The New Palgrave Dictionary of Economics Online

explanation

Julian Reiss
From The New Palgrave Dictionary of Economics, Second Edition, 2008
Edited by Steven N. Durlauf and Lawrence E. Blume

Abstract

Explaining socio-economic phenomena is one important aim of economics. There is very little agreement, however, on what precisely constitutes an adequate economic explanation. Starting from the very influential but defective 'deductive-nomological model' of explanation, this article describes and criticizes the major contemporary competitors for such an account (the probabilistic–causal, the mechanistic–causal and the unificationist models) and argues that none of them can by itself capture all aspects of a good explanation. When seeking to explain a socio-economic phenomenon it should therefore be borne in mind that different types of explanation serve different purposes.

Keywords

causality in economics and econometrics; economic laws; equilibrium; explanation; laws of nature; methodology of economics; positive economics; probability; statistical inference

Article

In the early 1950s Milton Friedman famously declared that the 'ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (that is, not truistic) predictions about phenomena not yet observed' (Friedman, 1953, p. 7). Today, after the demise of logical positivism in philosophy and positivistic trends in economics, economists tend to regard the explanation of phenomena as one legitimate aim of economics besides the more directly policy-oriented aims of prediction and control. Perhaps, following Friedman, explaining a phenomenon is primarily of instrumental value for the preparation and guidance of policy. But perhaps economists seek to explain in order to increase our understanding of the economic world, for purely cognitive reasons. Whether derivative or fundamental, explanation is a major goal that economists pursue and understanding what exactly is sought is an important task for economic methodology.
An adequate account of explanation in economics should satisfy at least three desiderata:

1. (a) it should be *descriptively adequate*; that is, it should be consistent with economic practice;
2. (b) it should be *epistemically adequate*; that is, it should give reason to believe that that which it identifies as an explanation is indeed explanatory; and
3. (c) it should be *empirically adequate*; that is, it should not identify something as an explanation unless it is based on sufficient evidence.

The so-called deductive–nomological or DN model of explanation (Hempel and Oppenheim, 1948) can rightly be regarded as the received view of scientific explanation. Although the theory is now generally regarded as untenable, it is it is useful to consider its guiding ideas as a starting point because its flaws motivate the alternative, more satisfactory accounts.

The deductive–nomological model

According to the DN model, an explanation is an argument whose premises constitute the so-called explanans (or 'that which explains') and whose conclusion constitutes the so-called explanandum (or 'that which is to be explained'). The explanandum will usually be a description of a noteworthy singular event (such as 'Black Monday', 'the rise of the dot.com industry' or 'the collapse of the Tiger economies') or a repeated pattern of events, which may be called a 'phenomenon' (such as 'hyperinflations', 'the J-curve effect' or 'the price drop of cars that have just left the showroom').
The adjectives 'deductive' and 'nomological' indicate that the argument must meet at least two criteria in order to count as an explanation. First, the argument must be deductively valid, that is, the explanandum must follow logically from the explanans. Second, among the premises of the explanans there must be at least one law of nature (the Greek word *nómos* means habit or law). Typically, it is also demanded that the premises of the explanans be true or at least verified. However, none of these criteria is individually necessary nor are the criteria jointly sufficient.
In many cases explanations are probabilistic rather than deterministic and thus the explanandum does not always logically follow from the explanans. John Doe's exposure to asbestos explains his contraction of

lung cancer but the statement 'John contracted lung cancer' is not entailed by the statement 'John was exposed to asbestos'. Second, and related, laws of nature in the sense of universal regularities are few and far between, especially in non-fundamental sciences such as economics. All so-called 'laws' in economics, such as the law of supply and demand, the iron law of wages, Okun's law, Say's Law and so forth are, at best, true *ceteris paribus*, that is, if nothing intervenes and relative to a specific institutional structure. For example, we can use the law of supply and demand to predict that demand for a good will decrease when a tax is imposed. However, depending on what else happens in the economy actual demand may or may not decrease. If disposable incomes rise sufficiently or if preferences change in the right way, demand may in fact increase.

Third, it is not clear whether laws in the sense used by proponents of the DN model are explanatory at all. Suppose that it is a law – a universal regularity – that economic expansions follow monetary expansions. Economists no doubt regard knowledge of this kind as very valuable, but unless more is told about the relationship it would hardly count as explanatory. The DN model is therefore neither descriptively nor epistemically adequate.

In response to these and other difficulties of the DN model (for a valuable discussion of many of the criticisms, see van Fraassen, 1980, pp. 103–29) philosophers have developed alternative accounts of explanation. One tradition holds that to explain a phenomenon means to cite the causes of the phenomenon. It therefore roughly agrees with the 'nomological' part of the DN model but replaces the notion of law with that of cause. Another tradition holds that to explain a phenomenon means to show how it fits into a systematization of our beliefs about the world. It agrees with the 'deductive' part of the DN model and insists that good explanations are those that unify diverse sets of beliefs. Both traditions can be found in economics, and both come in two variants.

The probabilistic–causal model

The chief difficulty for the causalist, who maintains that to explain a phenomenon is to provide information about its causes, is to elucidate the notion of cause. We believe that a tightening of the money stock explains the subsequent increase in interest rates; a change in minimum wages explains changes in the employment rate; veteran status explains earnings. In none of these cases is there a universal regularity between event-types; rather, earlier events appear to be probabilistic causes in the sense that they are *statistically relevant*.

One view thus held that event X explains event Y if the probability of Y in some population described by Z is different when X is present from when it is absent: $P(Y|X, Z) \neq P(Y|Z)$ (cf. Salmon, 1971). In econometrics this idea is akin to the notion of a multiple regression:

$$Y = \alpha X + \beta Z + \epsilon,$$

where Y is the explained variable, X is the explanatory variable and $\mathbf{Z}$ is a vector of background variables. X is statistically relevant to Y if and only α is different from zero, and can thus be used to explain Y.

Not all statistically relevant events appear to be explanatorily relevant, however. A drop in the barometer reading raises the probability of a storm but the barometer reading does not explain the storm. It is a common cause – the change in atmospheric pressure – that explains its joint effects, the barometer reading and the storm.

This suggests that X plus the set of background factors must constitute the full set of causes of the explained variable (cf. Cartwright, 1983, Essay 1): in any population that is homogenous with respect to atmospheric pressure, the barometer reading is statistically irrelevant to the occurrence of the storm. An obvious drawback of this account is that it asks for immense amounts of background knowledge for identifying explanatory factors from statistics. It requires that all other causes of a phenomenon (that is, all confounding factors) to be known or known to be distributed equally between a treatment and a control group, as in a randomized trial. Given the complexities of the social world, one can expect this requirement to be met only exceptionally.

There are also more principled difficulties. One problem arises because some factors may act differently depending on what other causes are present. An increase in the money stock may have different effects on the economy depending on the interest rate and investor behaviour. In extreme situations increasing money may have no effects on the economy at all, and so the government's ability to conduct monetary policy is thus incapacitated. It therefore seems exaggerated to demand that explanatory factors raise the probability of the explained variable in *all* causally homogeneous populations (which is presupposed by the linear models favoured by econometricians). But is it enough for a factor to raise the probability of the effect in one single population or should it raise its probability on average (cf. Dupré, 1984)?

Moreover, when factors act genuinely probabilistically, factors can be statistically relevant despite the fact that they are not explanatorily relevant, even when all other causes have been included. Suppose that in some causally homogeneous background Z, money M is a probabilistic cause of both nominal income Y as well as the level of prices L. Let $P(Y|M, Z)=P(L|M, Z)=0.8$. Now, let money cause income on precisely those occasions that it causes prices and vice versa. Then, $1=P(Y|L, M, Z)>P(Y|M, Z)=0.8$ even though the change in prices does not explain the change in nominal income – it is a mere correlate (cf. Cartwright, 1999, ch. 5).

But it is important to keep practical and epistemic issues apart. *If* one knows that *C* is a probabilistic cause of a phenomenon of interest *E*, then there is no reason to deny that one can use *C* in an explanation of *E*. The epistemic adequacy of the probabilistic–causal model derives from the general acceptance of causes as explanatory factors. However, finding out if *C* is a probabilistic cause of *E* will often face insurmountable practical difficulties.

The mechanistic–causal model

In philosophy of science, the mechanistic–causal model has been mostly associated with the name Wesley Salmon (see for instance Salmon, 1984). It attempts to improve upon the probabilistic model on two counts. On the one hand, for practical purposes it may be easier to find out whether *C* causes *E* by investigating whether or not there is a mechanism running from *C* to *E* than by statistical inference. For instance, Milton Friedman and Anna Schwartz (1963, p. 59) write:

> However consistent may be the [statistical] relation between monetary change and economic change, and however strong the evidence for the autonomy of the monetary changes, we shall not be persuaded that the monetary changes are the source [that is, cause] of the economic changes unless we can specify in some detail the mechanism that connects the one with the other.

Indeed, if *C* causes *E* we expect there to be a mechanism running from *C* to *E*. Evidence about a mechanism from *C* to *E* can thus provide evidence for a causal connection. In turn, according to this view, the mechanism can be used to explain *E*.
On the other hand, causal explanations that are based on statistical inferences often cite relationships among aggregate factors such as the money stock, the unemployment rate, inflation and so on, and can arguably be said to be somewhat shallow. Perhaps a monetary expansion can be used to explain a subsequent economic expansion because there is statistical evidence that the former is the cause of the latter. In this way one learns at best *that* the monetary change causes the economic change. Describing the transmission mechanism one further learns *how* the monetary change causes the economic change. The explanation is thus arguably more detailed, deeper.
It cannot be said, however, that the mechanistic account wins unequivocally over the probabilistic account on both fronts. In order to meet the empirical adequacy desideratum, a mechanistic explanation must be based on evidence no less than a probabilistic-causal explanation. A mere ‘sketch’ of a mechanism (such as the sketch of the transmission mechanism that follows above quotation by Friedman and Schwartz) does not explain anything. Usually, mechanistic explanations cite relationships among individuals, their preferences and external constraints. The argument that such hypotheses are more readily verifiable (for instance, because they may be verifiable by introspection) goes back at least to the writings of John Stuart Mill (1830). But it is not clear whether it is always easier to provide evidence for causal mechanisms that run at the micro level than for aggregate causal relationships. For instance, the problem of confounding factors is in no way confined to statistical inferences among aggregate variables, and, at the micro level, can only seemingly be alleviated by assuming away the operation of confounders a priori.
Moreover, although with some justification it can be said that mechanistic explanations are deeper than aggregate explanations, there are situations in which information about exactly how some variable influences another is entirely irrelevant. A policymaker, for instance, may be more interested in what is common among expansion episodes rather than in the exact processes that made them happen – which may be different on each occasion.

How models explain: unificationism

Let us now move from the applied side to the more theoretical side of economics. Consider the following quotation (Akerlof, 1970):

> From time to time one hears either mention of or surprise at the large price difference between new cars and those which have just left the showroom. The usual lunch table justification for this phenomenon is the pure joy of owning a ‘new’ car. We offer a different explanation.

Akerlof then describes an asymmetric-information model in which low-quality second-hand cars drive higher-quality cars out of the market, which leads to a decrease in average quality and prices. One way to interpret what Akerlof does is to regard the explanatory power of models such as his as consisting in an ability to suggest *schemas* that allow the description of a wide variety of different and seemingly unconnected phenomena. In Akerlof's original article, for instance, the model of the second-hand car market is regarded as a mere ‘finger exercise’ for further application in markets as diverse as insurance, labour, other goods and credit markets. Other economists invoke transaction costs to explain the existence of firms, intergovernmental collaboration, why crime rates are higher among the poor and ‘fair use’ doctrines about the use of copyrighted material among many other phenomena. Many other salient theoretical concepts in economics play a similar unifying role.
Philip Kitcher developed the idea that to explain a phenomenon means to derive a description of the phenomenon from an instance of an argument pattern, instances of which can be used for deriving

descriptions of many different kinds of phenomena into a formal account (Kitcher, 1989). Despite its intuitive appeal, however, it is quite clear that unification cannot be all there is to explanation. How could we tell whether those factors that are salient in a model are the ones that drive the results in the real world? This is a particular problem in economics as many of the concepts that do the alleged explanatory work in models such as Akerlof's are not very discriminatory. There is hardly any market transaction that is not characterized by asymmetric information because it is virtually always the case that one party knows more or something different about a contractually relevant property. Similar observations can be made about other concepts such as human capital or transaction costs or imperfect information. In some situations such factors will be the ones that drive the result, in others they will merely provide a background against which other factors operate. But this is a dominantly qualitative question that should be decided by empirical means, not by means of models alone. Moreover, it seems unlikely that unification is necessary for explanation. Many economic events will be explained with reference to very local and idiosyncratic processes such as wars, innovations and individuals' decisions that lack the power to unify whole classes of events (for further criticism of the unification model, see Woodward, 2003, ch. 8).

Nevertheless, unification plays at least two important roles in economic explanation. First, models such as Akerlof's suggest factors that may be causes of real phenomena. Unifying model schemas thus have an important heuristic role. Second, unifying explanations are in some sense desirable explanations. Even though the causal role a factor plays in bringing about a phenomenon is that which makes a model that describes the operation of this factor explanatory, it is its ability to systematize our beliefs and to reduce the number of 'brute facts' we have to accept as given that makes the explanation attractive to economists.

Equilibrium explanation

Economics is full of equilibrium notions such as the Nash equilibrium, evolutionarily stable equilibrium, sunspot equilibrium, partial and general equilibrium theories. For a variety of reasons, economists tend to downplay explanatory accounts if these accounts do not have a bearing on theory (Heckman, 2000, p. 85):

> Applications of this approach [that aims at the statistically analysis of 'natural experiments'] often run the risk of producing estimates of causal parameters that are difficult to interpret. Like the evidence produced in VAR [vector autoregression] accounting exercises, the evidence produced by this school is difficult to relate to the body of evidence about the basic behavioural elasticities of economics. The lack of a theoretical framework makes it difficult to cumulate findings across studies, or to compare the findings of one study with another. Many applications of this approach produce estimates very similar to biostatistical 'treatment effects' without any clear economic interpretation.

Equilibrium explanations, of course, have exactly this virtue: they show how some phenomenon can be systematized in a theoretical framework. Equilibrium explanations obtain at a level in between aggregate probabilistic explanations and causal-mechanical explanations. Unlike aggregate explanations, they are always formulated in terms of micro entities such as preferences, production possibilities and so forth. But, unlike causal–mechanical explanations, they rarely specify the exact details of how an equilibrium is reached or how an economy moves from one equilibrium to another. Equilibrium explanations abstract from the causal dynamics and focus on static end-points. Elliott Sober therefore argues (Sober, 1983, p. 204, emphasis in original):

> Equilibrium explanations present *disjunctions* of possible causal scenarios; the actual cause is given by one of the disjuncts, but the explanation doesn't say which.

Because of this, equilibrium explanations are more unifying than explanations that describe the actual causal mechanism that lead to the equilibrium. Nevertheless, equilibrium explanations (at least in economics) tend to cite a lot of information about causes such as preferences, productivity growth, technology and so on. Although abstracting from *some* causal detail, equilibrium explanations can thus safely be regarded as a species of causal explanation.

The greatest challenge for equilibrium explanations is, however, to meet the desideratum of empirical adequacy. In order to derive any results in an equilibrium model, usually a large number of highly distorting idealizations have to be made: consumers maximize utilities and producers their profits; they operate under perfect information; markets clear instantaneously; goods are infinitely divisible and so on and so forth. Furthermore, results derived from a model making such idealizations tend to be very sensitive to specification changes. There is therefore little reason to believe that those forces that drive the equilibrium results obtain also outside the model. Hence, unless it can be shown that this is the case, equilibrium models should be regarded as mere potential explanations.

Conclusion: the variety of causal explanations

The different types of explanation perform different epistemological roles. Very detailed causal–mechanistic explanations can be contrastive: they can provide information about what is special about the way in which a phenomenon came about, the way in which its causal history differs from the causal

histories of other, similar phenomena. Aggregate and unifying explanations, by contrast, are comparative: they provide information about what similar or different phenomena have in common (cf. Pettit, 1993, pp. 253–7). For those who are interested in explanation mostly for the practical goals of prediction and policy, aggregate explanations will often be the relevant type. However, mechanistic knowledge can be used to improve predictions, for instance, because it may provide information about the ways in which aggregate relationships sometimes fail to hold. Those with more purely cognitive goals will often prefer explanations that unify.

See Also

- causality in economics and econometrics
- models

Bibliography

Akerlof, G. 1970. The market for ‘lemons’: quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488–500.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cartwright, N. 1999. *The Dappled World*. Cambridge: Cambridge University Press.

Dupré, J. 1984. Probabilistic causality emancipated. In *Midwest Studies in Philosophy* 9, ed. P. French, T. Uehling, Jr. and H. Wettstein. Minneapolis: University of Minnesota Press.

Friedman, M. 1953. The methodology of positive economics. In *Essays in Positive Economics*. Chicago: University of Chicago Press.

Friedman, M. and Schwartz, A. 1963. Money and business cycles. *Review of Economics and Statistics* 45, 32–64.

Heckman, J. 2000. Causal parameters and policy analysis in economics: a twentieth century perspective. *Quarterly Journal of Economics* 115, 45–97.

Hempel, C. and Oppenheim, P. 1948. Studies in the logic of explanation. *Philosophy of Science* 15, 135–75.

Kitcher, P. 1989. Explanatory unification and the causal structure of the world. In *Scientific Explanation*, ed. P. Kitcher and W. Salmon. Minneapolis: University of Minnesota Press.

Mill, J.S. 1830. *Essays on Some Unsettled Questions of Political Economy*. London: Parker.

Pettit, P. 1993. *The Common Mind*. Oxford: Oxford University Press.

Salmon, W. 1971. Statistical explanation. In *Statistical Explanation and Statistical Relevance*. Pittsburgh: Pittsburgh University Press.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Sober, E. 1983. Equilibrium explanation. *Philosophical Studies* 43, 201–10.

van Fraassen, B. 1980. *The Scientific Image*. Oxford: Clarendon Press.

Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.

How to cite this article

Reiss, Julian. "explanation." The New Palgrave Dictionary of Economics. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008. The New Palgrave Dictionary of Economics Online. Palgrave Macmillan. 16 November 2009 <http://www.dictionaryofeconomics.com/article?id=pde2008_E000290> doi:10.1057/9780230226203.0534(available via http://dx.doi.org/)