Fall 12-19-2020

Understanding Without Explanation? A Philosophical Inquiry Into the Explanation Paradox of Economic Models

Eric Haupt

Understanding Without Explanation? A Philosophical Inquiry Into the Explanation Paradox of Economic Models

Eric Haupt

December 2020

Senior thesis submitted in partial fulfillment
of the requirements for a
Bachelor of Arts degree in Economics
at the University of Puget Sound

Author: Eric Haupt

Title: Understanding Without Explanation? A Philosophical Inquiry Into the Explanation Paradox of Economic Models

Keywords: Idealization, Abstraction, Models, Explanation, Economic Methodology, Philosophy, Instrumentalism, Positivism,

Abstract

Conventional economic modelling is frequently criticized for being "unrealistic" due to the variety of unrealistic assumptions that underpin many models. Critics frequently wonder how models that are unrealistic or false can accurately explain economic phenomena. This criticism and the problem it presents for economics is captured by Julian Reiss's "explanation paradox." This paper aims to evaluate Reiss's paradox and assess the problems it poses for economics as a positive science. To address this problem, I survey a variety of competing strategies offered by philosophers and economists before critically evaluating the validity of the paradox. I conclude that while economic models cannot be considered explanatory because they do not completely and accurately identify the causes of economic phenomena, they still offer understanding that is valuable for informing economic policy.

Introduction

In economic theory, a "model" is an abstract, visual representation of a theoretical relationship between multiple economic variables such as price and quantity, demand for goods x and y, inflation and unemployment, and so on. Popular criticisms of conventional economic theory often take aim at the characteristics of economic modelling which critics deem "unrealistic" or "false," meaning that they are not descriptively accurate representations of economic phenomena as they exist in the real world. If correct, this criticism appears to spell trouble for the epistemic value of economic modelling. Intuition tells us that if a description of a particular phenomenon is inaccurate, it cannot be the case that the description offers us an accurate explanation or understanding of what that phenomenon is, how it works, or how it could be used for practical purposes. The intuition that this criticism touches on is best captured by the "Explanation Paradox," an idea coined by the German philosopher and economist Julian Reiss.

Reiss's paradox can be summarized by the combination of the following three contradictory premises:

(1) Economic models are false.

(2) Economic models are nevertheless explanatory

(3) Only true accounts can explain[1]

If sound, this paradox demonstrates that economic models are doomed to fail to explain economic phenomena. The problems this poses for economics as an explanatory science are clear, but it must also be noted that the explanatory failure this paradox alleges carries considerable real-world consequences if the paradox is valid. Public policy is heavily influenced by prevailing economic expertise, which is in turn influenced by economic modelling.[2] If these models are neither true nor explanatory, it seems unreasonable to expect them to be effective in bringing about their intended outcomes. Addressing the paradox's validity thus carries importance both for resolving economics's methodological soundness and for guiding the appropriateness of conventional economic modelling's use in guiding public policy.

This paper is presented in four parts: first, we will elaborate on conventional criticisms of the completeness and accuracy of economic models and the problems they present for modelling's explanatory power. Once this is done, we will survey a variety of competing strategies various economists and philosophers have offered for either resolving the paradox or addressing the apparent problems it poses for economic methodology. With those scholars' insights in mind, we will then re-examine and attempt to resolve the paradox through a critical examination of its second and third points, "economic models are nevertheless explanatory" and

---

[1]Julian Reiss, "The Explanation Paradox.*" Journal of Economic Methodology, Vol. 19, No. 1, March 2012,* 43–62

[2] Sawhill, Isabel. "We Need to Rethink our Economic Assumptions." Brookings. The Brookings Institution, last modified May 7

"only true accounts can explain." Finally, we will conclude by considering our findings' relevance for policy.

The Falseness of Economic Models

When a critic of economics describes an economic model as 'false' or 'unrealistic,' what characteristics of the model is she referring to, specifically? The answer largely depends on who the critic is and which model she is referring to, but most criticisms, especially those directed at a model's explanatory power rather than its practical use, can be sorted into two categories: those that take aim at a model's *abstractness* and those that take aim at its *idealization*.[3] If a model carries these traits in abundance, it fails to be an accurate reflection of the world it describes. A model is abstract if it is deliberately mathematized or otherwise formalized to a degree to which it omits "certain properties or factors" and becomes only a true picture of an abstract world, but not of the real world which it purports to explain.[4] A model is idealized if it "exaggerates or distorts certain properties or factors" in its representation of the system it purports to explain.[5] Models often acquire this property through their use of assumptions to preserve their formal coherence. It is not difficult for one who has studied economics to name the ubiquitous unrealistic assumptions found in economic models. For instance, a model of consumer choice assumes that consumers possess perfect information about the relevant characteristics of what they want to buy and that they have complete, monotonic, transitive preferences between the various goods available to them. Similarly, a model of a perfectly competitive market assumes a large number of participants, homogeneous products, perfect access to information among

---

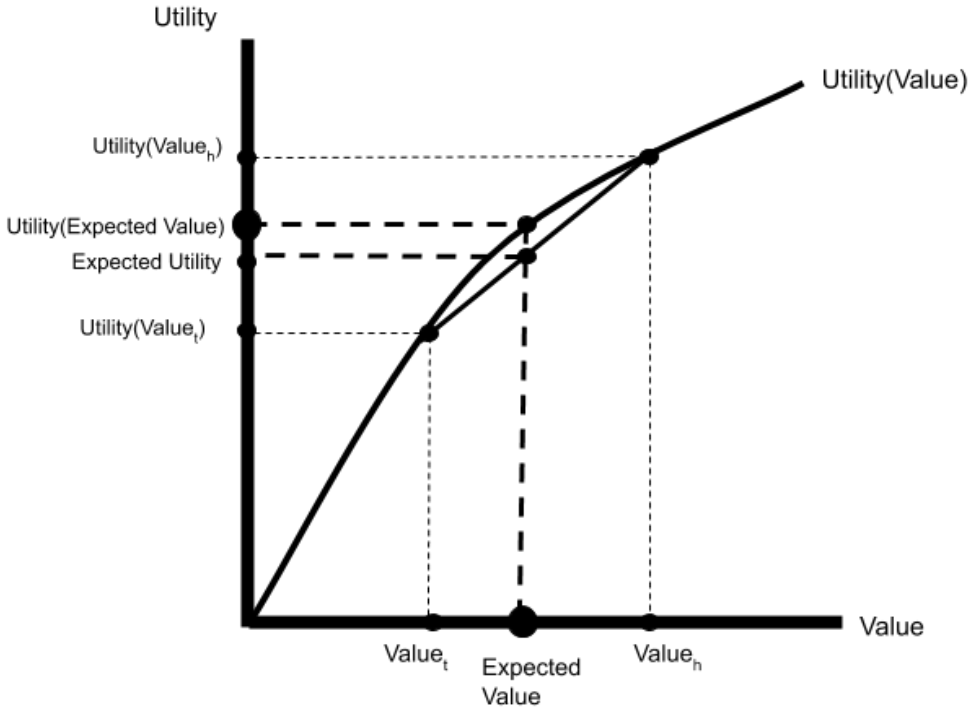
[3] D. Wade Hands, "Hypothetical Pattern Explanations in Economic Science: Hayek's Explanation of the Principle and Pattern Prediction Meets Contemporary Philosophy of Science." *In Research in the History of Economic Thought and Methodology.* Published online: 04 May 2018; 37-56.

[4] Reiss, 50

[5] Reiss, 47

consumers and sellers, and no barriers to entry into the market. It is by no means impossible to find real-world cases that satisfy assumptions like these, but problems with technical accuracy arise when models incorporating these assumptions are used to analyze cases where they do not apply perfectly, or when case-specific operating factors are not adequately accounted for.[6]

Consider the example of a model of rational consumer choice under conditions of risk for a risk averse individual:



This model diagrams a scenario where an individual, *x*, is considering making an expenditure under a condition of risk, such as purchasing a lottery ticket. The x-axis represents the expected value *x* will receive from the expenditure, which is calculated by multiplying the probability of a certain return by the monetary value of that return. The y-axis represents the personal utility a person receives from undertaking that risky expenditure. The *Utility(Value)* function represents the relationship between these two variables for the individual, *x*. We say that *x* is risk averse

[6] Rodrik, Dani. 2015. *Economics Rules*. 1. ed. ed. Oxford: Oxford Univ. Press, 96

because this model shows *x* gaining more utility from the expected value of this expenditure than the amount of utility they expect to receive after undergoing the expenditure. In other words, their risk preferences are such that they would prefer to receive the amount of money that this risky expense is "worth" than to run the risk of receiving less than its expected value in exchange for a chance of receiving greater than its expected value.

What is wrong with representing *x*'s choice situation this way? Like with many economic models, this representation is idealized and abstracted to a point where the story it tells about one's decision-making behavior is questionable. Its abstract qualities are clear: a well-behaved function for *Utility(value)* requires a complete account of this individual's preferences among different levels of risk and an assumption of perfectly rational preferences between those levels of risk. Additionally, a real person's utility function−if one could be constructed−would almost certainly not carry the elegant convexity or monotonicity of those seen in microeconomics textbooks. The model is heavily idealized as well. Factors that would influence an individual's risk preferences at a given moment or even render them irrational, such as mood or influence by advertisement, are not accounted for. Additionally, the notion of "expected value," how it's calculated, and how it's applied in judging an individual as "risk averse" or "risk loving" is strikingly normative, and it implies the obviously false premise that an individual either consciously or subconsciously compares the mathematical concepts of expected utility and expected value when considering a risky choice.[7] None of this is to say that this model is useless. In particular, an estimation of a risk preference function computed using values collected from a real person's decisions under conditions of risk would be helpful for making inferences about what that individual might do in choice situations not yet observed. However, the model's

[7]Hausman, Daniel M. 2007. The Philosophy of Economics. 3. ed. ed. Cambridge: Cambridge University, 79

heavily abstracted representation of the phenomenon it describes and its omission of crucial causal factors appears to severely limit its use in offering an accurate explanatory account of that phenomenon.

The failure of models such as this one to adequately capture all causal mechanisms operating in the systems which they hope to describe presents problems for their use in explaining those systems, especially in novel cases when such systems have operative parts which are unique to their own case and were not anticipated by the models corresponding to the kind of phenomenon that it is. That being said, discerning a model's success or failure at "explaining" something requires an established definition of what "explanation" means. The account of explanation discussed by Julian Reiss is one that principally deals with causes. Under this definition, an explanation is successful if and only if it identifies the causal mechanisms at play in a given phenomenon: "To explain a specific economic event isto cite its causes; to explain a general economic phenomenon is to describe the causal mechanism responsible for it."[8] Reiss goes on to note that this particular account of explanation presents a unique challenge for economics relative to other sciences:

> The requirement that causal accounts be true to be explanatory is in fact the great downside of causal explanation. When phenomena are complex, and economic phenomena are, truth is hard to come by. Accounts given of economic phenomena are usually dramatically simplified and features we know affect a result are represented in a systematically distorted way. Among economists, the slogan 'all models are wrong, but some are useful' is well-known.[9]

[8] Reiss, 43
[9] Reiss, 44

The question is thus raised if it's possible for economic models to offer explanatory power, and thus understanding, despite their systemic distortion, abstractness, and idealization.

Competing Strategies

Many economists' and philosophers' answer to the logical tension articulated in the explanation paradox proceeds by challenging the definition of explanation that it adopts: a model explains if and only if it gives a complete and accurate *causal* account of the phenomenon it targets. Suppose that it is not the case that for a model to provide explanation of a phenomenon, it must provide a complete and accurate description of that phenomenon's causes. If this is true, then it is not necessarily the case that a model fails to be explanatory merely because its idealized or abstract characteristics deprive it of insight into a phenomenon's causes.[10] Arguments such as these are a promising direction for resolving Reiss's explanation paradox.

In his famous essay, *The Methodology of Positive Economics*, Milton Friedman took this line of defense a step further by arguing that not only was a model's complete accuracy not necessary for it to be sufficiently explanatory, but that criticisms aimed at economics's level of "realism" are wholly irrelevant to the discipline's pursuit of developing a robust positive science.[11] According to Friedman, the true measure of a model's scientific power is its usefulness in generating accurate predictions about phenomena not yet observed. Indeed, a model's ability to produce these predictions is sufficient to call it explanatory, even if−and curiously−*especially if* it is descriptively false in an appropriate way:

> A hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone.

[10] Reiss, 50

[11] Milton Friedman, *"The Methodology of Positive Economics"* In Essays In Positive Economics (Chicago: Univ. of Chicago Press, 1966)

> To be important, therefore, a hypothesis must be descriptively false in its assumptions; it takes account of, and accounts for, none of the many other attendant circumstances, since its very success shows them to be irrelevant for the phenomena to be explained.[12]

It is worth noting that Friedman's tone suggests he is not convinced of the appropriateness of an "explanatory" standard for a model's importance in the first place. Still, his language in the above excerpt suggests that he regards a model's possession of predictive power as being *sufficiently* explanatory for the purposes of a positive science, even if that explanation is incomplete by the standards offered by other scholars such as Reiss and those Friedman responds to throughout his essay. The choice-under-uncertainty model we discussed earlier would qualify as explanatory by these standards despite its abstract nature and silence on the neurobiological factors that actually cause one's risk-taking behavior on a physical level. This is a dramatically different account of explanation from the causal account, and notably, if true, it dissolves the tension between the statements "economic models are false" and "economic models are nevertheless explanatory" by positing that a model's truth is largely irrelevant to the matter of its explanatory power.

I grant Friedman's argument on his own instrumentalist terms. In cases where it can be shown that an economic model yields sufficiently accurate predictions about phenomena not yet observed with a satisfactory degree of consistency, the model satisfies its purpose in the creation of a "filing system for organizing empirical material and facilitating our understanding of it."[13] However, verification of a model through examination of only its predictive power is unsatisfying for philosophers of science who follow stricter standards of explanation.

---

[12] Friedman, 15
[13] Friedman, 148

Collin Rice is one such contemporary philosopher who−contrary to Friedman−has called into question the appropriateness of treating the inaccurate parts of scientific models as appropriate "distortions" that do not compromise the accurate and important elements of the story that the model tells.[14] Rice calls this the "decompositional" strategy, which typically carries three common assumptions in application:

1. Target Decomposition Assumption: The real-world system is decomposable such that the contributions of the features that are relevant to the occurrence of the target phenomenon can be isolated from the contributions of features that are irrelevant (or are largely insignificant) to the target phenomenon.
2. Model Decomposition Assumption: The scientific model is decomposable such that the contributions of its accurate parts can be isolated from the contributions of its inaccurate (that is, idealized or abstracted) parts.
3. Mapping Assumption: When successful, the accurate parts of the model can be mapped onto the relevant parts of the real-world system and the inaccurate parts of the model only distort the irrelevant parts of the real-world system.[15]

In some form or another, assumptions like these are present in virtually all idealized models of the sort that Friedman himself embraces in his essay.[16] With the highly idealized ideal gas law in chemistry used as an example, Rice calls into question the supposed irrelevance of disturbing forces not included in the law's mathematical formulation: "PV = nRT where P is pressure, V is volume, T is temperature, n is the number of moles of gas, and R is the constant."[17] Since virtually no gas's activity is influenced exclusively by these variables, it's nonsense to suggest

[14] Rice, Collin. 2019. "Models Don't Decompose that Way: A Holistic View of Idealized Models." *The British Journal for the Philosophy of Science* 70 (1): 179
[15] Rice, 182
[16] Friedman, 34
[17] Rice, 190

that they are the only relevant factors at work. Any kind of decompositional strategy that treats the included variables as important and unincluded variables as irrelevant must therefore be misguided. A similar rebuke of economic models that makes similar attempts to decompose the variables at play would also follow along these lines, much in contrast to Friedman's defense. Rice instead argues that the value in these idealizations comes from their utility in aiding mathematical analyses, which often facilitate researchers' ability to make insights and inferences about the phenomenon being studied to an extent not possible if they were perfectly complete, accurate, and undistorted. This argument has notable merit. Models like the choice-under-uncertainty model enable the kind of mathematical analysis that Rice discusses by, for instance, enabling economists to estimate risk-utility functions based on incomplete information. This estimation could then be used to produce credible predictions about one's behavior in the future, a la Friedman's instrumentalist defense.

Rice's and Friedman's approaches differ in their attitudes about the acceptability or even necessity of unrealistic assumptions as vehicles for understanding, but they appear compatible with each other with regard to the instrumental value of models, and neither rely on a complete causal account of the models they describe. However, one neither needs to fully abandon a causation-based account of understanding nor turn to strict instrumentalism to defend the explanatory power of admittedly idealised and descriptively false economic models. Philosophers of science such as Reutlinger, Hangleiter, and Hartmann have proposed a "toy models" approach which accepts the falseness of models, but takes them to be useful vehicles for "grasping" an epistemically accessible, necessarily simplified understanding of otherwise inaccessibly complex phenomena.[18] Toy models allow one to "grasp" such a phenomenon if and

[18] Alexander Reutlinger, Dominik Hangleiter, and Stephan Hartmann, *Understanding (with) Toy Models,* 1070

only if the model enables one to "to visualize the behaviour of the target system of a scientific toy model or to have a 'mental model' of the toy model and its solutions."[19] Toy models are then divided into two categories: "autonomous" models which provide general, how-possibly explanations of phenomena which have some factor of interest in common, and "embedded" models which provide how-actually explanations of specific cases. How-possibly explanations offer a conceivable way that a phenomenon could occur, and how-actually explanations outline how a specific phenomenon actually does occur. To return to the example of individual choice under uncertainty, the elegantly convex function of risk preference we reviewed suits Reutlinger, Hangleiter, and Hartmann's category of autonomous models. If we developed a perfectly accurate utility function for a specific individual which accounts for all relevant variables left out in the standard construction of choice under uncertainty models, that function would represent an embedded, how-actually model. The general models featured in standard economic pedagogy almost exclusively fall into the category of autonomous, how-possibly models.

What is the value of autonomous toy models if they do not provide how-actually explanation? The answer lies chiefly in their utility in providing us with that sense of "grasping" the phenomenon being described. "There are central examples of autonomous models that are best interpreted as providing how-possibly explanations and how-possibly understanding. This sort of understanding is valuable because it has (what we call) important modal, heuristic, and pedagogical functions in scientific research and science education."[20] It is worth noting that Reutlinger, Hangleiter, and Hartmann's work is not tailored to economics in particular, so it is not immediately clear why the notion of autonomous models is helpful in addressing economic modelling's failure to offer a causal account of understanding or why it is meaningfully different

[19] Reutlinger, Hangleiter, and Hartmann, 1085
[20] Reutlinger, Hangleiter, and Hartmann, 1087

from the instrumentalist approaches discussed earlier. Toy models in other disciplines, such as the ideal gas law discussed by Rice, are not thought of as failing to identify causes.

To see how the autonomous/embedded model distinction can help address economic modelling's confusion over causes, consider the commonplace assumption of the rationality of economic agents. While the rationality assumption is descriptively false for most if not all individuals, it is an intuitively appealing, "graspable" norm in economics pedagogy that helps illustrate the dynamics at play in consumer choice. When introducing a concept such as choice under uncertainty, it makes sense to begin by introducing a theoretical risk-taking agent whose choices make the most rational sense from an economic efficiency standpoint. For a rational person, a comparison between *expected utility* and *utility(expected value)* would be a descriptively accurate account of the cause of that person's behavior. Irrational behavior from real individuals can be thought of as deviations from that norm, similar to how choice and judgment errors are conceptualized in behavioral economics. Still, any approach that restricts conventional economic modelling to only how-possibly explanations tacitly admits that all models but the most granularly, even impossibly specialized ones cannot offer descriptively accurate causal accounts. Even if they could, those causes would be expressed in the abstract imagery and formal mathematics of economic models instead of in terms of "deeper" causes of economic behavior, like neurological activity.

At this point, one may take away from the literature that it is futile to look to economic models for a causal account of understanding when it is clear that they do not offer one. The more promising inquiry lies in searching for an account of understanding that is not completely and accurately causal, but not entirely instrumental or pedagogical either. But is this a reasonable goal? Verrault-Julien describes philosophers' tendency to clutch to a complete and accurate

causal account of understanding in terms of two conventional epistemological tenets: that one possesses understanding of a phenomenon if and only if "1) one has knowledge of causes and 2) that knowledge is provided by an explanation."[21] These tenets jointly make up what Verrault-Julien calls the "Narrow Knowledge Account of Understanding." The "narrow" description of the account is appropriate; for an abstract, idealized economic model, meeting these two criteria seems to be an almost insurmountable challenge. However, Verrault-Julien provides good reasons to suspect that this account of understanding is untenable:

> First, the literature on non-causal explanations provides good reasons to believe that causal knowledge is not necessary for understanding. Second, as I have argued, there are cases of theoretical modelling that do not provide explanations and yet, according to practitioners and philosophers, afford understanding. This indicates that having an explanation is also not necessary for understanding.[22]

Theoretical modelling that does not give a full account of causes is prevalent in the natural sciences, but is still thought of as offering explanation. Verrault-Julien lists notable examples of these, namely phlogiston theory in chemistry, which is thought to explain combustion, and the "Hawk-Dove' model in evolutionary biology, which is thought to explain evolution under pressures of animal competition.[23] These are both how-possibly explanations because if the conditions outlined in the situations they describe were satisfied, it would be conceivable for them to result in the target systems we observe in the real world. However, Verrault-Julien notes that both, while carrying valid internal logic and offering coherent, possible explanations, are still clearly inaccurate representations of the target phenomenon that they hope to explain when additional empirical facts about that phenomenon are taken into account. Nevertheless,

[21] Phillipe Verreault-Julien, *Understanding Does Not Depend on (Causal) Explanation*, 1
[22] Verrault-Julien, 9
[23] Verrault-Julien, 7

philosophers and scientists think of them as affording explanation, despite clearly failing to meet the criteria necessary for the narrow knowledge account of explanation. How can this be?

Verrault-Julien contends that this is because the narrow knowledge account of understanding is too restrictive of an account of the necessary conditions to obtain understanding. He then supplements this account with another account of understanding: the broad knowledge account of understanding. This account is only slightly different from the one offered by Reutliner et al. According to the broad knowledge account of understanding, the relationship between an explanation and its target system affords understanding if and only if the following conditions hold:

- *Possibility condition*: The generalizations G1,...,Gm, the auxiliary statements S1,...,Sn, or the explanandum statement E are (im)possible according to the relevant modal interpretation and epistemic goal.
- *Implication condition*: Generalizations G1,...,Gm with the auxiliary statements S1,...,Sn logically entail E or a conditional probability P (E|S1,...,Sn).
- *Dependency condition:* G1,...,Gm support at least one counterfactual between S1,...,Sn and E.[24]

Put another way, all the conditions contained in a model must be internally, logically possible and also conceivably possible within the physical world, the conditions must imply the outcome-phenomenon the model hopes to describe, and there must be a conceivable case where that outcome-phenomenon does not take place. So, for instance, our choice-under-uncertainty model meets all of these conditions because the risk preference function in it is logically and physically possible, a rational person would act on the information captured in the model in the

[24] Verrault-Julien, 12

way it says he will, and one can easily imagine ways to falsify the dynamics the model describes in a Popperian sense.

This account differs from the narrow account in two ways: 1) It asserts that causal knowledge is not necessary and 2) It asserts that having an explanation is not necessary.[25] This is a very accommodating standard for understanding. Not only does it permit understanding from how-possibly explanations like the Hawk-Dove model, but it also provides an apparent opening for understanding from highly idealized and abstract economic models, even if they cannot provide explanations through analysis of causes. With the insights of Verrault-Julien and others in mind, we are prepared to re-examine the explanation paradox and judge the extent of the problem it poses for economic modelling's ability to explain, afford understanding, and inform policy.

The Paradox Reconsidered

Verrault-Julien's broad knowledge account of understanding offers a framework for descriptively false economic models to offer understanding, but not necessarily full causal explanation through how-possibly explanation. Let us reconsider each of the explanation paradox's three premises with Verrault-Julien's account of understanding in mind, particularly with respect to what it does and does not say we can gain from abstract, idealized modeling.

(1) Economic models are false.

If we consider economic models to be abstract and idealized how-possibly explanations, as suggested by Rice, Verrault-Julien, and Reutlinger et al, then it follows that they are false in the sense that they are not descriptively accurate or comprehensive of all factors at play in the vast majority of cases. Where models may be true is in their identification of general relationships between some number of variables, which are most easily demonstrated when other

[25] Verrault-Julien, 10

disturbing forces are omitted for the sake of clarity. Ergo, the model demonstrates a true relationship in an abstract, and therefore untrue world. This demonstration can carry valuable heuristic and practical purposes, but there is a difference between stating that something could be true and stating that something is, in fact, true.

Moreover, microeconomic modelling does not only omit or distort causes, it often substitutes them for false ones as well. Consider the fundamental principle of rational choice theory: rationality. The rationality assumption is essential for models to be mathematically coherent, but constructing a choice space for a rational individual relies on examining only their stated preferences and intentions as explanations for their actions. What is wrong with doing this? Rosenberg argued that this sort of thinking is a principle of outdated and incorrect folk psychology.[26] Consider a jogger who, when asked why she voluntarily incurs pain and exhaustion every morning to jog for 45 minutes, answers that she is doing so in order to stay healthy. She may well be answering honestly, and an economist would take her response to be a statement of what she is willing to pay in exchange for the benefits of improved physical health. This is what Rosenburg would say risks veering into folk psychology. The jogger's stated or even believed reasons for jogging may not be the real explanation. Perhaps the real reason she jogs is because her body produces endorphins while she exercises, and her addiction to them compels her to exercise each morning. A psychological pressure such as this can't be built into a model based on rational choice. In the jogger's case, taking her rational calculus between the expected costs and benefits of jogging to be explanatory would not just be a distortion of the causes of her actions, it would be a substitution of a true cause for a false one. Because economic

[26] Alexander Rosenberg, *Philosophy of Social Science,* second edition. Boulder, CO: Westview Press, 1995.

models are so idealized, abstract, and tell descriptively false stories about why people do what they do, I grant the premise that economic models are false.

(2) Economic models are nevertheless explanatory

"Explanatory" is the key word here. Despite intuitions to the contrary, there are good reasons to doubt that economic models can be considered explanatory by any conventional standard.[27] Reiss discusses this at length in his breakdown of the explanation paradox. Causal accounts of explanation are by far the most popular in the philosophy of science, but as we have already discussed, these tend to be "idealized away" in economic modelling to the point where they cannot hope to match the circumstances of any empirical situation which we hope to explain. While this account of explanation is the most important to address because of its popularity, Reiss also rejects accounts that follow Philip Kitcher's unificationist theory and notions of mathematical explanation. In both cases, Reiss contends that economic "explanations" are not empirically vindicated to the degree that these other accounts require. [28]

Recall that Verrault-Julien's broad knowledge account of understanding permits certain models to be used as vehicles for understanding even if they do not offer a causal explanation. A how-possibly economic model therefore can offer understanding regardless of if it is explanatory. However, stating that a model offers understanding merely because it tells us something about what could, but not necessarily does happen may strike some as so minor a concession that it is epistemologically redundant and pointless. This is not the case. To illustrate this, consider the famous Schelling checkerboard model of racial segregation, which is a how-possibly explanation.[29] The basic premise of Schelling's model is that different sets of unlike individuals who are distributed into different locations across an area represented by a

[27] Anna Alexandrova, *It's just a feeling: why economic models do not explain,* 1
[28] Reiss, 52
[29] Reiss 54

grid will relocate themselves to satisfy their preference for a certain minimum number of neighbors who are like themselves. So long as every individual's satisfaction with his location requires that there is some minimum number of neighbors who share a given characteristic with him (typically race), the opportunity for relocation in this community will result in the community's voluntary self-segregation. Is this a verifiable, causal explanation of why racial segregation continues to persist in officially desegregated communities? Of course not. To go even further, there is little reason to believe that Schelling's model explains any particular case of racial segregation. However, this model does produce the meaningful and important insight that it is possible for individuals to segregate themselves even when they are neither compelled by a third party nor have any explicit desire to do so. This is the sense in which a how-possibly explanation can improve our understanding of a phenomenon even without identifying its actual causes. Similar types of value can come from economic models, but this value is not sufficient for claiming we can accurately explain real-world phenomena. So while I reject the premise that "economic models are nevertheless explanatory," there are avenues for them to provide meaningful understanding. We will return to this idea later.

(3) Only true accounts can explain

Evaluation of this claim depends on which standard of explanation we adopt. To some degree, discussion of this point is moot since we have already reviewed Reiss's reasoning for why economic models do not offer how-actually explanations to begin with. Still, many economists seem to accept this point when justifying their use of models which they know to be false to achieve understanding. The instrumental approach advocated for by Friedman is the most famous defense of this. Another interesting example is Robert Sugden's "credible worlds" account, which argues that economic models are credible representations of the world as it could

be in accordance with a set of observed regularities.[30] The difference between this account and a general how-possibly explanation is that Sugden describes the relationships in economic models as credible *candidates* for truth based on observed empirical tendencies rather than mere possibilities for how truth could play out based on formal rules and assumptions. A credible world, he argues, overcomes a merely possible world's limitations for offering explanation.

Still, one cannot help but feel that these so-called 'credible' worlds are really just possible worlds dressed up in more confidence and intuitive appeal. A wide variety of subjective influences and judgments affect what happens to strike one as a credible account of some phenomenon. Furthermore, wouldn't anyone who proposes a how-possibly explanation base their hypothesis on some observed tendency or credible intuition of theirs? Clearly, additional criteria for evaluation of these worlds is needed.

We are left with the unavoidable conclusion that the necessary additional criterion for evaluating a possible world's explanatory power is the validation of its truth through rigorous empirical vindication and accurate identification of the causal mechanisms at play. Perhaps more importantly, maintaining that only a true account of a phenomenon can explain it is tantamount to common sense. If one's explanation of a phenomenon is simply not true, can we really say that they explained it?

Thus, the only claim of the explanation paradox we have rejected is its second premise: "economic models are nevertheless explanatory." This resolves the paradox, but it also appears to spell certain doom for the usefulness of economic modeling. This is not the case. There is still value to be found in accounts of understanding that can accommodate false economic models.

Imagine that the explanation paradox has a cousin, "the understanding paradox," which goes like this:

---

[30] Sugden, 18

(1) Economic models are false.

(2) Economic models nevertheless *afford understanding*

(3) Only true accounts can *afford understanding*

Is this paradox valid? We already evaluated the first premise, but the second and third hold up differently when we consider understanding instead of explanation. Per Verrault-Julien's broad knowledge account of understanding, we do not necessarily require identification of causes or how-actually explanation in order to achieve understanding. How-possibly explanation can still improve our understanding of a phenomenon through close logical analysis, just like we saw with the example of Schelling's checkerboard model of segregation. If our goal is to achieve understanding instead of explanation, considering "credible worlds," autonomous models, toy models, and other possible explanations becomes appropriate and fruitful. The third premise of this new "paradox," "only true accounts can afford understanding," is thus rejected. If the third premise is rejected, the falseness of economic modeling is not a barrier for it to afford understanding. Thus, the second premise, "economic models nevertheless afford understanding," is not challenged by economic modelling's systemic idealization, abstraction, and distortion in the same way that economic modelling's capacity to explain is.

We are left with the conclusion that economic models are abstract vehicles to achieve understanding, but not explanation of a phenomenon through close, mathematical analysis of postulated relationships between variables of interest.

On Policy

The implications of this analysis for policy are interesting, but fairly limited. A utilitarian approach to public policy does not require causal explanation of why its tools work (although that is helpful if it's available); it's chiefly concerned with achieving a satisfactory level of

confidence that they do work. This is where it may be helpful to reconsider Friedman's perspective on the methodology of positive economics. Friedman's account of the sufficient conditions for explanation is not compatible with Verrault-Julien's or Reiss's, but perhaps his instrumental approach is sufficiently explanatory as far as policymakers need be concerned. Recall that in his view, the purpose of economic study is not to develop an accurate record of the causes of economic phenomena, but rather to "serve as a filing system for organizing empirical material and facilitating our understanding of it."[31] This filing system's success is evaluated by its use in generating predictions about phenomena not yet observed. This style of positive economics may be all that's required to properly inform normative economics, i.e. policy opinions based on empirically vindicated economic relationships.

But this does not resolve the issue. Economic modelling's degree of success at empirically vindicating relationships between variables is not settled. One can think of any number of unexpected recessions or failed fiscal and monetary policy measures that theory suggested would have worked. Historically, economists such as John Stuart Mill have often attributed this to the typically enormous complexity of studied phenomena in the social sciences. But this just loops us back to a problem similar to the one we encountered when judging an abstract model's capacity for explanation. If so many variables are left out of a model that the model can no longer explain *or* predict phenomena for us, what good is it?

I contend that the answer here is not to place one's full faith in any one model when trying to analyze an economic event or craft policy, but rather to consider a wide body of economic models, empirical information, and careful intuition. Aydinonat and Rodrick advocate for a similar strategy when addressing economic modelling's woes with explanation. In their view, the optimal approach to explaining an economic event or fact is to assemble a menu of

[31] Friedman, 148

relevant models that share some factor in common with the target phenomenon.[32] Consideration of such a variety of models, they argue, will allow one to develop a more holistic explanation than they would have if they simply selected one, best model. Aydinonat even recommends something of a playbook for how to do this:

> i. Determine the set of models that are relevant for the explanatory task,
>
> ii. Assemble a list of possible explanations from the menu of possible explanatory factors which are suggested by this set of models,
>
> iii. Empirically verify which of these factors are actually causing the fact or event to be explained,
>
> iv. If available models fail to lead to a satisfactory explanation, look for other relevant models, or build new models to expand the menu of possible explanations, and
>
> v. Repeat the preceding steps until a satisfactory explanation is found.[33]

I do not claim that this avenue to explanation is any more successful than the others we have surveyed. Assembling many models into a wide menu does not fully compensate for the explanatory problems that each of them still possess, especially with regard to false and not merely distorted identifications of causes such as those embedded in rational choice theory. Instead, I bring up Aydinonat's suggestions because they offer a promising set of principles for one to follow to develop the sense of understanding necessary to craft economic policy. If a team of policy makers follows all of Aydinonat's instructions, will they have arrived at a correct explanation of the phenomenon they hope to address? By the various accounts of explanation laid out in this paper, probably not, but such a holistic approach−tempered by empirical scrutiny−will significantly enrich their understanding of it in accordance with Verrault-Julien's

[32] Aydinonat, N. Emrah. 2018. "The Diversity of Models as a Means to Better Explanations in Economics." The Journal of Economic Methodology, 237
[33] Aydinonat, 247

broad knowledge account of understanding. For instance, policy makers who want to raise the minimum wage in an effort to reduce poverty benefit from considering the diverse body of models that outline the possible effects of such a policy. Some models may suggest that poverty will increase because the increased minimum wage will raise unemployment, others may suggest that the measure will reduce poverty by giving workers more money to spend and encouraging greater productivity. With these possibilities considered, the policy makers can then compare the situations outlined in the models to the real-world history reflected in the empirical evidence and judge what is likely to take place if the minimum wage is raised. Mileage will vary since some phenomena are more complex than others, but the rich diversity of economic modelling offers promise for developing the best possible, if still imperfect sense of understanding for crafting the best possible, if still imperfect economic policy.

Conclusion

It is important to note that the debate over economic models' explanatory power can appear one-sided. A survey of the literature gives one the impression that almost anyone who writes about the problems captured by the explanation paradox will conclude that either economic models do not explain, or if they do, they only can with a long list of qualifications and assumptions attached. Still, there is no lack of economists who are content to apply models without getting caught up in methodological quandaries, so those who are confident in economic modelling's explanatory power may be underrepresented in this paper's coverage of the existing literature on the subject. Nevertheless, arguments by Reiss, Rice, and others make convincing cases that economic models do not provide explanation. This should not be taken to mean that economics is doomed to fail, however. Emerging subfields such as behavioral economics, neuroeconomics, and an increasing emphasis on empirical research are worthy supplements to

the insights offered by economic models. As we have discussed, tempering the understanding we can achieve from models with the knowledge we gain from empirical research may provide researchers with the insight necessary to develop effective policy. So while individual economic models cannot provide the complete and accurate explanations that economists want, the entire body of available models still offers promise for developing the understanding they need to do their jobs well.

Works Cited

Alexander Rosenberg, *Philosophy of Social Science,* second edition. Boulder, CO: Westview Press, 1995.

Alexandrova, Anna and Robert Northcott. 2013. "It's just a Feeling: Why Economic Models do Not Explain." The Journal of Economic Methodology 20 (3): 262-267.

Aydinonat, N. Emrah. 2018. "The Diversity of Models as a Means to Better Explanations in Economics." The Journal of Economic Methodology 25 (3): 237-251.

D. Wade Hands, "Hypothetical Pattern Explanations in Economic Science: Hayek's Explanation of the Principle and Pattern Prediction Meets Contemporary Philosophy of Science." *In Research in the History of Economic Thought and Methodology.* Published online: 04 May 2018; 37-56.

Hausman, Daniel M. 2007. The Philosophy of Economics. 3. ed. ed. Cambridge: Cambridge University Press.

Julian Reiss, "The Explanation Paradox." *Journal of Economic Methodology, Vol. 19, No. 1, March 2012, 43–62*

McCrown, Frank. "Schelling's Model of Segregation.", accessed December 14, 2020, http://nifty.stanford.edu/2014/mccown-schelling-model-segregation/.

Milton Friedman, *"The Methodology of Positive Economics"* In Essays In Positive Economics (Chicago: Univ. of Chicago Press, 1966), pp. 3-16, 30-43.

Palacio-Vera, Alfonso. 2019. "'Situational Analysis' and Economics: An Attempt at Clarification." Economics and Philosophy 35 (3): 479-498.

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2018. "Understanding (with) Toy Models." The British Journal for the Philosophy of Science 69 (4): 1069-1099.

Rice, Collin. 2019. "Models Don't Decompose that Way: A Holistic View of Idealized Models." The British Journal for the Philosophy of Science 70 (1): 179-208.

Rodrik, Dani. 2015. *Economics Rules*. 1. ed. ed. Oxford: Oxford Univ. Press, 96

Sawhill, Isabel. "We Need to Rethink our Economic Assumptions." Brookings. The Brookings Institution, last modified May 7,

Verreault-Julien, Philippe. 2019. "Understanding does Not Depend on (Causal) Explanation." European Journal for Philosophy of Science 9 (2): 1-20.