


# Causal attribution by the chain rule: unifying natural selection, learning, economics, and other disciplines

Steven A. Frank 

Department of Ecology and Evolutionary Biology, University of California, Irvine, United States  
Corresponding author: Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, United States. Email: [saf@stevefrank.org](mailto:saf@stevefrank.org)

## Abstract

Analysis often splits change into components. For example, how much of the observed variance is caused by genes or environment? In many cases, the split is ultimately made by the logic of the chain rule, which divides the difference of a product into two terms. Each term quantifies the partial difference associated with change in one component while holding the other component constant. The chain rule is of course widely known. However, this article argues that its deep fundamental role often goes unrecognized. The article shows how simply the basic chain rule unifies Fisher's fundamental theorem of natural selection, the Price equation description of evolutionary change, the Oaxaca–Blinder decomposition of wage differences in economics, the Kitagawa decomposition of mortality differences in demography, many expressions of thermodynamics, and most strikingly back propagation, the core optimization method of modern machine learning and artificial intelligence. The success in creating good designs and finding good solutions in both natural selection and artificial intelligence depends on how the chain rule propagates causes from instances of success or failure back to the underlying genes or parameters of the system. The mathematical analysis presented here shows that, for finite differences, the product rule form of the chain rule yields a basic decomposition of change into two components of a regression equation. That regression decomposition is purely a description of change with no explicit causal meaning. However, simple additional assumptions lead naturally to the modern counterfactual analysis of causality. From that perspective, we can easily understand the causal interpretation that Fisher gave to his fundamental theorem, and we can see the same causal structure in the Oaxaca–Blinder decomposition of economics and in causal analyses across many disciplines.

**Keywords** Fisher's fundamental theorem of natural selection, Price equation, Oaxaca–Blinder decomposition, counterfactual causal analysis, machine learning, back propagation

We often split observed variation into components. For example, how much of the change in a trait is caused by natural selection, and how much is caused by a change in the environment? This article shows that a simple common method often underlies such causal decomposition.

The general understanding of method leads to a broad conceptual unification of how one studies cause. And it leads to insight about the common basis by which natural selection and the complex computational algorithms of modern artificial intelligence gain information. Along the way, we also see the common basis of important decompositions for variation and causal analysis in other disciplines, including the famous Oaxaca–Blinder decomposition of economic analysis (Blinder, 1973; Fortin et al., 2011; Oaxaca, 1973).

The mathematical foundation arises from the chain rule, which divides the difference of a product into two terms. Each term quantifies the partial difference caused by change in one factor while holding the other factor constant. In standard calculus, the application of the chain rule to a product is called the product rule. This

simple identity turns out to be the common algebraic foundation for a remarkably wide range of analytical methods across the sciences.

An early example is Fisher's fundamental theorem of natural selection (Fisher, 1930, 1941, 1958). Fisher regressed the fitness of individuals on their genetic composition. He then split the total change in mean fitness into two terms. The first ascribed gene frequency changes to natural selection while holding constant the partial regression coefficients of individual genes on fitness. The second term ascribed regression coefficient changes to changes in the genetic context and the environment (Frank, 1997; Price, 1972b).

This article shows that Fisher's regression decomposition follows simply from the product rule for finite differences, which is the common name for the chain rule applied to a product when changes are not infinitesimally small.

My concern is not with the value of Fisher's theorem as a statement about mean fitness in populations. Rather, this article sets Fisher's theorem in the broader context of methods to decom-

Received: 25 December 2025. Revised: 27 March 2026. Accepted: 4 April 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE) and European Society for Evolutionary Biology (ESEB).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

pose change into parts and to consider the causal interpretation of the components. These methods clarify the history of Fisher's theorem. More importantly, the study of change and the analysis of cause are central goals in all of the sciences.

To develop that point, this article provides a very general mathematical derivation of the regression decomposition into a set of direct factors and a set of contextual factors. That decomposition arises simply from the product rule for finite differences. Fisher's theorem is used as an illustrative example.

Fisher argued that the first term of his regression decomposition isolates the direct force of natural selection from the broader complexity of evolutionary change, just as one might isolate gravity as a force from the actual trajectory of a body that depends on friction, constraints, and initial conditions. That causal assignment of credit for change to a specific factor while holding other factors constant is the logic of the modern counterfactual approach to causality (Holland, 1986; Pearl, 2009), which Fisher anticipated by decades.

Price and Ewens showed that Fisher's decomposition is an attempt at causal attribution (Ewens, 1989; Price, 1972b). But they both disputed his peculiar definitions and causal interpretation of natural selection. For this article, the primary point is Fisher's underlying setup and application of counterfactual causal analysis. Causal interpretation almost always has a subjective component, and it is Fisher's subjective choice for causal interpretation that many authors do not accept.

This article makes three contributions. First, by deriving the minimal algebraic structure, I show that Fisher's theorem, the Oaxaca–Blinder decomposition of economics (Fortin et al., 2011), the Kitagawa decomposition of demography (Kitagawa, 1955), the Price equation (Frank, 2012; Price, 1972a), and related results in thermodynamics all reduce to the product rule for finite differences, often applied to a regression equation.

Second, I show how the essence of modern counterfactual causal analysis arises in a simple and natural way from the chain rule. In particular, the chain rule assigns causal credit by tracing how a change in each component propagates through a system to affect the outcome. The basic regression decomposition that follows from the finite difference product rule form of the chain rule is purely a description of change with no explicit causal meaning. However, a simple interpretive perspective leads naturally to the modern counterfactual analysis of causality.

Third, the chain rule is how machine learning uses observed successes and failures to attribute cause to model parameters. Causal attribution provides parameter updates to improve performance (Goodfellow et al., 2016). Similarly, natural selection finds good designs by implicitly back propagating reproductive success through a chain-rule-like calculation that attributes cause to particular genes, leading to updated gene frequencies. In both cases, causal attribution is hindered by nonlinear interactions. But things typically work well enough to create the seeming miracles of natural and artificial life. I unified natural selection with other types of learning in a related article (Frank, 2025).

## Product rule for finite differences

Finite differences have the form  $\Delta y = y' - y$ . The prime denotes the value of the variable in a changed context. Let  $z = bx$  be the product of  $b$  and  $x$ . Then we can write the finite difference for  $z$  as

$$\Delta z = \Delta(bx) = b'x' - bx = (b + \Delta b)(x + \Delta x) - bx,$$

which simplifies to the product rule for finite differences

$$\Delta z = b\Delta x + x'\Delta b.$$

On the right side, the first product term is the change in  $x$  holding  $b$  constant, and the second product term is the change in  $b$  multiplied by  $x'$ , the value of  $x$  in the changed context. This is a particular type of chain rule expansion.

We often need the difference of sums. Let

$$z = \sum_{i=0}^n b_i x_i = \mathbf{b} \cdot \mathbf{x},$$

in which bold variables are vectors, and the dot denotes the dot product of two vectors, which as shown is the sum of the element-wise products of the vectors. Then the product rule for finite differences extends to

$$\Delta z = \mathbf{b} \cdot \Delta \mathbf{x} + \mathbf{x}' \cdot \Delta \mathbf{b}. \quad (1)$$

If the differences are small, we write the difference operator as a differential,  $\Delta \rightarrow d$ . Then, recalling that  $x' = x + \Delta x$  and replacing differences by differentials, we obtain

$$dz = \mathbf{b} \cdot d\mathbf{x} + \mathbf{x} \cdot d\mathbf{b} + d\mathbf{x} \cdot d\mathbf{b}.$$

For small differences, the final term is the second-order product of two small values, which is negligible relative to the first-order smallness of the prior terms. Thus, in the limit we drop the final term and get the product rule for differentials

$$dz = \mathbf{b} \cdot d\mathbf{x} + \mathbf{x} \cdot d\mathbf{b}.$$

However, many realistic analyses apply to finite differences with measurements at two separated times. So, the less known product rule for finite differences in equation 1 is often a useful expression, as shown in the next section.

## Difference in a regression model

Write the value of our focal variable as a regression equation

$$z_j = b_0 x_{0j} + \sum_{i=1}^n b_i x_{ij} + \epsilon_j$$

Here, we have a population of  $z$  values indexed by  $j$ . Each instance of  $z_j$  depends on the predictors,  $x_{ij}$ , with each predictor weighted by its associated partial regression coefficient,  $b_i$ . The term  $\epsilon_j$  is the residual or error.

In the intercept term,  $b_0 x_{0j}$ , going forward we set  $x_{0j} = 1$  for all  $j$ , so that  $b_0$  is the intercept. This allows us to write the regression more compactly in a way that matches the notation of other key steps in this article as

$$z_j = \sum_{i=0}^n b_i x_{ij} + \epsilon_j = \mathbf{b} \cdot \mathbf{x}_j + \epsilon_j. \quad (2)$$

Each instance  $j$  depends on the vector of predictors,  $\mathbf{x}_j$ , with each predictor weighted by its associated partial regression coefficient in  $\mathbf{b}$ . Subsequent sums for  $i$  also run over  $i = 0, 1, \dots, n$ .

In the underlying population of  $z$  values, suppose each value  $z_j$  occurs with frequency  $q_j$ . Then the mean value of  $z$  is

$$\bar{z} = \sum_j q_j z_j = \mathbf{q} \cdot \mathbf{z} = \sum_i b_i \sum_j q_j x_{ij} = \sum_i b_i \bar{x}_i = \mathbf{b} \cdot \bar{\mathbf{x}}.$$

Here, by standard frequency-weighted least squares for regression, the average residual is zero,  $\sum_j q_j \epsilon_j = 0$  (Rao, 1973; Searle, 1971; Seber & Lee, 2003).

For the regression of  $z$ , we want the difference between the current mean value and the mean value in a changed context,  $\Delta \bar{z} = \bar{z}' - \bar{z}$ , which we obtain by applying the general expression for finite differences in equation 1, yielding

$$\Delta \bar{z} = \Delta(\mathbf{b} \cdot \bar{\mathbf{x}}) = \mathbf{b} \cdot \Delta \bar{\mathbf{x}} + \bar{\mathbf{x}}' \cdot \Delta \mathbf{b}. \quad (3)$$

It is of course possible to interpret particular predictors as interactions between causal factors. But the point here concerns the general structure of the problem rather than the specific causal model.

## Fisher's fundamental theorem

Fisher's theorem is simply the application of the general expression for the difference in the mean value of a regression variable, given in equation 3. Let the variable of interest be fitness, denoted  $w \equiv z$ , and the difference in mean fitness be  $\Delta \bar{w} \equiv \Delta \bar{z}$ .

For simplicity, assume a haploid genetic model, in which the multiple loci in an individual are denoted by  $x_{ij}$  for the  $i = 1, \dots, n$  loci in the  $j$ th individual, with  $x_{0j} = 1$  as the indicator variable for the intercept. Suppose that at each locus there is an allele that takes on values of either 0 or 1. More complex genetics can be handled within similar regression expressions, but the added notation does not provide additional insight (Frank, 1997).

With those assumptions, the average value of a predictor in the regression is the frequency of the allele with value 1, or in standard genetic language,  $p_i = \bar{x}_i$  is the gene frequency at the  $i$ th locus. For the intercept,  $p_0 = 1$ .

Now the change in mean fitness follows immediately from equation 3 as

$$\Delta \bar{w} = \Delta(\mathbf{b} \cdot \mathbf{p}) = \mathbf{b} \cdot \Delta \mathbf{p} + \mathbf{p}' \cdot \Delta \mathbf{b}. \quad (4)$$

This expression describes the total change exactly. Fisher emphasized that the first term focuses on the change in gene frequencies caused by natural selection,  $\Delta \mathbf{p}$ , while holding constant the average effect of each gene on fitness, given by the partial regression coefficients,  $\mathbf{b}$ .

## Main result

If, without loss of generality, we assume that in the initial population  $\bar{w} = 1$ , and, following Fisher, define the marginal fitness of allele  $i$  as  $w_i = p'_i/p_i$ , then

$$\Delta p_i = p_i(w_i - 1) = p_i \alpha_i,$$

for which Fisher defined  $\alpha_i = w_i - 1$  as the average excess in fitness. Then the first term of the finite difference of the regression in equation 4 is

$$\Delta \bar{w}_{ns} = \mathbf{b} \cdot \Delta \mathbf{p} = \sum_i b_i \Delta p_i = \sum_i p_i \alpha_i b_i, \quad (5)$$

which is often called the additive genetic variance in fitness, as shown in the following subsection. Fisher called this the fundamental theorem of natural selection. The result is that the partial change in mean fitness caused by the change in gene frequencies while holding constant the average effects is the additive genetic variance in fitness.

In Fisher's causal perspective, changes in gene frequencies,  $\Delta \mathbf{p}$ , are caused directly by natural selection, and changes in the partial regression coefficients,  $\Delta \mathbf{b}$ , are caused by change in context.

## Details on genetic variance

This subsection shows why the expression on the right side of equation 5 is a genetic variance. These details can be skipped on first reading.

From equation 2, we can write the regression expression for fitness as

$$w_j = \sum_i b_i x_{ij} + \epsilon_j = g_j + \epsilon_j,$$

in which  $g_j = \sum_i b_i x_{ij}$  is the genetic value of an individual, and here we are indexing fitness by  $j$  rather than  $i$  as above. Thus, compactly,  $w = g + \epsilon$ , the sum of the modeled genetic and residual effects. Next, recall that  $q'_j = q_j w_j$  for our assumed mean fitness of  $\bar{w} = 1$ , and  $p_i = \sum_j q_j x_{ij}$ , so that

$$\begin{aligned} \Delta p_i &= \sum_j q'_j x_{ij} - \sum_j q_j x_{ij} = \sum_j q_j (w_j - 1) x_{ij} \\ &= \text{Cov}(x_i, w). \end{aligned}$$

Then, in equation 5, we can equivalently write

$$\sum_i b_i \Delta p_i = \text{Cov}(\sum_i b_i x_i, w) = \text{Cov}(g, w) = \text{Var}(g),$$

because  $g$  and  $\epsilon$  are uncorrelated by least squares regression (Rao, 1973; Searle, 1971; Seber & Lee, 2003). Thus, the expression in equation 5 is the genetic variance for the particular model of genetic effects.

## Interpretation

Fisher's theorem states, from the first term of equation 4, that the partial change in mean fitness caused by natural selection is  $\mathbf{b} \cdot \Delta \mathbf{p}$ , freezing the partial regression coefficients  $\mathbf{b}$  at their initial values while letting gene frequencies  $\Delta \mathbf{p}$  change. Everything that changes the mapping from the given genetic predictors to fitness is pushed into the second term,  $\mathbf{p}' \cdot \Delta \mathbf{b}$ .

It is obvious that Fisher's theorem does not fully account for how natural selection changes total mean fitness. The average effects of genes, which are the partial regression coefficients, typically depend on the gene frequencies that are changed by natural selection. So what does it mean to say that the direct cause of natural selection is only through changes in gene frequencies while holding constant the average effects of those genes?

There is also the problem that interactions between genes, such as dominance and epistasis, influence fitness. Those interactions depend on genotypic arrangements, which are altered by natural selection.

What was Fisher thinking? No one knows for sure. Certainly, Fisher understood all of this. He essentially invented the theory of classical statistical inference (Fisher, 1925, 1922). And his 1918 article was the first to express clearly the various components of variance in the expression of traits, including the components associated with dominance and epistasis (Fisher, 1918).

Our strongest clues about Fisher's thoughts come from his own words (Fisher, 1930). The first sentences of his great book on natural selection are: "Natural Selection is not Evolution. Yet, ever since

the two words have been in common use, the theory of Natural Selection has been employed as a convenient abbreviation for the theory of Evolution by means of Natural Selection, put forward by Darwin and Wallace. This has had the unfortunate consequence that the theory of Natural Selection itself has scarcely ever, if ever, received separate consideration.” Put another way, Fisher’s goal was absolutely focused on separating natural selection as a specific force from its wider context of evolutionary change.

Further, Fisher emphatically stated in 1941 that he had never written about the evolution of mean fitness, and wondered why Sewall Wright kept making the mistake of linking his fundamental theorem of natural selection to an overall conclusion about mean fitness (Fisher, 1941).

Fisher’s way of isolating natural selection was to focus on gene frequency changes while holding constant the average effects of the genes. That does not work if one’s goal is to calculate the dynamics of mean fitness evolution, as noted at the start of this subsection. But there is a certain logic to it that fits the common theories of causal analysis that developed subsequent to Fisher’s writings (Pearl, 2009).

In particular, Fisher noted that natural selection changes the frequencies of genes. That change in gene frequencies is directly transmitted to the next generation. Natural selection also changes genotype frequencies. But in standard Mendelian genetics, recombination, recombination, and mating all change genotype frequencies while holding constant gene frequencies. So the pure direct effect of natural selection is only in the gene frequency changes. All else is context.

Thus, a natural causal counterfactual that one may use to isolate natural selection from its broader evolutionary context is the change in gene frequency. That approach does not make the complexity of calculating the evolution of mean fitness go away. But it is consistent with the modern theory of causal analysis by counterfactual perturbation (Pearl, 2009), which Fisher understood many decades earlier.

In particular, Fisher recognized that the partial regression coefficients (average effects) compute the counterfactual causal effect of randomized gene substitution, averaging over nonlinear genotypic interactions as they occur in the initial population, thereby isolating each gene’s population-specific effect on fitness.

There is no resolution between those who find such attempts at isolating a force from its system as profoundly misleading and those who find such isolation as a good approach to partial causal understanding. My own view is that one should understand both perspectives.

The benefit of full exact calculation of total change is easy to understand. In favor of partial analysis, most learning and optimization methods work by partial perturbation, evaluating changes in one set of parameters while holding others fixed (Frank, 2025; Goodfellow et al., 2016; Nocedal & Wright, 2006). One calculates partial perturbations because one can rarely do complete calculations for how changing a system alters its performance. For example, gradient descent in machine learning updates parameter weights by their partial derivatives with respect to a loss function, which is similar to Fisher’s isolation of gene frequency changes while holding average effects constant.

Ultimately there are benefits in separating how one thinks about forces from how one analyzes dynamics. Lanczos emphasized this separation in his classic treatment of the variational principles in mechanics, in which he isolated individual

forces from the full complexity of constrained motion (Lanczos, 1986).

How closely does Fisher’s approach match modern counterfactual causal analysis? In an earlier article (Frank, 1997), I showed that the Price equation combined with regression provides a common path-analytic causal framework that unifies Fisher’s theorem, Robertson’s (1966) covariance theorem for quantitative genetics, the Lande-Arnold (1983) model for the analysis of natural selection, and Hamilton’s (1970) rule for kin selection.

Lee and Chow (2013) later showed that Fisher’s average effect corresponds to a causal intervention under Pearl’s do-operator (Pearl, 2009), a formal method for counterfactual analysis. Otsuka and Okasha extended a causal interpretation of the Price equation through Pearl’s counterfactual methods (Okasha & Otsuka, 2020; Otsuka, 2016). Other studies similarly developed formal counterfactual analyses for selection gradients and kin selection (Fromhage & Jennions, 2019; Henshaw et al., 2020).

These various studies develop increasingly elaborate formal machinery to arrive at the insight that Fisher’s chain-rule partition can provide counterfactual causal attribution. The present article shows how simply this point arises from first principles.

## The Price equation

The Price equation is another product rule expression for finite differences (Frank, 2012; Price, 1972a). This equation is widely used in ecology, evolutionary biology, and other disciplines. Define the population mean of a variable as  $\bar{z} = \sum_j q_j z_j = \mathbf{q} \cdot \mathbf{z}$ , in which  $q_j$  is the frequency of the value  $z_j$ . Then, by our standard expression for finite differences

$$\Delta \bar{z} = \Delta \mathbf{q} \cdot \mathbf{z} + \mathbf{q}' \cdot \Delta \mathbf{z}. \quad (6)$$

which is my generalization of the Price equation (Frank, 1997, 2012). The Price equation was originally expressed in terms of covariance and expectation functions (Price, 1972a), which can be obtained by noting that we can write frequency changes in terms of fitness as in the prior section as  $\Delta q_j = q_j(w_j - 1)$ , again assuming that  $\bar{w} = 1$  in the initial population. Also note that from the prior section we have for the relation between frequencies and fitness  $q'_j = w_j q_j$ . Then we can expand equation 6 as

$$\begin{aligned} \Delta \bar{z} &= \sum_j q_j (w_j - 1) z_j + \sum_j q_j w_j \Delta z_j \\ &= \text{Cov}(w, z) + E(w \Delta z). \end{aligned}$$

## Partitions in other disciplines

### Economics

A simple example of the Oaxaca–Blinder decomposition (Blinder, 1973; Fortin et al., 2011; Oaxaca, 1973) can be written for the difference in average income between two groups. For the first group, we start with the regression of income,  $z$ , on a vector of predictors,  $\mathbf{x}$ , as

$$z_j = \mathbf{b} \cdot \mathbf{x}_j + \epsilon_j$$

in which predictors can be variables such as education level. The equation follows the same assumptions as equation 2. Because

the average of the error term is zero, we have as before  $\bar{z} = \mathbf{b} \cdot \bar{\mathbf{x}}$ . Then, we get the difference in mean incomes between groups as the same expression we found in [equation 3](#) as

$$\Delta \bar{z} = \mathbf{b} \cdot \Delta \bar{\mathbf{x}} + \bar{\mathbf{x}}' \cdot \Delta \mathbf{b}.$$

For example, suppose the first group has a particular relation between education and income, described by the regression coefficients,  $\mathbf{b}$ . The second group has a different average for education level,  $\Delta \bar{\mathbf{x}}$ . The first term predicts the change in income using the income–education relations in the first group. The second term describes how changes in the income–education relations in the second group alter the predicted change in income between the two groups.

Overall, we decompose the total change into the difference ascribed to changes in education holding constant the income–education relationship from the first group and the difference ascribed to changes in the income–education relationship holding the education levels constant from the values of the second group.

This decomposition was originally developed to study wage discrimination between groups defined, for example, by gender or race. The key question was how much of the observed wage gap could be attributed to differences in observable characteristics such as education, and how much might be attributed to differences in the relation between income and education, potentially reflecting discrimination.

The parallel with Fisher's theorem is direct but inverts emphasis. Fisher focused on the first term—the change in gene frequencies holding constant the regression coefficients—as the quantity of causal interest. Fisher's second term treats the shift in regression coefficients as a remainder reflecting changed context.

By contrast, economists often care most about the second term, because one cause for the shift in regression coefficients may be hidden discrimination, capturing the portion of the wage gap that cannot be explained by the differences in observable predictors. The different foci arise entirely from the causal question being asked.

## Demography

Kitagawa decomposed differences between two demographic rates (Kitagawa, 1955). For example, the difference between the death rates of two populations can be partitioned into components. The first is the differences in age composition, holding age-specific mortality rates constant. The second is the differences in age-specific mortality rates, evaluated at the changed age composition.

This partition follows [equation 3](#), with fraction of individuals at each age playing the role of predictor means and age-specific mortality rates playing the role of regression coefficients. This decomposition is widely used in demography (Fortin et al., 2011).

## Thermodynamics

Nicholson et al. decomposed the change in the mean value of a thermodynamic ensemble into a term for changes in the probability distribution holding observable values constant and a term for changes in observable values holding the distribution constant (Nicholson et al., 2020). They used this decomposition to unify

several results in irreversible thermodynamics and information theory.

Frank and Bruggeman showed that this thermodynamics decomposition is identical to the Price equation, given in [equation 6](#) (Frank & Bruggeman, 2020). All of these decompositions are applications of the simple product-rule partition in [equation 1](#), which leads directly to the regression partitioning commonly used in applications.

## Discussion

The product-rule partition and its application to regression are purely algebraic and hold without any causal assumptions. Causality enters only when one interprets the regression coefficients as effects, such as Fisher's average effects, and when one chooses a what-if counterfactual comparison that specifies what is held fixed across contexts (Holland, 1986; Pearl, 2009).

Fisher's theorem corresponds to the counterfactual in which the regression coefficients linking individual genes to fitness are held constant while gene frequencies change. The remainder term collects changes in that mapping due to genetic interactions, genetic background, and environment. This perspective explains both the usefulness of Fisher's partition for a particular causal question and why disagreements persist when readers prefer a different counterfactual analysis.

The same partition supports different causal questions depending on which term one treats as the quantity of interest. Fisher focused on the first term to isolate the direct force of natural selection. Economists often focus on the second term, in which shifts in regression coefficients may reveal discrimination. Demographers use both terms symmetrically to separate compositional change from rate change. In each case the mathematics is identical. What differs is the substantive question that motivates the partition.

Fisher, the economists who developed the Oaxaca–Blinder decomposition, and the machine learning engineers who use back propagation all apply the same chain-rule mathematics. They differ only in the causal question they ask. Decades of econometric work on the Oaxaca–Blinder decomposition have addressed the choice of reference group, the sensitivity of the partition to the specific predictors used, and the conditions under which the decomposition supports causal inference (Fortin et al., 2011). Those lessons apply to any chain-rule causal decomposition, including Fisher's theorem.

The chain rule is an elementary mathematical identity. The main point of this article is that much of causal analysis across the sciences reduces to this identity. Fisher saw it in 1930 and built his fundamental theorem of natural selection on it. The economists rediscovered it independently. Back propagation exploits the same principle to train the large models on which modern artificial intelligence depends. Recognizing the common foundation clarifies understanding, connects previously isolated disciplines, and reveals the deep simplicity beneath seemingly complex methods of causal attribution.

## Data and code availability

This work did not generate new data or code.

## Funding

The Donald Bren Foundation and National Science Foundation grant DEB-2325755 support my research.

## Conflict of interest

The author declares no conflict of interest.

## Acknowledgments

I thank Jeff Zabel for pointing out the similarity between the Oaxaca–Blinder decomposition and some classic equations in biology, and for his general insights about causality across disciplines.

## References

- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.
- Ewens, W. J. (1989). An interpretation and proof of the fundamental theorem of natural selection. *Theoretical Population Biology*, 36, 167–180.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399–433.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A*, 222, 309–368.
- Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22, 700–725.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon.
- Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics*, 11, 53–63.
- Fisher, R. A. (1958). *The genetical theory of natural selection* (2nd ed.). Dover Publications.
- Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. *Handbook of Labor Economics*, 4, 1–102.
- Frank, S. A. (1997). The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution*, 51, 1712–1729.
- Frank, S. A. (2012). Natural selection. IV. The Price equation. *Journal of Evolutionary Biology*, 25, 1002–1019.
- Frank, S. A. (2025). The Price equation reveals a universal force-metric–bias law of algorithmic learning and natural selection. *Entropy*, 27(11), 1129.
- Frank, S. A., & Bruggeman, F. J. (2020). The fundamental equations of change in statistical ensembles and biological populations. *Entropy*, 22(12), 1395.
- Fromhage, L., & Jennions, M. D. (2019). The strategic reference gene: An organismal theory of inclusive fitness. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20190459.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, 228, 1218–1220.
- Henshaw, J. M., Morrissey, M. B., & Jones, A. G. (2020). Quantifying the causal pathways contributing to natural selection. *Evolution*, 74, 2560–2574.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272), 1168–1194.
- Lanczos, C. (1986). *The variational principles of mechanics* (4th ed.). Dover Publications.
- Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 37, 1212–1226.
- Lee, J. J., & Chow, C. C. (2013). The causal meaning of Fisher's average effect. *Genetics Research*, 95, 89–109.
- Nicholson, S. B., García-Pintos, L. P., del Campo, A., & Green, J. R. (2020). Time–information uncertainty relations in thermodynamics. *Nature Physics*, 16, 1211–1215.
- Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). Springer Series in Operations Research and Financial Engineering. Springer.
- Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693–709.
- Okasha, S., & Otsuka, J. (2020). The Price equation and the causal analysis of evolutionary change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375, 20190365.
- Otsuka, J. (2016). Causal foundations of evolutionary genetics. *British Journal for the Philosophy of Science*, 67, 247–269.
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Price, G. R. (1972a). Extension of covariance selection mathematics. *Annals of Human Genetics*, 35, 485–490.
- Price, G. R. (1972b). Fisher's "fundamental theorem" made clear. *Annals of Human Genetics*, 36, 129–140.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). Wiley.
- Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Production*, 8, 95–108.
- Searle, S. R. (1971). *Linear models*. Wiley.
- Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Wiley.

Received: 25 December 2025. Revised: 27 March 2026. Accepted: 4 April 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE) and European Society for Evolutionary Biology (ESEB). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.