Journal of Economic Methodology, 2013
Vol. 20, No. 3, 262–267, http://dx.doi.org/10.1080/1350178X.2013.828873

It's just a feeling: why economic models do not explain

Anna Alexandrova[a]* and Robert Northcott[b,1]

[a]*Department of History and Philosophy of Science, Cambridge University, Cambridge, UK;*
[b]*Department of Philosophy, Birkbeck College, University of London, London, UK*

(*Received 8 July 2012; accepted 1 May 2013*)

Julian Reiss correctly identified a trilemma about economic models: we cannot maintain that they are false, but nevertheless explain and that only true accounts explain. In this reply we give reasons to reject the second premise – that economic models explain. Intuitions to the contrary should be distrusted.

Keywords: models; explanation; understanding; idealization

Introduction

Julian Reiss has written a masterful paper that brings together the insights of many philosophers of economics of the last three decades (Reiss, 2012). In doing so, he exposes an uncomfortable and important truth – that the consensus view, to the extent it exists, is not just unstable but also contradictory. It tries to serve too many gods, have its cake and eat it too. We fully agree with most of what Reiss says and we hope that his paper reaches a wide audience. The issues he raises extend beyond philosophy of economics and should be of concern to anyone interested in scientific models and their role in explanation.

Reiss poses the following trilemma:

(1) Economic models are false
(2) Economic models are nevertheless explanatory
(3) Only true accounts explain

We will focus here on our main point of disagreement, namely the trilemma's second horn. In particular, we deny that economic models are explanatory.[2] It seems to us that this is much the weakest link in Reiss's argument. We begin by reiterating briefly why economic models are indeed not explanatory, then give reasons why intuitions to the contrary should be distrusted, before exploring why such mistaken intuitions might arise in the first place.

Economic models do not explain

Reiss himself makes clear why explanations offered by economic models do not satisfy the criteria laid down by any current theory of scientific explanation. First, they do not qualify as causal explanations because they are false and therefore do not identify any actual causes. This is the most important claim since, as Reiss rightly notes, causal explanation is by far the most popular candidate for the notion of explanation appropriate to economics. But economic models equally fall foul of other theories of explanation too. An insightful section of Reiss's paper (pp. 56–59) shows just why such models do not explain according to a unificationist theory such as Philip Kitcher's. Meanwhile, not stating laws, or at least

*Corresponding author. Email: aa686@cam.ac.uk

not any that are empirically vindicated in the necessary way, they also clearly do not explain in the deductive-nomological sense.

There is no refuge, either, in the notion of mathematical explanation. Perhaps, for instance, it might be thought that the Hotelling model demonstrates the mathematical reason why two firms locate next to each other, much as statistical mechanics demonstrates the mathematical reason why with overwhelming probability heat will flow from hot air to cold. But, first, the notion of mathematical explanation of physical facts is contentious and is the subject of much current debate.[3] And, second, in any case it is agreed by all that to be considered seriously, mathematical explanations require empirical confirmation of precisely the kind that is typically absent in economic cases.

In one way, therefore, the discussion ends already: according to all relevant philosophical theory, economic models do not count as explanatory. We are therefore licensed, indeed obliged, to reject intuitions to the contrary. Nevertheless, as Reiss observes, in a sense this is just to restate his trilemma rather than to resolve it, for even if pro-explanatory intuitions are mistaken, still they themselves need to be explained away. To this end, we begin by exploring further why they should indeed be distrusted.

Pro-explanatory intuitions are suspect

The 'models are isolations' story long defended by Uskali Maki and 'models state capacities' view of the early Nancy Cartwright is of no help here, as Reiss correctly points out. He states three reasons for why idealizing assumptions in economic models are not sufficiently similar to Galilean idealizations (pp. 51–52): (1) the former are conspicuously present in the model, while the latter are absent in a Galilean thought experiment; (2) Galilean idealizations are quantitative, not categorical and (3) Galilean idealizations have a natural zero. Reiss concludes: 'therefore we do not know where to look for "truth in the model"' (p. 52).

We agree with the conclusion but not the reasons. Why should (1), (2) and (3) a priori preclude an isolation or a capacity interpretation of a model? They do not. Instead, the fundamental problem is that we have no empirical evidence for thinking that the models are successful at isolating capacities. If models did succeed in doing so, we would be able to *do* the things with economic models that we are able to do with Galilean-thought experiments, that is, combine their insights with our knowledge of disturbing factors in a given environment to predict results. An honest look at experimental and design economics, our only opportunities for genuine tests of microeconomic models, reveals that this is not what happens. Whenever model-based causal claims are made, experimentalists quickly find that these claims do not hold under disturbances that were not written into the model. Our own stock example is from auction design – models say that open auctions are supposed to foster better information exchange, leading to more efficient allocation. Do they do that in general? Or at least under any real-world conditions that we actually know about? Maybe. But we know that introducing the smallest unmodelled detail into the set-up, for instance complementarities between different items for sale, unleashes a cascade of interactive effects. Careful mechanism designers do not trust models in the way they would trust genuine Galilean-thought experiments (Alexandrova, 2008; Alexandrova & Northcott, 2009). Nor should they. This is why economic models do not deserve the honorific 'capacity'.

When successful causal economic explanations are achieved, it turns out that it is not economic models that do the explaining but rather good old-fashioned experiment-tested causal hypotheses. Sure, these hypotheses are inspired by models, and models do get credit for this inspiration. But they do not get credit for explanation. Reiss thus correctly classifies our view as denying models an explanatory role. He notes: 'In the context of preparing experiments for policy, models may well serve the heuristic function Alexandrova describes. To be fair, she does not claim more than that.' (p. 54). In fact, though, we are fully prepared to claim more than that. The use of models in mechanism design is a very good test for a philosophical account of models, because it is such a high-stakes case (as well as for other reasons – see below). It is really important to get an auction right. And when it does not go right, its failure is much more obvious and costly than a failure of the Hotelling model to explain, say, the polarization of US politics. When push comes to shove, as it does in mechanism design, models are not treated as explanatory.

As a general matter, the economics profession is known for its 'casual empiricism'. As the name suggests, it involves scoring explanatory victories casually rather than by relying on econometric or experimental tests. Often this involves nothing more than drawing a vague and intuitively appealing analogy between the model and the phenomenon. For example, the famous Prisoners' Dilemma game is often invoked in cases of a price war between, say, two gasoline stations located opposite each other. In cases like this, appeal to vague similarities between the model and the phenomenon is often the beginning and end of the 'explanation'. A more high-profile example of casual empiricism can be found in the press release of the Nobel prize committee, which stated that Thomas Schelling's 'analysis of strategic commitments has *explained* a wide range of phenomena, from the competitive strategies of firms to the delegation of political decision power' (Nobelprize.org 2005, italics added). Yet Schelling's models, for all their importance, have not scored any major predictive or experimental successes.

The auction case, by contrast, is significant as it is a rare one within economics of precisely such success. Rather than casual empiricism, instead a particular intervention demonstrably led to markedly increased revenues plus a range of other benefits. So we should take much more seriously the lessons from the auction case than from any number of casual Hotelling ones.

The details of the Hotelling model, moreover, show well the flimsiness of any understanding or explanation it is alleged to provide. First, its predictions are not borne out fully. For instance, two competing political parties typically do not have identical platforms, nor do competing stores typically locate right next to each other. At best, then, the model is incompletely or partially explanatory. Second, the predictions are typically qualitative rather than quantitative. This makes it hard to assess precisely to what degree the model is explanatory.[4] But perhaps the most important worry, third, is one that Reiss himself discusses extensively (pp. 52–53), namely the lack of robustness of the model's predictions with respect to variation in its assumptions. For instance, the very small tweak of changing from a linear to a quadratic cost function completely reverses the Hotelling model's predictions regarding firm location![5] This lack of robustness is worth dwelling on. Consider the model is alleged to yield us an intuitive understanding of why firms locate close to each other, or of analogous phenomena in other fields such as politics. Yet a minor tweak to an assumption presumably peripheral to that intuition completely reverses the result. This suggests that in fact we have a rather poorer intuitive grasp of what really explains the result than we had thought.

Another issue supports our general scepticism here. Economic models frequently invoke, as Reiss notes, entities that do not exist, such as perfectly rational agents, perfectly inelastic demand functions and so on. As economists often defensively point out, other sciences too invoke non-existent entities, such as the frictionless planes of high school physics. But there is a crucial difference: the false-ontology models of physics and other sciences are empirically constrained. If a physics model leads to successful predictions and interventions, its false ontology can be forgiven, at least for instrumental purposes – but such successful prediction and intervention is necessary for that forgiveness (Northcott, 2013). The idealizations of economic models, by contrast, have not earned their keep in this way. So the problem is not the idealizations in themselves so much as the lack of empirical success they buy us in exchange. As long as this problem remains, claims of explanatory credit will be unwarranted.

Finally, perhaps lessons can be learnt from the many areas of biology which are also characterized by idealized mathematical models. The explanatory status of these models is unclear in just the same way as in economics. One view is that in practice the models in biology serve to structure and inspire subsequent research by providing concepts and ideas, but that they do not themselves tell us what to be realist about nor do they themselves explain (Pincock, 2012, see also work by Jay Odenbaugh and Patrick Forber). Rather, it is only this subsequent research, often featuring close empirical study, that achieves explanations. This picture, of course, is more or less exactly the one we have argued is true for economics as well. Perhaps there can thus be a consilience between these two areas of philosophy of science.

Where do the mistaken intuitions come from?

So far, we have argued that economic models do not explain and that intuitions to the contrary are suspect. This combination implies the need for an error theory. In particular, why might such mistaken intuitions arise in the first place?

Several overlapping possibilities suggest themselves. The first is that the word 'explanation' has many connotations. Most notable here is the longstanding distinction between epistemic and ontic conceptions of it (Salmon, 1984). Very roughly, epistemic views emphasize that explanations reduce our surprise at an outcome, making it more evident to us why that outcome occurred. The ontic view, by contrast, analyses explanations purely in terms of impersonal objective features. Causal explanation is usually taken to be the classic example of the latter kind, as it consists in identifying an effect's cause, i.e. its place in the causal structure of the world, quite independent of any subjective or epistemic aspects. In the case of philosophy of economics, the causal view of explanation predominates, for good reason. It is common ground between Reiss and us that this is as it should be, so we will not defend that predominance here.

The particular error theory is then that because economic models seem to lessen the surprise of an outcome, this is erroneously taken to imply that they explain it. This appears to be a common reaction to the Hotelling model and the issue of firm location, for instance. Surprise being a subjective matter, perhaps the model does indeed lessen it. But such reactions are quite unreliable, as we have seen, as a guide to whether we have achieved any explanation in the causal sense. And if we are committed to causal explanation, this renders irrelevant mere surprise lessening in itself – even if our intuitions have yet to take that on board.[6]

Move on now to a second possible source of mistaken intuitions: the notorious ease with which humans conjure up after-the-fact rationalizations and illusions of success. Whenever we observe something *consistent* with the stylized claims of an idealized model, it is correspondingly all too tempting to leap to the conclusion that that thing is *explained* by the model. Situations that force us to go beyond such treacherous psychological triggers are therefore essential test cases. This is the value of the auction case study, as the need to construct a successful intervention short-circuited all lazy talk about what the theoretical models might be achieving.

Reiss himself discusses the shakiness – because insufficiently constrained – of judgements of a world's credibility (p. 56). Yet it seems to us that exactly his complaints apply equally to judgements of explanatoriness too! Indeed, his reasons for dismissing credibility intuitions as explanatory arguably go a long way towards providing the error theory that we are seeking. Reiss insists that credibility judgements arise out of economists' training and socialization in the discipline. Yet surely so too do judgements of explanation. Consider the evidence that the more game theory one studies the more one begins to see social interactions as games.[7] And once one sees social interactions as games, it is all the more tempting to treat game theory as explanatory.

As Reiss correctly points out, many economic models, Hotelling's included, are widely seen to be explanatory and 'feel' explanatory. But it is worth asking who, apart from economists and those close to the discipline (for example, philosophers and historians of economics), share these feelings? The relevant contrast class here is not laypeople but rather other social scientists who study the same phenomena but with different theoretical tools. Do they also feel the explanatory pull of economic models? Do they feel this pull to the same degree as economists? We do not know.

Our third strand of thinking in this section is to reflect on the origin of explanatory intuitions in general, not just within economics. The empirical study of these feelings by cognitive science is in its infancy. Still, it appears that there exists a kind of 'explanatory phenomenology' – a phrase coined by Alison Gopnik to describe the 'aha' feeling shared by children and scientists alike (Gopnik, 2000). This phenomenology might even constitute a basic emotion that, like other basic emotions, has an evolutionary purpose. Gopnik's view (p. 300) is that 'explanation is to theory-formation as orgasm is to reproduction. It is the phenomenological mark of the fulfilment of an evolutionarily determined drive … we experience orgasms and explanations to ensure that we make babies and theories.'[8] (By 'theories' Gopnik has in mind, roughly, maps of causal dependency relations.) This comparison of the feeling of explanation to orgasm motivates our scepticism about the evidentiary value of these feelings. We know better than to look for orgasm to make sure that reproduction happened. Similarly, we should know better than to look for 'aha' feelings to make sure that actual explanation happened. Especially in theoretical economics.[9]

Conclusion

We conclude that, given the many well-founded worries above, there is more reason than ever to trust well-established theories of explanation over mere intuitions about it, and to trust battle-tested auction successes over hazy after-the-fact rationalizations. And the verdict of these theories and trustworthy cases is clear – and negative. Economic models may give us orgasms, but they do not give us explanations.

Notes

1. The authors are jointly and equally responsible for the content of this comment.
2. Throughout, by 'economic models' we mean the idealized rational choice models characteristic of contemporary mainstream economics.
3. See, for instance, recent work by Bob Batterman, Chris Pincock, Mark Colyvan and Otavio Bueno.

4. Northcott (in press) discusses the notion of degree of explanation invoked here. It argues, among other things, that in order for such degree of explanation even to be assessed, a model must make a quantitative prediction about the value of an effect variable. This is a different complaint to Reiss's own objection to the non-quantitative nature of the Hotelling model's predictions.
5. Even if a model's main results were robust to changes in its assumptions, still it is dubious that this kind of robustness would vindicate it in any important way (Odenbaugh & Alexandrova, 2011). But the point here is that the Hotelling model cannot manage even this.
6. It seems to us that Reiss's paper itself conflates these two senses of explanation, or at least deviates from its commitment to the causal view, when it endorses the intuition that the Hotelling model is indeed explanatory (pp. 48–49).
7. For evidence, see Marwell and Ames (1981) and Frank, Gilovich, and Regan (1993). Or Reiss himself, approvingly (p. 49): 'We begin to see Hotelling situations all over the place.'
8. It may be that only male orgasm has this evolutionary purpose (Lloyd, 2005).
9. This is not to say that such feelings are entirely uninformative and should be discounted. Rather, they should be thought evidentiary but fallible. So when there is some evidence in their favour (that they are widely shared by intelligent economists) and some not (which we have marshalled in this paper), we must weigh things up. It will be clear by now where in our opinion the balance lies.

References

Alexandrova, A. (2008). Making models count. *Philosophy of Science*, *75*, 383–404.

Alexandrova, A., & Northcott, R. (2009). Progress in economics: Lessons from the spectrum auctions. In H. Kincaid & D. Ross (Eds.), *Oxford handbook of philosophy of economics* (pp. 306–336). New York: Oxford University Press.

Frank, R., Gilovich, T., & Regan, D. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives*, *7*, 159–171.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal understanding. In F. Keil & R. Wilson (Eds.), *Cognition and explanation* (pp. 299–323). Cambridge, MA: MIT Press.

Lloyd, E. (2005). *The case of the female orgasm*. Cambridge, MA: Harvard University Press.

Marwell, G., & Ames, R. (1981). Economists free ride, does anyone else? *Journal of Public Economics*, *15*, 295–310.

Nobelprize.org (2005). The Prize in Economic Sciences 2005 – Press Release. Nobel Media AB 2013. Retrieved August 9, 2013, from www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2005/press.html

Northcott, R. (2013). Verisimilitude: A causal approach. *Synthese*, *190*, 1471–1488.

Northcott, R. (in press). Degree of explanation. *Synthese*. doi:10.1007/s11229-012-0124-9

Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: Robustness analyses in economics and biology. *Biology and Philosophy*, *26*, 757–771.

Pincock, C. (2012). Mathematical models of biological patterns: Lessons from Hamilton's selfish herd. *Biology and Philosophy*, *27*, 481–496.

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

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.