# Philosophy Of Economics

## Is There an Explanation Paradox?

Paradox According to Julian Reiss (2012), all economic models are false—in the sense that they "misrepresent their targets" (49)—in one of the following ways (see Wimsatt, 2007, 101-02): (i) they are of only *local applicability*; (ii) they are *idealizations*, whose conditions of applicability are never realized; (iii) they are *incomplete*, leaving out causally relevant factors; (iv) they *misdescribe* (omit or add) *interactions*; (v) they give "a totally wrong-headed picture of nature", by not only featuring wrong interactions, but also *inexistent entities/properties*. Nonetheless, (many) economic models are explanatory, or so economists say. At the same time, intuitively, only true accounts of target phenomena explain such phenomena.[1] For instance, the causal account of explanation, which is nowadays most popular, requires a true story on the actual causes of phenomena. In sum, we have, for Reiss, three claims—(1) economic models are false; (2) economic models are explanatory; and (3) only true accounts explain—that are individually plausible but jointly inconsistent. This situation Reiss labels the "explanatory paradox".

To illustrate, Reiss considers the Hotelling model (Hotelling, 1929). Hotelling uses the model to derive the so-called principle of minimal differentiation, which says that (under certain conditions) it is rational for producers to make their products as similar as possible. Here is a brief description of the model. Assume that price is not the only determinant of sales; also distance from the vendor matters, as buyers incur transportation costs. Assume buyers have perfectly inelastic demands for some good. Assume they are located on a line segment, and that there are two vendors of that good at two fixed points on the segment. Then, the two vendors (who incur no production costs) can rationally set their price higher to that of the competitor and still not lose all of their customers to the competitor, provided prices are not raised to the point that it becomes profitable for buyers to travel longer distances to the farther vendor. In this way, each vendor can secure all of the customers located on the portion of the line segment opposite to the location of the other vendor, plus some of the customers located between the two vendors, depending on the exact prices and distances.

The most interesting result comes when Hotelling relaxes the assumption that the location of the vendors on the segment is fixed. What would happen if the vendors were allowed to move? Answer: they would tend to get closer to each other to gain more of the customers located in between the two vendors. Eventually, they would occupy the centre of the line segment (but not the exact same point), as that is the stable equilibrium that guarantees maximum profits, namely an equal sharing of the business.

The equilibrium is not socially optimum. If the vendors were located at equal distances from the opposite ends of the segment, they would equally share the business, and in addition customers would have to travel less. However, the equilibrium would not be stable, as it would be profitable for the vendors to move closer to each other, until the occupy the centre of the segment again. Hotelling extrapolates from this result to explain why, for instance, political parties (e.g., democrats and republicans) find it profitable to adopt very similar platforms rather than different platforms, between which voter would find it easier to choose. Also, the model would explain the concentration of business districts in large cities.

Reiss emphasizes how the model involves all

[1] Truth may not always suffice to explanation (as was claimed for instance by Cartwright, who said that "truth does not explain much"). However, it seems uncontroversial that it is at least necessary to it (but see below).

aforementioned kinds of "falsehood". (i) It is intended to only apply to settings where producers can erect quasi-monopolies by differentiating their product from competitors and can set prices in the light of maximising profits. (ii) It assumes that the two producers move along a one-dimensional line (no breadth, slope, etc.). (iii) It ignores factors that may influence the customers' decision (e.g., friendship, ideologies). (iv) It assumes specific functional forms, such as perfectly inelastic demand. (v) It assumes that minimal product differentiation is always the result of the desire to create a spatial monopoly to maximize profits (and never of, say, imitation or chance). Yet, although it involves falsehood, the model "feels" explanatory. But how is that possible? If it were really explanatory, it would be true—which is not, by assumption. Contradiction!

True? To remove the inconsistency, one (or more) of (1) to (3) must be rejected. Reiss reviews attempts to reject (1), (2), and (3) in the literature, starting with arguments to the point that models are true "in the abstract", namely they misrepresent their target in some respects but correctly represent it in other respects. In other words, they do not tell the truth, and yet they tell what would be true in the absence of interferences. Reiss attributes this view to Cartwright and Mäki. In particular, for Mäki a model is "true" when it adequately *isolates* the effect of a factor of interest and *idealizes away* anything else. So, for instance, a Galilean thought experiment studying the law of falling bodies involves imagining a body's rate of fall as subject to the Earth's gravity in isolation from air resistance, other gravitational fields, etc.

Reiss agrees that isolation (and, in consequence, explanation by isolating models) works well in natural domains—but, he asks, does it work in economics, too? No. He sees an important disanalogy between non-economic and economic models. Economic models' assumptions do not eliminate, or assume away, disturbing causal factors (e.g., the influence of geography, of transportation costs, of demand) but rather they specify their functional form (e.g., a specific one-dimensional geography, linear transportation costs, perfectly inelastic demand).

In response, one might insist that "assuming away" in the natural sciences (e.g., assuming away air resistance) works by assigning a specific value—zero—to certain quantities; analogously, economic models assume away by assigning specific parameter values. Still, Reiss sees three differences between the two cases. First, in a Galilean thought experiment, the assumed-away factor does not appear, whereas in a non-Galilean model such as Hotelling's it figures explicitly, as the result would not be deducible without the explicit assumption. Second, Galilean assumptions involve quantitative causal factors, whose causal effect comes in degrees, whereas economic assumptions are categorical (e.g., one-dimensional geography is qualitatively different from a two-dimensional one). Third, the causal factors in Galilean assumptions have a natural zero, contrary to the quantities in economic models (e.g., geography, the functional form of transportation costs). As a result, whereas Galilean thought experiments teach us what some factors do in the presence of the assumed-away factors, such that their results are exportable, non-Galilean assumptions make the result specific to particular situations, and not generalizable.

Many claim that robustness analysis takes care of the latter problem. By varying controversial auxiliary assumptions, one can test whether the result depends on them rather than on the target hypothesis, and thus infer whether they result would still hold in real world situations where auxiliaries are most likely not satisfied. While Reiss agrees that this would be a solution in principle, and that economists do like to present their models together with robustness analyses of their results, he maintains that in practice robustness tests are often impossible, and when they are, they often give negative results, that is, the results are *not* insensitive to controversial auxiliaries. For instance, the Hotelling result is sensitive to the number of vendors (e.g., when the ven-

dors are three, there is no stable equilibrium) and to the exact functional form for transportation costs (e.g., the equilibrium is not stable around the linear form). This is detrimental to the explanation that is supposed to ensue from the model: if the model applies *only in the abstract*, it explains *nothing in reality*.

Not Explanatory? Perhaps economic models aren't explanatory, after all?

Hausman, for instance, thinks that only theories may be true (that is, interpreted models, or models with associated theoretical hypotheses), from which he infers that models (themselves) cannot explain. Models are only useful for "conceptual explorations", namely for exploring the possible consequences of a theory. The problem with this position, for Reiss, is that it shifts the issue from false models *to false theories*. Hausman wouldn't want to deny that economic theories are explanatory. Whence, the question: what make false theories (if not false models) explanatory?

By contrast, Alexandrova claims that models have a heuristic and not an explanatory function. They are "open formulae", in that they contain free variables, and thus are not truth-evaluable. They become truth-evaluable only when the free variables are interpreted so as to get causal hypotheses, which are tested experimentally. The problem with this view, for Reiss, is that it flies in the face of intuition. Economists routinely invoke models to explain real-world phenomena in the absence of any experimental test.

Finally, Grüne-Yanoff maintains that economic models are used to establish possibility or necessity hypothesis. For instance, Schelling's (1969) famous segregation model establishes that outright racism is not necessary to housing segregation. The possibility of housing segregation in the absence of racism does not entail that observed segregation is not the result of racism, and thus the model does not explain actual segregation. However, Reiss argues, economic models' ability to establish modal hypotheses is not incompatible with another ability, namely that of providing explanation. Some economic models do explain, or so it seems.

No Need of Truth? Finally, someone might claim that explanation does not require truth. Who? Arguably not an advocate of a causal account of explanation. Reiss attributes the view to Sugden, who believes that explanation depends on the model's "credibility", that is, on the model's world being "a parallel or counterfactual world that, to a greater or lesser extent, resembles aspects of our own world" (55). To learn about a real-world phenomenon, one needs an inductive inference from the model to the world, analogous to those from one instance of a type to another. For instance, to learn about segregation in Cleveland one may infer from what happens in Baltimore, Philadelphia, New York, etc. Or one may infer from what happens in a model, provided this is an instance of the same type as Baltimore, Philadelphia, New York, etc.—that is, provided the model world is *credible*, in the sense that (in Sugden's own words) "it is compatible with what we know, or think we know, about the general laws governing events in the real world" (Sugden, 2009, 18). Sugden is explicit that credibility is not the same as truth. It is closer to "truthlikeness".

From a descriptive point of view, Reiss agrees with Sugden that economists do believe that their models are explanatory because credible. From a normative point of view, however, he contends that credibility *per se* is not a good reason to believe that they are, in fact, explanatory. What makes the model credible are considerations that depend on the scientist's values and education, on

[2] It has also been argued that credibility alone is not sufficient for explanation. For instance, for Grüne-Yanoff (2009) it is also necessary to show that the differences between model world and real world do "not give reason to judge as incredible in the real-world situation what was judged as credible in the model" (91). Moreover, one cannot appeal to the truth of counterfactual conditionals of the form "if the initial and boundary conditions of the model were the case, then the results would be the case" to lend credibility to the model, because their truth evaluation *must follow from* the model's

social norms, on theoretical preferences, etc. and not necessarily on explanatory facts.[2] What facts, exactly? To elaborate his claim, Reiss invokes the most popular alternative to the causal account of explanation, namely the unificationist account. In the unificationist account, Reiss claims, explanation does not depend on truth. If economic models explained in the unificationist sense, they would not need to be true. However, they don't. And if economic models don't explain by unification, it doesn't follow that the explanation can dispense with truth.

More precisely, the argument goes as follows. Economic models are like argument patterns in the unificationist account. The more conclusions can be derived from using them, the more stringent they are; and the fewer patterns are necessary to derive the conclusions, the greater their unifying power. The idea seems endorsed by Friedman, who claims that the virtues of a theory are its simplicity (few assumptions) and fruitfulness (large scope) (1953, 10). This, in turn, would make sense of the economists' use of mathematical models, of rational choice theory, and of equilibrium concepts—all of which would form the unifying argument patterns, which grant economic models their explanatory power. Reiss finds the view that unifying accounts are explanatory defensible. However, he claims that "the argument patterns economics tends to produce are at best spuriously unifying" (58). The reason is their lack of stringency. In principle, the more arguments a pattern disallows to be recovered from it, the more stringent it is. The argument patterns of economics, however, are not stringent, in that they disallow very little. In other words, they have *little empirical content*, in a Popperian sense. Utility maximization, for instance, imposes few if any restrictions—utility is often assumed to derive from material gain, but non-material gains can in principle be relevant, too. Analogously, profit maximization is often interpreted in a material sense, but it need not be. For instance, when Hotelling's model is applied to politics, profit maximization is read as "votes maximization". Not even the notion of an economic agent imposes serious restriction: not only human agents but also animals or sub- or super- personal entities may count as economic agents. (To wit, Ross has recently invoked a non-anthropocentric view of economic agents to avoid time-inconsistent preferences.) Nor does the rule "Solve the model using an equilibrium concept" add much empirical content. There are many equilibrium concepts, and relative to each one of them, the range of arguments that can be generated from it is still vast. So, economic models do not explain by unification. Therefore, it does not follow that they don't need to be true in order to explain. So, Reiss concludes, "[w]hatever economists think when they say they provide explanations of this or that phenomenon, the accounts they give are not explanatory qua the unifying power of the argument patterns from which they are derived" (59).

And if none of the three claims generating the inconsistency can be rejected, the paradox is genuine. Reiss' article has sparkled a hot debate (see, e.g., the contributions to "Symposium on the Explanation Paradox", *Journal of Economic Methodology*, 20(3), 2013). The jury on whether the paradox is genuine is still out.

credibility, and thus it cannot be used to justify it (92).

## References


Friedman, M. (1953). The Methodology of Positive Economics. In Hausman, D., editor, *The Philosophy of Economics: An Anthology*, pages 145–178. Cambridge University Press, 2008.

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, 70(1):81–99.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, 39(153):41–57.

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, 19(1):43–62.

Schelling, T. C. (1969). Models of segregation. *American Economic Review*, 59(2):488–93.

Sugden, R. (2009). Credible Worlds, Capacities and Mechanisms. *Erkenntnis*, 70(1):3–27.

Wimsatt, W. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.