns out to include certain kinds of change due to random drift (Okasha 2006, §1.4.1). For reasons of space I do not discuss drift in this paper. The models can be treated as simplifications of real-life processes in which drift is stipulated not to occur.

2. Here and throughout, terms describing selection do not distinguish between selection-for and selection-against. In other words, the change in population share is associated with trait values but is not necessarily caused by them (Sober 1984, §3.2). It is well known that the Price equation cannot by itself determine whether trait values were causally responsible for selection; this will also be true of the categorical version presented below. For the purposes of this paper I sideline this issue.

Here is a worked example. Suppose the shopkeeper has three apples, of good, medium, and bad quality: $z = 0.9, 0.5, 0.1$. A shopper comes along and picks the good apple. Before selection, each apple has the same population share: $p = \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. The average trait value is $\sum pz = \frac{0.9+0.5+0.1}{3} = \frac{1.5}{3} = 0.5$. After selection, the shopper has chosen just the good apple. Population shares are now $p' = 1, 0, 0$, and the average trait value is $\sum p'z' = (1 \times 0.9) + (0 \times 0.5) + (0 \times 0.1) = 0.9$. The total change in average trait value is $\sum p'z' - \sum pz = 0.9 - 0.5 = 0.4$. The change in average trait value due to selection is $\sum p'z - \sum pz = 0.9 - 0.5 = 0.4$. The total change is equal to the change due to selection. They are equal because in this example apples don't independently change in quality over time. All the change is accounted for by selection. Suppose instead apples *do* change their quality over time. In particular, suppose the shopper roughly handles all apples in order to determine which is best. Rough handling degrades apple quality by 0.05. Then $z' = 0.85, 0.45, 0.05$. Now the overall change is $0.85 - 0.5 = 0.35$. The change due to selection remains the same, 0.4. But the change due to other factors is -0.05 . Figure 2 depicts the two sources of change contributing to the overall change.³

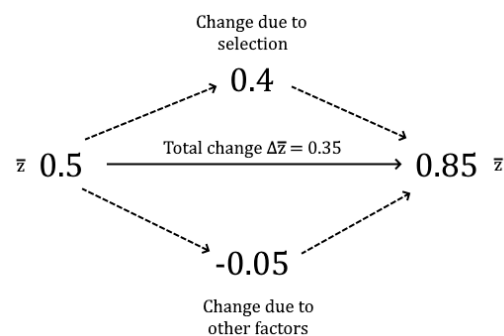


Figure 2: The numeric Price equation partitions change in average trait value $\Delta \bar{z}$ into two parts: change due to selection, and everything else.

2.4. The Price Equation Assumes Variables Are Numeric

The Price equation is almost universally treated as applying to numeric variables.⁴ This can be seen from the fact that the change due to selection, $\Delta_s \bar{z}$, is almost always manipulated into the following statistical form (Okasha 2006, 22):

$$\Delta_s \bar{z} = \text{cov}_p(W, Z)$$

3. What if the shopper chooses randomly, but happens to choose the best apple by accident? This would also lead to a positive value for the selection term, but we would not think of it as 'selection'. This kind of case is akin to drift. Again, I stipulate that drift does not occur in these models. All changes that fall under the selection term are due to inherent properties of the entities in question (or at least relational properties between the entities and some aspect of the selection process, such as the shopper's visual field). How to model the difference between selection and drift is discussed by Okasha (2006, §1.4.1). How to determine whether a real-life process is a process of selection or drift – and whether there is always a clean-cut distinction – is a huge question not to be treated here.

4. One possible exception is Okasha (2006, 24), who mentions using the Price equation to model selection on blue-coloured entities in a population. However, he does this by converting the trait 'blueness' into a numeric variable using a dummy-coding scheme (to be introduced later in this section), so perhaps this does not count. I am unaware of discussion of the Price equation and categorical traits that does not counsel translating those traits into numeric variables. Even Price (1995, 394), when talking about selection for "attributes of mood and subject matter" in pieces of music, suggests that they must be "quantitatively evaluated".

The term $\text{cov}_p(W, Z)$ is the *covariance* between selection coefficients W and trait values Z (which are now being treated as random variables, hence are capitalised). Covariance is a statistical measure of two numeric variables that describes their linear relationship. If they are both high for the same packages and both low for the same packages, we say they positively covary, and their covariance is positive. If one is low while the other is high and vice versa, we say they negatively covary, and their covariance is negative. If the high and low values do not pattern together in any way, covariance is zero.

The covariance between trait values and selection coefficients quantifies the change in the average trait value that is due to selection. This makes intuitive sense: if higher trait values are associated with higher selection coefficients, there will be a greater population share for packages with higher trait values in the next generation. So long as positive covariance is maintained, average trait values will increase over time. On the other hand, if *lower* trait values are associated with higher selection coefficients, average trait values will decrease over time. The covariance formulation is possible because z is assumed to be numeric.⁵

The Price equation also enables more nuanced descriptions of the action of selection. When the average trait value changes due to an association with fitness, we say there has been directional selection. However, there can be selection even in cases where the trait value average itself does not change at all. Stabilising selection occurs when a specific trait value is favoured. Suppose that the current average trait value is optimal. Selection might act to prune away values both higher and lower than the optimal, thus reducing the variance of the distribution while keeping the average the same.⁶ Capturing stabilising selection with the Price equation is as simple as defining a trait value $z^* = (z - \bar{z})^2$. When defined this way, \bar{z}^* is the variance of the distribution. Applying the Price equation to z^* tells you how \bar{z}^* changes, thus telling you how the variance of z changes: when selection causes the variance to decrease, $\text{cov}_p(W, Z^*)$ is negative, indicating stabilising selection. The average and the variance are called *moments* of the trait distribution, and higher moments can be captured by defining further traits $(z - \bar{z})^3$, $(z - \bar{z})^4$ and so on. Applying the Price equation to each of these constructed traits reveals how selection changes the moments of the distribution (Rice 2004, 178). And since the moments of a distribution collectively define its overall shape, the set of Price equations describing changes in moments collectively describe the overall change in shape of the distribution. All of this is possible when z is numeric, and so the moments of the distribution are well-defined.⁷

So far, so familiar. The rest of the paper concerns a constraint on the Price equation that has remained virtually unacknowledged.⁸ Covariance is a relationship between numeric variables. But there can presumably be selection on traits that are best represented as categorical variables. Indeed, the very notion of change in the average value of a trait does not apply when the trait

5. An overlooked issue arises in the derivation of the covariance term, which requires substituting p'_i for $p_i w_i$. It is possible that p_i and p'_i are well-defined while w_i is not. For example, if $p_i = 0$ and $p'_i > 0$, the substitution is problematic. Intuitively this combination of values represents migration into the population, which I discuss further in section 4.3. Halting the derivation before reaching the covariance term enables migration to be represented.

6. In this simplified example, directional and stabilising selection are mutually exclusive. Depending on the definition used, directional selection may be compatible with stabilising selection (Rice 2004, 176).

7. In the rest of the paper I will continue to speak of the Price equation describing the change in the average value of a trait. This tacitly includes all the moments of a distribution, because the moments are defined as averages of the constructed traits $(z - \bar{z})^2$, $(z - \bar{z})^3$, and so on (Rice 2004, 176–8).

8. Precursors include Jablonka and Lamb (2020, 69), who call for an “appropriately extended Price equation, reframed in informational terms,” and Bourrat (2024, 14), who suggests extending the work of Frank (2012a) on the relationship between the Price equation and information theory. My primary motivation is also understanding the informational perspective on selection; the categorical form of the Price equation described herein is a preliminary step toward that goal.

in question is categorical, yet it seems that these traits can still be selected.⁹ An example will illustrate this problem. Suppose a child's favourite colour is red, and they are choosing gifts from a shop's collection of balls. There are red, green and blue balls in the shop bin, and the child chooses all the red balls. By analogy with the apple case, there has clearly been selection for colour. Yet there is no natural way to assign the different colours to numbers such that we should say there has been selection for an 'increase in average trait value'. That is because an equally acceptable assignment would yield the result that there has been a *decrease* in average trait value, or no change at all. For example, if we assign red = 0, green = 1, blue = 2, then there has been a decrease in average trait value; if on the other hand we assign red = 1, green = 0, blue = 2, then there has been no change (assuming equally many starting balls); finally, if red = 2, green = 1, blue = 0, then there has been an increase in average trait value. How, then, might Price's framework support a truly general definition of selection?

As mentioned in the introduction, there is a workaround to this problem (discussed by Okasha 2006, 24). Instead of a single categorical trait 'colour' with values red, green, and blue, we can define three different binary traits 'red', 'green', and 'blue', each of which can be either 1 or 0 (in statistics this is known as dummy coding). A ball gets the value 1 for the red trait if it is red, and 0 for the other traits. Then three different Price equations can be applied, each of which describes selection on one of these dummy traits. If the child chooses the red balls, the three Price equations will reveal that there has been selection in favour of being red, selection against being green, and selection against being blue. This seems to be exactly the result we want, albeit using three equations instead of one. However, the equation I present in section 3 is, I believe, a more principled solution to the problem of representing selection on categorical traits. It is more principled because it does not treat a single trait as comprising separate dummy traits. It therefore requires a single equation rather than multiple equations, regardless of how many trait values there are. I compare my solution to the dummy coding approach in section 5.2 – they are in fact quite closely related.

In order to adequately capture selection on categorical traits in a formal framework, we have to change our interpretation of just one of Price's terms. The next section shows how.

3. A Categorical Price Equation

3.1. *Capturing Selection: Distributions, Not Averages*

I propose that the Price equation can be used to capture selection on categorical traits if we change our interpretation of one of its central terms. We will end up with a description of the *change in the distribution of types* in a population due to selection, as opposed to the change in average trait value. As with change in average trait value, change in the distribution of types can be segregated into change due to selection and change due to other factors. The goal is to

9. Several commentators on this paper have independently suggested that what is most significant about the Price equation is its quantification of selection in terms of covariance. The decomposition into two terms is for them of less interest – a stepping stone to the genuinely impressive result. I was somewhat surprised by this view as it so directly contrasts with my own: I believe the focus on covariance has (at least partially) obscured the fact that selection can operate on traits that are not numeric, while the distinction between change due to selection and change due to transmission remains a fundamentally interesting formal achievement. Because the manuscript was largely complete before the breadth and depth of the opposing view became apparent to me, I have not dedicated space here to defending my own perspective. Instead I will demonstrate how to derive a categorical version of what I still think is worth calling the (or a) 'Price equation', even though it does not contain a covariance term. I hope that the reader will recognise value in this representation, and that through its derivation and the subsequent discussion more light will be shed on the relative value of these two aspects of the Price equation.

represent this change, and this decomposition, in some manner that does not require the *average* value to change in order that there be selection.

The good news is that we can still use key features of the Price equation to do this. To see how, first let each z_i be the vector whose values are all zero except for the i th entry, whose value is 1. This is called a one-hot vector, a common way of representing categorical variables. For example, the red, green and blue balls in the shop would be represented by the vectors $z_1 = \langle 1, 0, 0 \rangle$, $z_2 = \langle 0, 1, 0 \rangle$ and $z_3 = \langle 0, 0, 1 \rangle$. The assignment of colours to vector entries is arbitrary, but none of the calculations we will carry out are affected by these assignments. In effect, the vector representation condenses multiple dummy-coding variables into a single vector variable. This enables us to derive a single Price equation for the whole population.

Now the distribution of balls in the shop bin can also be described by a vector. If they are initially evenly distributed, the population distribution is $Z = \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$. This can be thought of as describing the probabilities of choosing each package when picking at random from the population. And if we let the individual p_i be the individual population proportions as before, the overall distribution is equal to the sum $\sum pz$:

$$\begin{aligned}\sum pz &= \frac{1}{3} \langle 1, 0, 0 \rangle + \frac{1}{3} \langle 0, 1, 0 \rangle + \frac{1}{3} \langle 0, 0, 1 \rangle \\ &= \left\langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right\rangle = Z\end{aligned}$$

In short: when z is a numeric variable, $\sum pz$ is a population average \bar{z} ; when z is a categorical variable (i.e. a one-hot vector), $\sum pz$ is a population distribution Z . In both cases we are dealing with a weighted sum, it's just that weighted sums are averages for numeric values z and distributions for categorical values z . Since the Price equation expresses relationships among a collection of weighted sums it looks as though it can be written exactly as before. The only difference is it no longer describes the change in average trait value $\Delta\bar{z}$ but the change in population distribution ΔZ .

In order to write down the Price equation using vectors to represent categorical types, we need to define the remaining two terms. Fortunately, they are as intuitive as the original definitions. Let each z'_i be the vector whose values are all zero except for entries descended from the i th parent, whose values are the proportions descended from that parent. For example, a red ball typically does not change colour of its own accord, so $z'_1 = \langle 1, 0, 0 \rangle$ which represents the fact that every red ball stays red after selection regardless of whether it was picked (figure 3). On the other hand, if half the red balls spontaneously turn green, the vector would be $z'_1 = \langle \frac{1}{2}, \frac{1}{2}, 0 \rangle$ (figure 4). Let p'_i be the population share of descendants of i . If only red balls are chosen, then $p'_1 = 1$. This is true whether or not they change colour. (As before, p'_i ignores information about trait values, and z'_i ignores information about population share.)

Now $p'_i z'_i$ is a vector describing the proportions of descendants of package i , weighted by their population share in the subsequent population. And $\sum_i p'_i z'_i$ is the vector describing the new population: essentially, it sums all the non-green balls that turned green together with the green balls that stayed green in order to get a total population share for green balls, and it performs a corresponding sum for red and blue balls. The result is the population distribution after selection, Z' .

3.2. A Price Equation for Vectors

Having defined the key terms I'll drop the subscripts for ease of reading. We just saw that $\sum pz$ is a vector describing the distribution of types in the population before selection and $\sum p'z'$ is

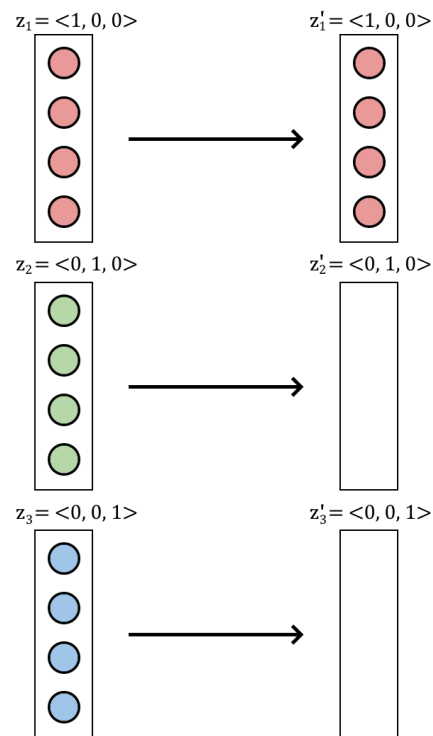


Figure 3: **Selection on a categorical trait.** A shop contains equal numbers of red, green and blue balls. A child selects only the red balls. There has been selection on colour, which can be represented by defining each trait value z_i as a one-hot vector.

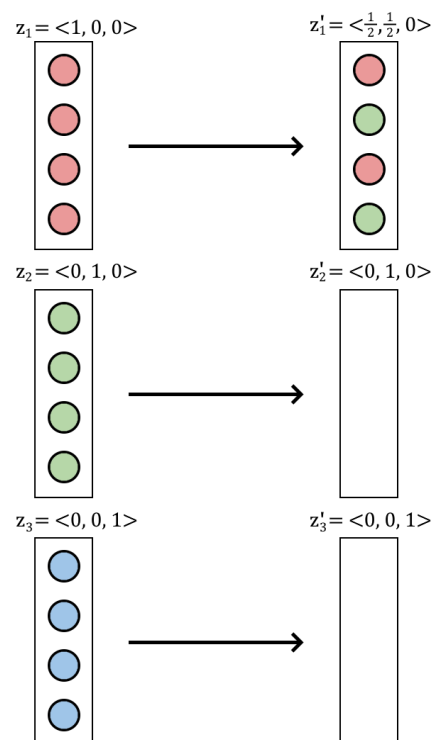


Figure 4: **Selection on a categorical trait with mutation.** See the main text for a calculation of the two terms of the categorical Price equation in this case.

the equivalent vector after selection. Their difference is a vector describing the total change in population proportions. This just is the left hand side of the Price equation, except that here the weighted sums are distribution vectors rather than averages.

Does the Price derivation still work? Yes: assuming $\sum p'z$ is well-defined, the derivation is perfectly valid. Do the derived terms still capture our pretheoretic distinction between change due to selection and change due to other factors? Consider the selection term on a categorical interpretation:

$$\Delta_s Z = \sum p'z - \sum pz \quad (4)$$

Because each z is a one-hot vector indexing its package, this describes how much the population share of each package changes. When the child chooses the red balls:

$$\begin{aligned} \Delta_s Z &= \sum p'z - \sum pz \\ &= 1 \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &\quad - \left(\frac{1}{3} \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{3} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{3} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \end{aligned}$$

The resulting vector represents the fact that the red balls have increased their population share by $\frac{2}{3}$, while the other two colours have each lost $\frac{1}{3}$. These gains and losses happened purely due to selection (in this case the child's choice of balls).

The transmission term works too:

$$\Delta_t Z = \sum p'z' - \sum p'z \quad (5)$$

This says: imagine each package had reproduced perfectly ($\sum p'z$), and look at how different that is from what each of them actually did ($\sum p'z'$). If half the red balls spontaneously turn green (and the other balls retain their colours), this sum becomes:

$$\begin{aligned} \Delta_t Z &= \sum p'z' - \sum p'z \\ &= 1 \times \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &\quad - \left(1 \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} \end{aligned}$$

The resulting vector represents the fact that red balls have lost a population share of $\frac{1}{2}$, and green balls have gained an equivalent share, due to sources of change other than selection (see again figure 4). The total change in this case is $\left[\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}\right] + \left[-\frac{1}{2}, \frac{1}{2}, 0\right] = \left[\frac{1}{6}, \frac{1}{6}, -\frac{1}{3}\right]$. Indeed, both red and green balls have increased their population share by $\frac{1}{6}$ (they initially had four out of twelve, or one third of the population, and now have two out of four, or one half) and blue balls have decreased their population share by one third (they initially had four out of twelve and now have zero). The categorical Price equation decomposition is depicted in figure 5.

Because the categorical equation makes no assumptions beyond splitting the population into different types, it can capture changes in distribution even when applied to numeric traits. As long as the population can be segregated into discrete types, the categorical equation can be applied. Discretization can be achieved by considering ranges of numeric variables; for example, the height of an individual can be taken to fall into ranges of 1–1.1 metres, 1.1–1.2 metres, and so on. Then any changes in the distribution of the population with respect to these discrete categories will be represented by the categorical equation, regardless whether or not the overall population average height changes. The question of which ranges to use is a modelling decision that depends on the purposes to which the equation is being put in a particular case. The take-home message is that the categorical equation is broad enough to apply to both categorical traits and appropriately discretized numeric traits (the difference between categorical and numeric traits is sometimes captured in terms of ‘levels of measurement’; see section 5.3 for further discussion).

Vectors have been employed in the Price equation before. Lande and Arnold (1983) explored correlated selection between traits using a vector where each entry corresponds to a single numeric trait. Their goal was to partition selective forces acting on a trait into direct effects (where selection favours a trait because of some phenotypic effect it has) and indirect effects (where selection favours a trait because it is correlated with another trait upon which selection is having a direct effect). The resulting statistical apparatus has a further application: to capture complex multi-trait fitness effects. For example, Brodie (1992) employed it to discover a correlation between the colour patterns and anti-predator behaviour of garter snakes. This is not due to an accidental genetic correlation, but because the fitness of a garter snake’s phenotype is a complex combination of both traits. It appears that non-striped snakes ought to perform an evasive behaviour called a ‘reversal’ while striped snakes should avoid doing so; the fitness function of each trait is modulated by the trait value of the other.

Frank (2012b) and Frank and Godsoe (2020) also employ vectors as part of their exposition of the Price equation, but in that case the vector entries are different values of a particular trait (corresponding to different individuals in the population) rather than different traits. There might be connections with my one-hot approach; Frank often describes the vector z as a “coordinate system” that evolves in tandem with population proportions. I will have to leave a more detailed comparison for another time.

I believe the foregoing constitutes a *prima facie* case for the usefulness of the categorical form of the Price equation. Objections are discussed in section 5; before then, section 4 describes further properties of the equation, in particular its application to multi-level selection.

4. Properties of the Categorical Price Equation

4.1. Multi-level Dynamics: Expanding the Selection Term

So far I’ve assumed packages have multiple individuals within them, all of which are the same type. We can relax this assumption to start to understand how multi-level selection can be

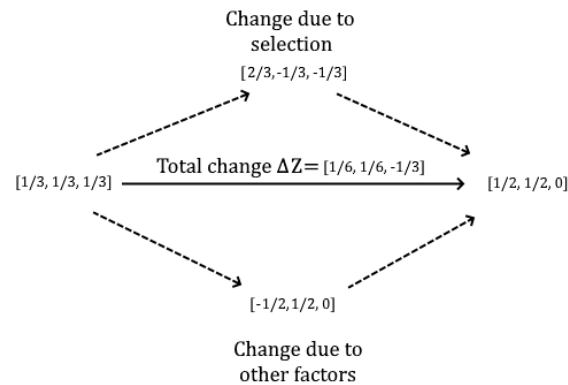


Figure 5: The categorical Price equation partitions change in population distribution ΔZ into two parts: change due to selection, and everything else.

represented in the framework. Multi-level selection concerns processes operating at hierarchical levels, with a population of entities at a lower level belonging to discrete entities at a higher level. Following Okasha (2006), we'll call the lower-level entities *particles* and the higher-level entities *collectives*. This will entail two different systems of packaging, with lower-level packages collected into higher-level packages (figure 6). For the formal results below to hold, each particle must belong to exactly one collective, hence each particle package must be nested within exactly one collective package. Additionally the descendant package of a particle package must belong to the corresponding descendant package of the collective package to which it belongs (as in figure 6).

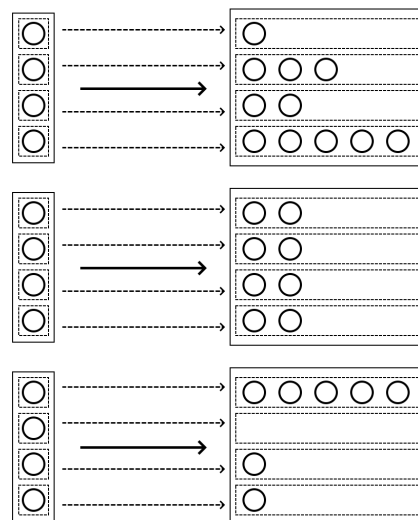


Figure 6: Price's framework applied to multi-level selection. Circles are particles. Dashed boxes are particle packages. Solid boxes are collective packages. As in the original framework, each initial package corresponds to exactly one subsequent package after selection (shown by dashed arrows for particles and solid arrows for collectives). In this example, there has been selection within collectives 1 and 3 (different particles have different numbers of descendants) but not collective 2 (all particles have exactly two descendants). There has also been selection between collectives, because each initial collective has the same population share while the subsequent collectives have different population shares. The framework requires that the population share of a collective be defined in terms of the number of particles it contains; see the main text for discussion.

| Global index i | Local index kj |
|------------------|------------------|
| 1 | 11 |
| 2 | 12 |
| 3 | 13 |
| 4 | 21 |
| 5 | 22 |
| 6 | 23 |
| 7 | 31 |
| 8 | 32 |

Table 1: A population of eight particles has two indexing systems. The particles are divided into collectives, with collectives 1 and 2 possessing three particles each and collective 3 possessing two particles. The global index is the same as the normal package index, with one particle per package. The local index specifies the collective k to which the particle belongs, followed by its unique index within that collective j .

Building on previous work, Okasha (2006, §2) discusses two concepts of multi-level selection, corresponding to two kinds of fitness that can be attributed to collectives. In the first concept, collective fitness is simply the average fitness of the individual particles within the collective. Since particle fitness is defined in terms of number of offspring particles, this first concept of collective fitness is a measure of the number of offspring *particles* produced by a collective. In the second concept, collective fitness is the number of offspring *collectives* produced by a collective. For the second concept to be distinct from the first, it must be possible to count offspring collectives without weighting them by size (Okasha 2006, 54). However, as I am employing the Price framework, the one-to-one package assignment implies that the number of offspring collectives of any collective is fixed at 1. The fitness of a collective is determined by its present and future population share rather than its absolute offspring, and so the second concept of collective fitness cannot be applied without it collapsing to the first. Future work could investigate whether the framework can be expanded to recapture the distinction.

Turning to the question of multi-level selection, recall that the framework treats both particles and collectives as packages. By using two different packaging systems we find that the selection term (equation (4)) of the categorical Price equation can be decomposed into parts that describe *selection between collectives* and *selection within collectives*. This corresponds to the decomposition discussed by Okasha (2006, §2.3.1), but does not require invoking numeric terms like covariance or expectation. Let's see how it works.

First we'll index particles across the whole population with i . Recall that z_i is a one-hot vector picking out the category particle i belongs to, and p_i is the particle's population share. Now divide all the particles into k collectives. Within each collective, particles are indexed with the letter j . The collectives don't need to be evenly sized. Each particle gets its unique global index i , and its unique collective index kj . Table 1 shows two sets of indices for a population of eight particles arranged into three unevenly sized collectives.

Now we can define trait values and population shares according to the different indexing schemes. First, since every particle has its one-hot vector describing its categorical trait value, this is the same in both indexing schemes: $z_i = z_{kj}$. But the population share is not the same as the particle's collective share. p_i says how much of the population is composed of this particle, but p_{kj} only says how much of the local collective it makes up. I might be a small fish relative to the ocean, but a large fish relative to my local pond. A particle's global population share and its local collective share can be related via the collective's own population share: how much of the population the entire collective takes up (how big the pond is relative to the ocean). We'll

denote this by P_k . Then a particle's population share is a product of its collective share and the collective's population share. In other words, how well-represented I am in the population is equal to how well-represented I am in my collective multiplied by how well-represented my collective is in the population:

$$p_i = p_{kj}P_k$$

One more definition and then we can derive the multi-level result. A collective's trait value, Z_k , is simply the vector describing the proportion of types that make it up. If it contains only one type of particle it will be one-hot. But it might contain multiple types. Because the particles' shares in the collective are constrained to sum to 1, the collective's trait vector will be a list of numbers that sum to 1. Each entry in the vector describes what proportion of particles of that category are in the collective. The trait vector can therefore be defined as $Z_k = \sum_j p_{kj}z_{kj}$.

Once these definitions are on the table, how can we distinguish selection within a collective from selection between collectives? Noting that the selection term for particles in the whole population, $\sum_i p'_i z_i - \sum_i p_i z_i$, can be written more concisely as $\sum_i z_i (p'_i - p_i)$, here is an interesting identity whose derivation can be found in the appendix:

$$\begin{aligned} \Delta_s Z &= \sum_i z_i (p'_i - p_i) \\ &= \underbrace{\sum_k Z_k (P'_k - P_k)}_{\text{Selection between collectives}} + \underbrace{\sum_k P'_k \sum_j z_{kj} (p'_{kj} - p_{kj})}_{\text{Weighted selection within each collective}} \end{aligned} \quad (6)$$

The easiest way to understand this equation is to think about the values of the component terms in a few different situations. First, imagine the collectives are all homogeneous, each containing exactly one type of particle (figure 7). Since all particles within a collective will have exactly the same number of offspring, they will all have exactly the same collective share after selection. And since they are the only particle type in that collective, this will be equal to their original collective share. (There are three red balls and each splits into two; each originally had a collective share of $\frac{1}{3}$ and each of their descendants has a collective share of $\frac{2}{6} = \frac{1}{3}$.) Therefore $p'_{kj} = p_{kj}$ for every particle in every collective, and the second term is zero. This makes perfect sense: when collectives are homogeneous, the only selection that can take place is between collectives, and so $\sum_i z_i (p'_i - p_i) = \sum_k Z_k (P'_k - P_k)$.

Now imagine that collectives all contain within them a microcosm of the whole population. Say for example the population has three red, three green and three blue balls, and there are three collectives with one of each type of ball (figure 8). There can be no selection between collectives because they are all identical to each other. No matter the fitnesses of the particles within, each collective's population share after selection P'_k must be identical to its population share before selection P_k . The first term is therefore zero, and all the selection takes place within each collective. In general there can be a mixture of selection effects between and within collectives (as in figure 6); then both terms are non-zero.

These results have previously been derived for the numeric Price equation. They are exciting regardless of whether the trait in question is numeric or categorical, because they display a kind of self-similarity. The term describing selection between collectives has the same form as the term describing selection across the entire population. We can therefore gather collectives together into super-collectives, and further split equation (6) into three terms: one describing

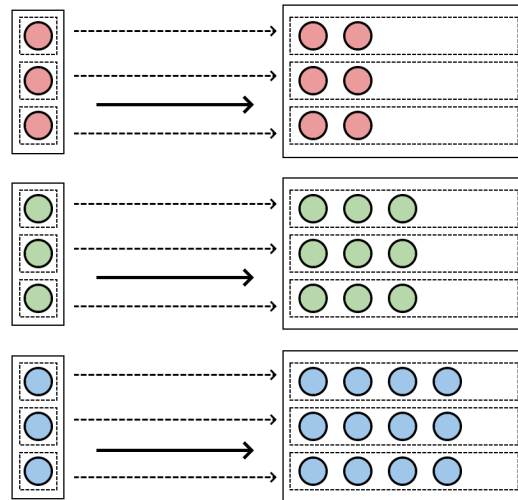


Figure 7: **Multi-level selection with homogeneous collectives.** There is between-collective selection because different collectives have the same initial population share but different subsequent population shares. The first component of equation (6) is therefore a vector with some nonzero entries. There is no within-collective selection since every particle within a given collective has the same number of offspring. The second component of equation (6) is therefore a vector whose entries are all zero.

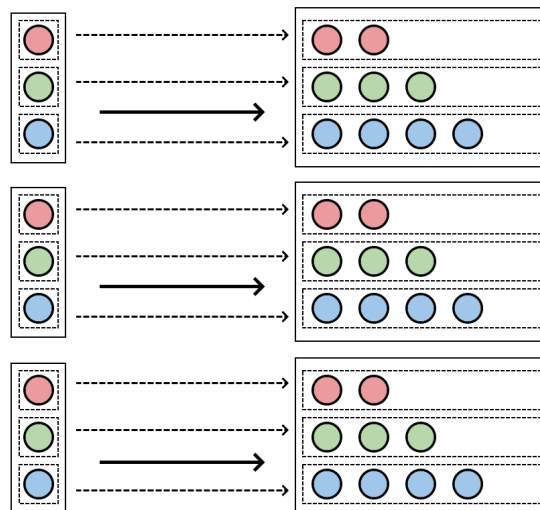


Figure 8: **Multi-level selection where each collective is a microcosm of the population.** There is no selection between collectives because each collective has equal population share before and after selection. The first component of equation (6) is therefore a vector whose entries are all zero. There is selection within each collective because the different particles' descendants have different collective share within each descendant collective. The second component of equation (6) is therefore a vector with some nonzero entries.

selection between super-collectives, one describing selection among first-order collectives within super-collectives, and one describing selection among particles within a collective. As has been noted for the numeric Price equation, there is no principled restriction on the number of levels that can be represented.

4.2. Multi-level Dynamics: Decomposing the Transmission Term

Consider the Price equation at the collective level (omitting subscripts k):

$$\underbrace{\sum P'Z' - \sum PZ}_{\text{Total change at collective level}} = \underbrace{\sum Z(P' - P)}_{\text{Selection at collective level}} + \underbrace{\sum P'(Z' - Z)}_{\text{Transmission bias at collective level}} \quad (7)$$

The transmission bias is the average across subsequent collectives, $\sum_k P'_k$, of how the aggregate trait changed within each collective, $Z'_k - Z_k$. But how the aggregate trait changed within each collective is just the Price equation for particles within collectives:

$$\begin{aligned} Z'_k - Z_k &= \sum_j p'_{kj} z'_{kj} - \sum_j p_{kj} z_{kj} \\ &= \underbrace{\sum_j z_{kj} (p'_{kj} - p_{kj})}_{\text{Particle selection within collective } k} + \underbrace{\sum_j p'_{kj} (z'_{kj} - z_{kj})}_{\text{Particle transmission bias within collective } k} \end{aligned}$$

(The equivalent decomposition for the numeric Price equation has been noted by Frank (1998, 15) and Bourrat (2021, 18) among others, and is alluded to by Okasha (2006, 70).) What this means is that a hierarchy of self-similar equations appears, describing non-selective sources of change at one level as a combination of selection and non-selective sources at the next level down. This self-similarity continues until the trait value can no longer be appropriately described as an aggregate of the trait values at a lower level. While a collection of coloured balls can be described as having some proportion of red, green and blue balls, it's not clear that a single red ball can reasonably be described as having some proportion of coloured ball-parts. (Of course, if there were some process by which portions of balls could change colour, it may well be appropriate to represent them this way, and the decomposition would extend one level further.)

In sum, both the selection term and the transmission term display a kind of self-similarity. The difference between them has to do with the direction in the hierarchy in which groupings are considered. Starting from a certain hierarchical level, the selection term is self-similar when we consider groupings above this level (by collecting packages into collectives, super-collectives, and so on). The transmission term is self-similar when we consider groupings below this level (by looking at each package and treating it as if it were its own population, with its own mini-packages, subject to its own Price equation).

4.3. Migration

Because the categorical Price equation does not contain w_i , it is possible to consider packages for which $p_i = 0$. Of course, if both $p_i = 0$ and $p'_i = 0$ then the component of the equation for package i is 0, representing a situation in which the package is not part of the population at all. Since a modeller would simply not include such a package in their calculations, we can ignore this case. A more interesting situation occurs when $p_i = 0$ but $p'_i > 0$. This represents package i migrating into the population: initially it was absent (had zero population share), subsequently it is present and has nonzero population share.

One might think a problem would arise when assigning a value to z_i here. Supposing that the migratory package in fact had a different trait value before it migrated versus after it arrived, an accurate representation of the facts would attribute different values to z_i and z'_i . And this would seem to entail overcounting change: the initial population does not contain the migratory package at all, so it shouldn't matter for the calculation of populational change what z_i happened to be. Happily, the fact that the original trait value of the migratory package shouldn't matter is reflected in the mathematics. Here's what happens to component i of the categorical Price equation when we set $p_i = 0$:

$$\begin{aligned}(\Delta Z)_i &= z_i (p'_i - p_i) + p'_i (z'_i - z_i) \\&= p'_i z_i + p'_i z'_i - p'_i z_i \\&= p'_i z'_i\end{aligned}$$

The change to the population distribution Z caused by a migratory package falls squarely within the transmission bias term. This is further motivation to interpret the term as 'other sources of change'; transmission bias being just one kind of non-selective change.¹⁰

5. Problems, Objections, Omissions

5.1. Are There Categorical Traits?

I've used colour as an example of a categorical trait, but one might wonder whether this is a good idea, given that particular colours can be represented as a numeric combination of some set of basis colours. For example, the red, green and blue shades I've been using can be assigned numbers from the traditional RGB colour model. The red colour in my figures has an RGB value of (234, 153, 153), the green is (182, 215, 168), and the blue is (159, 197, 232). These colours could be represented using other 'basis' features too, with different numeric combinations placing each in a region of the resultant space. Doesn't this suggest that I am wrong to complain about the lack of a categorical Price equation, if these traits are tacitly numeric after all?

To respond, perhaps colour is not the best example of a categorical trait. A reviewer suggested flavour instead. Suppose the child is choosing between flavours of ice cream: vanilla, strawberry and chocolate (we may assume Neapolitan is unavailable). The change in flavour distribution from the vendor's trays to the child's cone reflects a selection process that implements a flavour preference. The vector representation applies as before, and in this case there is no familiar decomposition of flavours into a numeric flavour space.

10. Kerr and Godfrey-Smith (2009) present a different generalization of the Price equation that also allows migration to be represented. See section 5.4 for details of their approach and a potential extension of the categorical framework in light of their results. It should also be noted that while the numeric Price equation can similarly represent migration (by halting the derivation before substituting the covariance term), Price (1995) did not mention migration in his wide-ranging discussion of the equation and its applications. I thus raise the issue here not to claim a novel result, but to emphasise that the framework's scope of application is broader than Price himself perhaps realised.

However, there is a further worry that *any* putative categorical trait could be represented in a numeric space of high enough dimension. Techniques developed in artificial intelligence to deal with categorical data seem to reveal that any feature can in principle be placed into a numeric space. For example, the semantic values of words can be extracted from huge text corpora by placing each word into a position in a high-dimensional ‘semantic space’. Nearby words have similar semantic contents, while operations on numeric vectors sometimes yield surprisingly consistent relationships, such as ‘king - man + woman = queen’. Perhaps even flavours could be represented this way, if the appropriate bases were found for a space of sufficiently high dimension. If seemingly categorical traits can be placed into numeric spaces, this undercuts the motivation for seeking a categorical Price equation in the first place.

To respond to this further worry, it is of course true that some traits which at first seem categorical can be represented in high-dimensional numeric spaces. Let’s concede, for the sake of argument, that any categorical trait can be represented within such a space. Then one could apply a version of the Price equation with high-dimensional numeric vectors, rather than (or in addition to) the categorical one-hot vectors I’ve been advocating. As has already been shown, the equation is perfectly well-defined in this case: its derivation does not make any assumptions about what form z takes, only that it obey a few basic algebraic rules. Just as z can be a number or a one-hot vector, it can be a vector of any kind. What happens in this case is that $\sum pz$ is still a vector, but each of its entries is the average value of the entries for each of the individuals in the population. In other words, for a population of red, green and blue balls represented within a three-dimensional RGB space, the population vector will define a point in that space which takes the average value of the R component (continuing the example, this would be 192), the average value of the G component (188), and the average value of the B component (184). The result is a rather uninteresting light grey. Supposing that after selection we are left with red balls only, the average RGB value of the subsequent population is just the red balls’ RGB value. The Price equation tells us that we have moved from a light grey to a red. This information could well be useful depending on the purposes we want to put the Price equation to; similarly, we might get use from this multidimensional numeric Price equation with respect to any categorical trait once it is represented in high dimensional space.

The question is whether this is the only correct or useful application of the Price equation. If the objection is that categorical traits *can* be represented this way, it doesn’t harm my main point, because categorical traits can also be represented as one-hot vectors. The question the theorist should ask is not “which is the uniquely correct representation?”, because both are available; instead, the question is “which representation is most useful for my purposes?” My contention is that we can always apply the categorical version of the Price equation – should we have reason to – in order to see how population distributions change. We can apply it to seemingly categorical traits regardless of whether they have a high-dimensional numeric representation, and we can apply it to traits that are obviously numeric, by ignoring magnitude relationships between numbers and discretizing the population into unique categories (as described in section 3.2). We can always ask about a change in distribution rather than a change in average numeric value. All of these points remain true regardless of whether the premise of the objection holds; that is, whether or not it turns out that every seemingly categorical trait can be represented in a high-dimensional numeric space.

5.2. *Are One-Hot Vectors Really Better than Dummy Coding?*

By stacking 1s and 0s into vectors, one might object, I have done nothing more than repackage the dummy coding approach. Recall that dummy coding treats each categorical variable as a

collection of binary variables. Instead of the variable 'colour' with values red, green and blue, we have three variables 'red', 'green' and 'blue', each with values 1 and 0. Each individual has a value of either 1 or 0 for each of the three dummy variables depending on its colour. Rewriting the example of red balls being selected using dummy coding reveals three Price equations. I argued that the vector approach is preferable because it captures the changes in population distribution in a single equation. But it turns out that the vector equation simply stacks the three dummy equations on top of each other. The vector components correspond to the trait values red, green and blue, and the numbers that appear in each of these slots are exactly the numbers that appear in the corresponding dummy equations for the 'red', 'green' and 'blue' dummy variables.

Furthermore, as a reviewer suggested, the dummy coding approach captures the fact that the covariance between a trait value and fitness exactly measures the change due to selection of that trait in the population. One might think it is the equality between the covariance term and the change due to selection that is most notable about Price's discovery, rather than the algebraic rearrangement of terms (see also footnote 9 above). Since the categorical equation is obtained by merely rearranging terms, and since it does not capture the correspondence between covariance with fitness and change due to selection, it seems as though the equation loses what is most insightful about the Price equation. Given these facts, by what right do I claim the vector approach as superior, when it is little more than a repackaging of the dummy coding approach?

I am willing to concede that the vector representation of categorical variables is a relatively slight change from the dummy coding approach. Yet it is an extension that builds positively upon existing work. For one thing, it is aesthetically pleasing to employ a single equation with N-dimensional vectors rather than N equations. The population of balls does not, as a matter of fact, harbour three overlapping traits: it has a single trait with three values. The dummy coding approach misses that fact, while the vector approach captures it explicitly. As a result, the dummy coding approach doesn't allow explicit representation of the constraint that individuals can only have one value of a particular categorical variable. Using one-hot vectors to represent individuals embodies this constraint.

Another reason to prefer the vector representation is that it allows us to access geometric properties of populations. In the numeric multivariate case discussed earlier in this section, representing traits as vectors would allow us to measure the distance between individuals in the resulting multidimensional space. Various measures can be employed in multidimensional settings to describe the distance between two composite colours in RGB space, for example the Euclidean distance. Distinguishing the individual dimensions – decomposing the vector into component parts – misses the structure of the overall space. These distances may or may not be theoretically important for future applications, but I contend we should at least keep them in play rather than neglecting them without cause. Since categorical traits are an extreme case of multidimensional representation (one in which each individual lies exclusively on an axis of the space, hence is picked out by a one-hot vector), for mathematical consistency it is appropriate to use vectors when representing them.

Finally, emphasising that selection changes population distributions, not just trait value averages, prompts us to think more broadly about the Price equation's application. The fact that the interesting properties described in section 4 still hold should be sufficient to warrant further interest. While the categorical equation loses the relationship with covariance, fitness is still what is causing trait distribution changes due to selection. It is just that, in my presentation, covariance has been dethroned as the ultimate expression of that change. To those who take Price's key insight to be the covariance term, it may seem that something has been lost in my formulation. I think that what has been lost needed to be lost in order to capture the general case. Selection does not just change averages, it changes distributions.

| | | Level of measurement | | | |
|----------|-----------------|----------------------|------------|------------|------------|
| | | Categorical | Ordinal | Interval | Ratio |
| Equation | Categorical | Yes | Yes | Yes | Yes |
| | <i>Ordinal?</i> | <i>No</i> | <i>Yes</i> | <i>Yes</i> | <i>Yes</i> |
| | Numeric | No | No | Yes | Yes |
| | <i>Ratio?</i> | <i>No</i> | <i>No</i> | <i>No</i> | <i>Yes</i> |

Table 2: Levels of measurement to which the categorical and numeric versions of the Price equation apply. It is an open question whether there are further versions of the equation, one that applies at the ordinal level and below, and one that applies solely at the ratio level.

5.3. Levels of Measurement

A reviewer suggested applying this treatment to different levels of measurement. ‘Levels of measurement’ here refers to a four-way distinction between types of variable: categorical (also called nominal), ordinal, interval, and ratio. Categorical variables distinguish categories into which a value can fall, but do not have any numerical meaning. Ordinal variables define a ranking scale on which values can be higher or lower, but do not assume consistent distances between the rankings. So for example the finishing positions in a race can be represented by an ordinal variable: 1st place finished before 2nd and 2nd finished before 3rd, but the time discrepancies between the finishers might be radically different and these discrepancies are not captured by the variable. Interval variables do capture differences between their values. Temperature measured in degrees Celsius is an interval variable because the difference between 20 degrees and 30 degrees is in some sense the same as the difference between 30 degrees and 40 degrees. However, interval variables do not enable comparisons in terms of ratios: it does not make sense to say 40 degrees is twice as hot as 20 degrees, because the location of zero is arbitrary on this scale. Ratio variables are those for which both numeric differences and ratios are meaningful. Height is a ratio variable because it makes sense to say that one person is twice as tall as another. What I have been calling ‘numeric’ subsumes the interval and ratio categories.

We now have versions of the Price equation describing change in population properties relating to two out of these four levels (table 2). The numeric Price equation describes changes in population average. Because both interval and ratio variables have meaningful averages, the original Price equation applies to any variable of those two kinds. The version of the Price equation introduced in this paper describes changes in population distributions with respect to categorical variables. It applies to any categorical variable that can be defined over a population. Because each level of measurement has the same properties as those of the levels above, each Price equation applies to variables at its level and all levels below. So the numeric equation applies at the interval level and below (encompassing the ratio level), and the categorical equation applies at the categorical level and below (encompassing the ordinal, interval, and ratio levels).

Are there versions of the Price equation that apply at the ordinal level and the ratio level – do the italicised rows of table 2 correspond to genuine options? For the answer to be yes, there must be a variable z of the relevant kind (i.e. ordinal; ratio) for which a legitimate population-level property $Z = \sum pz$ can be defined. To my mind there is no particular reason why such traits must exist. The levels of measurement are a convenient categorisation for statisticians who need to distinguish variable types for the purposes of applying appropriate statistical tests. There is no guarantee that this four-way distinction will deliver four unique types of trait z that each define

a meaningful population property $Z = \Sigma pz$. For ordinal data in particular it seems unlikely that Σpz could be meaningful. Even if p could be defined, the average would tend to be a number somewhere in between the ordinal values. Such a number would be meaningless on an ordinal scale, because there is no guarantee that the differences between ordinal values are the same. Obtaining a value of 2.5 could entail something wildly different about where between 2 and 3 this value lies, compared to a value of 3.5 and what that entails about where between 3 and 4 the value lies. For this reason statisticians working with ordinal data tend to employ the median rather than the mean. Sometimes a weighted median is employed. But the median of an ordinal dataset, whether weighted or not, cannot be written as Σpz . Changes in the median cannot be captured by a Price equation.¹¹

In sum, while there might be a role for other versions of the Price equation that apply at different levels of measurement, it seems to me unlikely that there will be a proprietary equation for the ordinal level. I leave the question of the ratio level to future work.

5.4. *Multi-parental Inheritance and Neighbour-Modulated Fitness*

There are of course a great deal more processes relevant to evolutionary theory than have been discussed here. I have omitted discussion of random drift entirely.¹² Another significant process is multi-parental inheritance. To represent multi-parental inheritance we must either avoid the assumption that each descendant particle is associated with exactly one ancestor, or treat individuals as comprised of multiple particles. The latter would allow us to decompose each individual into a collection of discrete particles, each of which derives from a unique ancestor particle. This might be what Price (1995, 393 fig. 5) had in mind when he applied his framework to genetic selection, and it might even be warranted in that case. However, it's not clear that such an approach will be sufficiently general if the goal is (as I take it to be) carving out a concept of selection that plays the same role in every possible evolutionary system. It seems as though there can be selection in cases where particles have multiple parents, and cannot obviously be broken down into particles on a lower level, all of which necessarily have a unique parent particle.

The question of multiparental inheritance has been tackled by Kerr and Godfrey-Smith (2009). They introduce formalism explicitly describing which particles in the initial and subsequent population are connected to each other. Representing connections allows for particles in the initial population that are not connected to any in the subsequent population (corresponding to extinction) and particles in the subsequent population that are not connected to any in the initial population (corresponding to migration). Because multiple connections are allowed, multiparental relationships can be represented. In this sense, Kerr & Godfrey-Smith's approach encompasses a broader set of cases than I have considered. However, they retain the traditional focus on numeric traits, such that a descendant inherits the average trait value of its parents. Thinking about how their connectionist approach could be employed to represent multiparental inheritance of categorical traits, clearly the key question is how exactly the trait can be said to be 'inherited' from multiple parents if it is supposed to be categorical. There are several possibilities here. In some systems the child might inherit a blend of the two traits, while in others it might inherit one trait probabilistically depending on the strength of the connection. These considerations may be especially relevant in cultural evolution settings, where any individual in a population can in principle inherit a cultural trait from any other individual with whom

11. Frank and Godsoe (2020, 2) suggest that the median can be written in this form, but they appear to be referring to either interval or ratio variables rather than ordinal.

12. The effects of drift can appear in either of the two terms of the Price equation, depending on the cause of drift in the case in question. Thanks to an anonymous reviewer for pointing this out.

it socially interacts. Future work should investigate prospects for representing categorical traits using a connection-based approach.

The second huge topic not yet addressed is neighbour-modulated fitness. No account of selection is going to be complete unless it takes into account the contextual effects of nearby particles on the future population share of a focal particle. The most obvious biological application of this phenomenon is social behaviour (Gardner, West, and Wild 2011), but even the simplest cases demonstrate neighbour-modulated fitness: whether an apple is chosen by the shopper may depend on whether it is the best apple in the shop, which in turn depends on the trait values of the other apples. Some of the questions that arise here, but by no means all, can be answered by reference to the multi-level approach (Gardner 2015). Future work should draw on existing approaches to understand how selection operates on categorical traits when fitness is affected by a particle's neighbours.

6. Recapitulation

George Price introduced a deceptively simple framework with deep significance for theoretical biology and beyond. Starting with basic definitions of package, trait value, and population share, Price derived an equation which describes how the population average value of a trait changes due to selection and other factors. The insight at the heart of this result is often taken to be the quantification of selection as the covariance between trait value and fitness. However, this step of the derivation requires that the trait value in question be numeric. I have argued that the Price equation need not be restricted to traits that can be represented numerically. A conceptually satisfactory version of the equation can be derived, making essentially the same assumptions, by treating the trait which is subject to change as categorical. This is done by using one-hot vectors to represent categorical traits, and treating weighted sums as population distributions rather than averages. Phenomena relevant to questions of populational change, such as multi-level selection and migration, can be represented in the resulting framework. There is a great deal of work left to do.

Appendix

Derivation of the Multi-level Selection Equation

This derivation is based on a suggestion from Wade (1985, 63). The selection term with particles indexed across the whole population is:

$$\Delta_s Z = \sum_i z_i (p'_i - p_i)$$

Instead of counting particles across the population, we want to segregate them into k collectives. Within each collective, particles are indexed by j . Since P_k is the population share of collective k , it follows that $p_i = p_{kj}P_k$. Similarly, $p'_i = p'_{kj}P'_k$. Trait values stay the same regardless of indexing, so $z_i = z_{kj}$. The selection term therefore becomes:

$$\Delta_s Z = \sum_k \sum_j z_{kj} (p'_{kj}P'_k - p_{kj}P_k)$$

From here onwards we will remove the subscripts for ease of reading. Every upper-case letter can be considered to have the subscript k , and every lower-case letter can be considered to have the subscript kj . We can rearrange terms to get the following:

$$\begin{aligned} \Delta_s Z &= \sum_k \sum_j zp'P' - \sum_k \sum_j zpP \\ &= \sum_k P' \sum_j zp' - \sum_k P \sum_j zp \end{aligned}$$

At this point we make a similar move as in the original Price equation derivation, by adding and subtracting identical terms:

$$\Delta_s Z = \sum_k P' \sum_j zp' - \sum_k P \sum_j zp + \sum_k P'Z - \sum_k PZ$$

We can restate some of these terms, switching between expressions pertaining to particles and expressions pertaining to collectives:

$$\Delta_s Z = \sum_k P' \sum_j zp' - \sum_k PZ + \sum_k P'Z - \sum_k P' \sum_j zp$$

Then, as in the derivation of the original Price equation, we put like terms together:

$$\begin{aligned} \Delta_s Z &= \left(\sum_k P'Z - \sum_k PZ \right) + \left(\sum_k P' \sum_j zp' - \sum_k P' \sum_j zp \right) \\ &= \sum_k (P'Z - PZ) + \sum_k P' \sum_j (zp' - zp) \end{aligned}$$

And the equation as reported in the text appears when we do some factoring (with the subscripts explicitly included):

$$\Delta_s Z = \underbrace{\sum_k Z_k (P'_k - P_k)}_{\text{Selection between collectives}} + \underbrace{\sum_k P'_k \sum_j z_{kj} (p'_{kj} - p_{kj})}_{\text{Weighted selection within each collective}}$$

Glossary

Indices

- i : Index for counting packages; index for counting particles in the whole population
- j : Index for counting particles within a collective
- k : Index for counting collectives

Basic terms

- p_i : population share of package i in the initial population. Constrained so that $\sum_i p_i = 1$
- z_i : trait value of package i in the initial population
- p'_i : population share of descendants of package i in the subsequent population. Constrained so that $\sum_i p'_i = 1$
- z'_i : average trait value of descendants of package i in the subsequent population

Derived terms

- w_i : selection coefficient of package i . Defined as $\frac{p'_i}{p_i}$
- \bar{z} : Average trait value in the initial population. Defined as $\sum_i p_i z_i$
- \bar{z}' : Average trait value in the subsequent population. Defined as $\sum_i p'_i z'_i$
- $\Delta \bar{z}$: Change in average trait value. Defined as $\bar{z}' - \bar{z}$
- $\Delta_s \bar{z}$: Change in average trait value due to selection
- $\Delta_t \bar{z}$: Change in average trait value due to factors other than selection

Statistical terms

- W : Selection coefficient considered as a random variable
- Z : Trait value considered as a random variable

Vector terms

- z_i : Categorical trait for type i , represented as a one-hot vector
- Z : Population distribution of categorical traits before selection. Defined as $\sum_i p_i z_i$
- z'_i : Distribution of trait values over type i 's descendants after selection
- Z' : Population distribution of categorical traits after selection. Defined as $\sum_i p'_i z'_i$

Multi-level terms

- p_i : Population share of particle i . Constrained so that $\sum_i p_i = 1$
- p_{kj} : Collective share of particle j in collective k . Constrained so that $\sum_j p_{kj} = 1$ for all k
- z_i : Trait value of particle i . Equal to z_{kj} when i and kj index the same particle
- z_{kj} : Trait value of particle j in collective k . Equal to z_i when i and kj index the same particle
- P_k : Population share of collective k . Defined as the sum of the population shares of its constituent particles
- Z_k : The trait value associated with collective k . Defined as $\sum_j p_{kj} z_{kj}$. In categorical settings it is a vector describing the proportions of different particle types in collective k
- P'_k : Population share of descendant collective of collective k . Defined as the sum of the population shares of its constituent particles
- Z'_k : Trait value associated with descendant collective of collective k . Defined as $\sum_j p'_{kj} z'_{kj}$. In categorical settings it is a vector describing the proportions of different particle types in the descendant of collective k

Acknowledgments

A great deal of discussion and input led to the writing of this paper, most prominently from Hedvig Skirgård, Angela Chira, Sandra Auderset, Cristian Juárez, Viktor Martinović, and the participants of the workshop “Predicting Evolution: The Price Equation and its Applications” in Hanover in March 2023. The manuscript was much improved thanks to comments from Manolo Martínez, Sergio Balari, Aida Roige, Oriol Roca-Martín, Josh Myers, Karl Bergman, and especially Cameron Rouse Turner. Lastly, a great deal of good advice and healthy criticism was provided by three anonymous reviewers, and has been gratefully incorporated into the final paper.

This work was supported by Juan de la Cierva grant FJC2020-044240-I and María de Maeztu grant CEX2021-001169-M funded by MICIU/AEI/10.13039/501100011033.


Literature cited

- Bourrat, Pierrick. 2021. *Facts, Conventions, and the Levels of Selection*. Elements in the Philosophy of Biology. Cambridge University Press. <https://doi.org/10.1017/9781108885812>.
- Bourrat, Pierrick. 2024. “Adding Causality to the Information-Theoretic Perspective on Individuality.” *European Journal for Philosophy of Science* 14 (1): 9. <https://doi.org/10.1007/s13194-023-00566-1>.
- Brodie, Edmund D., III. 1992. “Correlational Selection for Color Pattern and Antipredator Behavior in the Garter Snake *Thamnophis Ordinoidea*.” *Evolution* 46: 1284–98. <https://doi.org/10.1111/j.1558-5646.1992.tb01124.x>.
- Campbell, Donald T. 1956. “Perception as Substitute Trial and Error.” *Psychological Review* 63: 330–42. <https://doi.org/10.1037/h0047553>.
- Campbell, John O. 2016. “Universal Darwinism As a Process of Bayesian Inference.” *Frontiers in Systems Neuroscience* 10. <https://doi.org/10.3389/fnsys.2016.00049>.
- Csányi, V. 1980. “General Theory of Evolution.” *Acta Biologica Academiae Scientiarum Hungaricae* 31: 409–34.
- Dennett, Daniel Clement. 1995. *Darwin’s Dangerous Idea: Evolution and the Meanings of Life*. Penguin UK.
- El Mouden, C., J.-B. André, Olivier Morin, and D. Nettle. 2014. “Cultural Transmission and the Evolution of Human Behaviour: A General Approach Based on the Price Equation.” *Journal of Evolutionary Biology* 27 (2): 231–41. <https://doi.org/10.1111/jeb.12296>.
- Fisher, Sir Ronald Aylmer. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press.
- Frank, Steven A. 1998. *Foundations of Social Evolution*. Princeton University Press.
- Frank, Steven A. 2012a. “Natural Selection. III. Selection versus Transmission and the Levels of Selection.” *Journal of Evolutionary Biology* 25 (2): 227–43. <https://doi.org/10.1111/j.1420-9101.2011.02431.x>.
- Frank, Steven A. 2012b. “Natural Selection. IV. The Price Equation.” *Journal of Evolutionary Biology* 25: 1002–19. <https://doi.org/10.1111/j.1420-9101.2012.02498.x>.
- Frank, Steven A., and William Godsoe. 2020. “The Generalized Price Equation: Forces That Change Population Statistics.” *Frontiers in Ecology and Evolution* 8. <https://doi.org/10.3389/fevo.2020.00240>.
- Gardner, Andy. 2015. “The Genetical Theory of Multilevel Selection.” *Journal of Evolutionary Biology* 28: 305–19. <https://doi.org/10.1111/jeb.12566>.

- Gardner, Andy. 2020. "Price's Equation Made Clear." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1797): 20190361. <https://doi.org/10.1098/rstb.2019.0361>.
- Gardner, Andy, Stuart A. West, and G. Wild. 2011. "The Genetical Theory of Kin Selection." *Journal of Evolutionary Biology* 24: 1020–43. <https://doi.org/10.1111/j.1420-9101.2011.02236.x>.
- Haldane, John Burdon Sanderson. 1932. *The Causes of Evolution*. Harper & Brothers.
- Hodgson, Geoffrey M. 2004. *The Evolution of Institutional Economics: Agency, Structure, and Darwinism in American Institutionalism*. Psychology Press.
- Hull, David L. 2001. *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*. Cambridge University Press.
- Jablonka, Eva, and Marion J. Lamb. 2020. *Inheritance Systems and the Extended Evolutionary Synthesis*. Cambridge University Press. <https://doi.org/10.1017/9781108685412>.
- Jäger, Gerhard. 2010. "Applications of the Price Equation to Language Evolution." In *The Evolution of Language*, 192–97. Proceedings of the 8th International Conference (EVOLANG8). Utrecht, Netherlands: WORLD SCIENTIFIC. https://doi.org/10.1142/9789814295222_0025.
- Kerr, Benjamin, and Peter Godfrey-Smith. 2009. "Generalization of the Price Equation for Evolutionary Change." *Evolution* 63 (2): 531–36. <https://doi.org/10.1111/j.1558-5646.2008.00570.x>.
- Knudsen, Thorbjørn. 2004. "General Selection Theory and Economic Evolution: The Price Equation and the Replicator/Interactor Distinction." *Journal of Economic Methodology* 11 (2): 147–73. <https://doi.org/10.1080/13501780410001694109>.
- Lande, Russell, and Stevan J. Arnold. 1983. "The Measurement of Selection on Correlated Characters." *Evolution* 37: 1210–26. <https://doi.org/10.2307/2408842>.
- Luque, Victor J. 2017. "One Equation to Rule Them All: A Philosophical Analysis of the Price Equation." *Biology & Philosophy* 32 (1): 97–125. <https://doi.org/10.1007/s10539-016-9538-y>.
- Okasha, Samir. 2006. *Evolution and the Levels of Selection*. Oxford: Oxford University Press.
- Okasha, Samir, and Jun Otsuka. 2020. "The Price Equation and the Causal Analysis of Evolutionary Change." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1797): 20190365. <https://doi.org/10.1098/rstb.2019.0365>.
- Popper, Karl Raimund. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.
- Price, George R. 1970. "Selection and Covariance." *Nature* 227 (5257): 520–21. <https://doi.org/10.1038/227520a0>.
- Price, George R. 1995. "The Nature of Selection." *Journal of Theoretical Biology* 175 (3): 389–96. <https://doi.org/10.1006/jtbi.1995.0149>.
- Rice, Sean H. 2004. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer.
- Sober, Elliott. 1984. *The Nature of Selection*. MIT Press.
- Strand, Paul S., Mike J. F. Robinson, Kevin R. Fiedler, Ryan Learn, and Patrick Anselme. 2022. "Quantifying the Instrumental and Noninstrumental Underpinnings of Pavlovian Responding with the Price Equation." *Psychonomic Bulletin & Review* 29 (4): 1295–306. <https://doi.org/10.3758/s13423-021-02047-z>.
- Van Veelen, Matthijs. 2020. "The Problem with the Price Equation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1797): 20190355. <https://doi.org/10.1098/rstb.2019.0355>.

Wade, Michael J. 1985. "Soft Selection, Hard Selection, Kin Selection, and Group Selection." *The American Naturalist*. <https://doi.org/10.1086/284328>.

© 2025 Author(s)

This is an open-access article, licensed under
Creative Commons Attribution 4.0 International 

This license requires that reusers give credit to the creator(s). It allows reusers to
distribute, remix, adapt, and build upon the material in any medium or format.

ISSN 2475-3025