Journal of Economic Methodology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjec20

Genuineness resolved: a reply to Reiss' purported paradox

Till Grüne-Yanoff [a]
[a] Department of Philosophy and the History of Technology , Royal Institute of Technology (KTH) , Teknikringen 78 B, 100 44, Stockholm , Sweden
Published online: 23 Sep 2013.

To cite this article: Till Grüne-Yanoff (2013) Genuineness resolved: a reply to Reiss' purported paradox , Journal of Economic Methodology, 20:3, 255-261, DOI: 10.1080/1350178X.2013.828866

To link to this article: http://dx.doi.org/10.1080/1350178X.2013.828866

Journal of Economic Methodology, 2013
Vol. 20, No. 3, 255–261, http://dx.doi.org/10.1080/1350178X.2013.828866

Genuineness resolved: a reply to Reiss' purported paradox

Till Grüne-Yanoff*

Department of Philosophy and the History of Technology, Royal Institute of Technology (KTH), Teknikringen 78 B, 100 44 Stockholm, Sweden

(*Received 28 August 2012; accepted 1 May 2013*)

This response to Reiss 'explanatory paradox' argues that some economic models might be true, and that many economic models are not intended for providing how-actually explanations, but rather how-possibly explanations. Therefore, two assumptions of Reiss' paradox are not true, and the paradox disappears.

Keywords: models; isolation; how-possibly explanation

Reiss' paradox is formed by three mutually inconsistent claims: (1) economic models are false; (2) economic models are explanatory and (3) only true accounts can explain.

Reiss' main effort in the paper is to show that the paradox is genuine, i.e. none of the claims can be successfully rejected. If at all, his main hope for resolution lies with finding an alternative account of explanation that allows doing away with claim (3). But 'before such a new way of thinking about explanation … can be shown to fit contemporary economic modeling, the rational response to the paradox is to remain baffled' (Reiss, 2012, p. 59).

What baffles me, rather, is why Reiss muses about the most awkward way out of his purported conundrum, instead of admitting what seems to me intuitively obvious: (i) that some economic models are true and (ii) that many economic models are not intended as how-actually explanations. Reiss does not like such a solution. He calls such quibbles with underlying claims 'resolv[ing] a paradox by fiat' (Reiss, 2012, p. 59). While not wishing to decree anything or coerce anybody, in the following I hope to offer some arguments in favour of my claims (i) and (ii).

Claim (i): Economic models are false in the sense that they misrepresent their targets in one way or another. On that I agree with Reiss. But why cannot at least some economic models be true in the sense that they represent some aspects of their targets correctly, while misrepresenting others? This is an issue of *stability* in the first place. The model describes the behaviour of a causal factor in a specific model environment. We learn from the model about this causal factor in this environment. But for the model to be true, the behaviour it describes must be stable across different environments. Specifically, the model must correctly represent the behaviour of the factor in the environment of its target, even if that environment considerably differs from the model environment.

Granted, many economic models describe factors that are not stable in this sense. Hotelling's model, which Reiss uses as an example, might be a case in point. But the question of stability is ultimately an empirical one: by studying a real-world situation, we might discover that some model correctly represents the behaviour of a causal factor in that (type of) situation. It then turns out that the represented factor behaviour is stable, and the model is true. How can one deny a priori the possibility that things turn out this way?

*Email: gryne@kth.se

At most, it seems, one could argue that as things have turned out so far, currently extant economic models are false. Yet even that seems dubious: many economists and philosophers argue that some models describe stable tendencies.[1] So there is little justification for the claim that economic models do not represent stable causal factors generally and therefore are false.

Anyway, Reiss does not directly argue for the general instability of causal claims modelled by economic models. Rather, he argues that economic model assumptions lack the property of being Galilean, and this lack undermines any warrant for the model to be true.

It is worthwhile looking at this claim in more detail. In a Galilean thought experiment, Reiss claims – following Cartwright – a situation is contemplated after 'mentally removing disturbing factors' (Reiss, 2012, p. 50). Galilean assumptions proceed by 'assuming away' (Reiss, 2012, p. 51) the influence of disturbing factors; non-Galilean assumptions, in contrast, proceed by 'determining a specific functional form' (Reiss, 2012, p. 51) of these factors. Reiss admits that this distinction might not be as clear as desirable, because 'assuming away' is just a special case of 'determining a specific form'.[2] But he insists that Galilean assumptions are distinguished from non-Galilean ones by certain features of the model in which they occur. In particular, Galilean assumptions do not normally appear in the model, they concern quantitative causal factors, and they concern factors with a natural zero.

According to Reiss, most economic models use non-Galilean assumptions. Because they do, the behaviour of the modelled causes is likely dependent on the specific form of the assumed model environment. Consequently, our expectations about the stability of the modelled factors – when only considering the model – should be low:

> There is no way to tell from just inspecting the model that it is one subset of assumptions that is driving the result rather than another .... And therefore we do not know where to look for 'truth in the model' .... (Reiss, 2012, p. 52)

In contrast to arguing for general instability directly, Reiss thus mounts a methodological argument against the probability of economic models representing stable factor behaviour. But is there really such a close link between non-Galilean assumptions and non-stability? I do not think so.

First, there are economic models whose assumptions satisfy Reiss' Galilean conditions. Take, for example, Fehr and Schmidt's (1999) model of inequality aversion. The model assumes that subjects' evaluation of an outcome to a group depends not only on their own material gain or loss, but also on the difference between their own and other group members. They experience inequity if they are worse off in material terms than other group members, and they also feel inequity if they are better off. For a two-person setting, the utility function of player i for group outcome $x = \{x_1, x_2\}$ is then given by:

$$u_i(x) = x_i + \alpha \max[x_j - x_i, 0] + \beta \max[x_i - x_j, 0] \quad i,j\{1,2\}. \tag{1}$$

The second term in Equation (1) measures the utility loss from disadvantageous inequality, while the third term measures the loss from advantageous inequality.

In the light of Fehr and Schmidt's model, all standard utility models that only take into account an individual's own material gains or losses *isolate* from inequality aversion. More specifically, they isolate from inequality aversion by *assuming away* the advantageous and disadvantageous inequality effects. These assumptions do not explicitly appear in the standard models; one needs to contrast them against the Fehr–Schmidt model to recognize them. They concern quantitative causal factors, in particular the material gains x for each individual and the weights α and β. Finally, these factors have a natural zero, namely the case where $x_i = x_j$, or where $\alpha = 0$ or $\beta = 0$. Thus, according to Reiss' conditions, the idealization all standard utility models make for the purpose of isolating utility from inequality effects is Galilean in character. Whether that makes them true I do not know. But it shows that the Galilean/non-Galilean distinction does not justify the general claim that economic models are false.

Furthermore, it is not clear that Galilean assumptions are a *necessary* condition for the truth of a model. Let us take the above example of price-generating properties of auctions being stable with respect to changes in the time interval between rounds. If a game-theoretic model representing a multiple-round auction design assumes a uniform time interval between rounds (e.g. by setting the time unit to 1 h), then the model might correctly represent the auction's price-generating properties even in situations where the intervals are not uniform or not 1 h each. But is the model's idealizing assumption (of uniform 1 h intervals) Galilean? Not according to Reiss' criteria: the model explicitly sets the time intervals to a specific value – to 1 h, and to uniformity. Furthermore, although the causal factor – the interval – is quantifiable, it does not have a natural zero.[3] Thus, there might be models that correctly represent aspects of a situation – and are true in this sense – although they isolate the relevant aspect with the help of non-Galilean assumptions.

Thus, although I have my own reservations about the isolating claims of economic models (see Grüne-Yanoff 2011), I am sceptical about Reiss' claim of a strong link between non-Galilean assumptions and non-stability. It does not hold up as a justification of the general claim that economic models are false. But maybe that is not what Reiss is after? Recall the above-quoted passage, where he argues that non-Galilean assumptions prevent us from knowing 'where to look for the truth' because they disallow identifying 'from just inspecting the model' that subset of assumptions 'that is driving the result'.[4] This is a claim about what we can learn 'from just inspecting the model', rather than about the truth of the model. These two claims must be clearly separated: in the above auction model case, the model might turn out to be true – but there might be no way of telling this 'from just inspecting the model'. Perhaps, Reiss is correct that with non-Galilean assumptions mere inspections of a model do not teach us much about its truth. However, that does not exclude the possibility that, in empirical investigation, it turns out that the behaviour of a tendency described in the model is indeed stable – and the model true in that sense. And that is all that is needed to reject claim (1) of Reiss' paradox.

Claim (ii): Of course, the above arguments do not do away with the concern that economic models are often – some would say too often – false, and that given claims (2) and (3), the success record of economic modelling is rather poor. If economic models are intended to provide how-actually explanations, and if how-actually explanations can be had only from true models, then even the fact that sufficiently many models are false should create discomfort among economic modellers.

But this presupposes that economic models are indeed intended to provide how-actually explanations – a presupposition that I now show is not correct. I will start with discussing some of the authors who Reiss cites as evidence for his claim that economic models intent to explain.

One author Reiss quotes as offering a model for explanation is Abhijit Banerjee. In his 1992 paper, Banerjee compares his model to an alternative, principle-agent model of herd behaviour. This model is substantially different to his own: herd behaviour in that model arises from agents getting rewards for convincing a principal that they are right. In Banerjee's model, in contrast, herd behaviour arises without such incentive distortions. Thus, for a given explanandum, it is impossible for both models to be true. Yet that does not faze Banerjee:

> In any case, this [principle-agent] approach is not inconsistent with our approach. This kind of principal agent problem ... seems common enough, especially in the context of asset markets. On the other hand, in many of the other potential instances of herd behavior, such as fertility choices, adoption of innovations, voting, etc., there is no obvious principal agent problem. (Banerjee, 1992, p. 801)

If Banerjee aimed at providing a how-actually explanation of some specific explanandum or type of explananda, such a position would not make sense. If only true accounts explain, and the explanatory aim is unambiguous, then only one of two distinct models can give a how-actually explanation. Consequently, Banerjee must be intending his model for another purpose. As he writes, he focuses on *potential* instances of herd behaviour and gives accounts of how some of these potential instances could have come about. In other words, he offers how-possibly explanations of herd behaviour. Whether certain phenomena are instances of herd behaviour, and whether Banerjee's how-possibly explanation is the how-actually explanation for certain of these instances, is not of his interest. Rather, he leaves it for others to decide.[5]

The second author I wish to discuss is Thomas Schelling, whom Reiss quotes as offering explanatory models. Most attention in this discussion has focused on Schelling's segregation model. In the original 1971 paper, Schelling justifies his modelling exercise as an exploration, not an explanation. In fact, he mentions explanation only once in the whole paper, and then not in relation to his model. Instead, the modelling effort is directed at discovering a possibility:

> To understand what kinds of segregation or integration *may* result from individual choice, we have to look at the processes by which various mixtures and separations are brought about. (Schelling, 1971, p. 147, my emphasis)

Schelling is even more explicit than this, characterizing his modelling efforts as 'an *abstract exploration* of some of the quantitative dynamics of segregating behaviour' (Schelling, 1971, p. 148, my emphasis). Beyond the brief mentioning of anecdotal examples, no specific explanandum is identified at all. When, in the final section, the question of application arises, Schelling writes: 'The foregoing analysis can be used to *explore* the phenomenon of "neighborhood tipping"' (Schelling, 1971, p. 181, my emphasis). Notably, Schelling does not advertise his model as an explanation of this phenomenon. Rather, to the contrary: his model produces 'a number of possibilities' (Schelling, 1971, p. 182) that show that 'in none of the cases shown does any important discontinuity necessarily occur at the modal or typical tolerance value' (Schelling, 1971, p. 182). The model, rather than explaining the phenomena of interest, shows possibilities of its production that had not been considered before. Instead of a how-actually explanation, Schelling's model provides us with a how-possibly explanation.

What is, of course, curious is that these authors often use the concept of explanation and explanatory model, even though an analysis of their modelling practices reveals that they are not aiming at how-actually explanation. What I conclude from this observation is that these economists – and many other economic modellers with them – are confused about the difference between how-actually explanations and how-possibly explanations. Admittedly, committing this conceptual confusion is hardly a crime, as no generally accepted account of this distinction is extant. Yet Reiss' paper shows how important it is not to confound these two concepts, at pain of paradoxical conclusions.

Unfortunately, Reiss himself is not very clear on this distinction, which leads him to uphold claim (2) and insist that his purported paradox is genuine. But once the distinction is in place, the paradox disappears. In particular, this distinction allows one to agree to claim (3) but, at the same time, to insist that models which do not represent any actual target (and hence in Reiss' parlance are false) might still offer valid how-possibly explanations.

Defenders of Reiss' positions might reply that how-possibly explanations are merely a preliminary step for how-actually explanations. To build models for the purpose of how-possibly explanations thus already implies the aim of how-actually explanations, and that is what the modelling should be assessed by. Reiss himself seems to follow this line, when he argues that any such account merely 'ignore[s] rather than solve[s] the problem' (Reiss, 2012, p. 54).

Two replies to such a counterargument seem convincing to me. The first is to argue that although how-possibly explanations might just be a preliminary step on the way to how-actually explanations, the model only provides the preliminary step, and some other practice provides the how-actually explanation. In this way, the model functions only as a heuristic device, with whose help a set of causal hypotheses can be formulated more conveniently or more precisely. These causal hypotheses are then tested with the help of empirical and experimental practices, and potentially further modified. In the end, a sufficiently well-supported hypothesis will function as the explanation of the phenomenon in question, without recourse to a model representation of this hypothesis. In such a case, the model itself is never explanatory, but it contributed to the explanation in various ways. Such a view, as I believe Alexandrova (2008) proposes, seems to capture important modelling practices in economics, and I fail to see why such an account ignores the problem of explanation in economics.

I believe that this first reply does not capture all relevant aspects of how-possibly explanations. In the following, I will sketch three accounts that interpret how-possibly explanations in contrast to how-actually explanations, rather than as their mere precursors.

First, Dray (1957) claims that how-possibly explanations have a different aim and a different structure from how-actually explanations. How-possibly explanations aim at giving an account how events that are considered impossible could have happened (for a related discussion regarding economic models, see Grüne-Yanoff 2009). How-actually explanations, in contrast, aim at accounting for how actual events have happened. Furthermore, Dray argues that how-possibly explanations rebut the impossibility of the explanandum by giving a necessary condition for its occurrence. He contrasts this with actual explanations offering sufficient conditions for their explananda.

Dray-type how-possibly explanations focus on identifying some *conditions* that show the possibility of the explanandum. Another kind of how-possibly explanation instead focuses on indicating the sort of *process* through which the explanandum took place (Reiner 1993). Consecutive authors point out that this may consist in a mere proposal of a possible mechanism, or alternatively in providing a partial mechanism that in fact had the explanandum as outcome. In the latter case, the actual mechanism that produced the explanandum is identified, but in a way insufficient 'to see more how the explanandum phenomenon was produced' (Persson 2011). Both purposes are served by economic models – the first by a model presenting a possible process, the second by a model presenting an actual process without sufficient causal detail, under possible background conditions.

Finally, Forber (2010) distinguishes between *global* and *local* how-possibly explanation. Global how-possibly explanations account for the possibility that an idealized object has a certain property, produced by a possible process from possible background conditions. Their purpose is to investigate the capabilities of general model processes (Forber, 2010, p. 33). Local how-possibly explanations, in contrast, account for the possibility of a real target object having a certain property, produced by a possible process from actual background conditions. Their purpose is to guide speculation on how a particular model process can produce actual target properties. Forber's distinction thus points to a difference between models with an abstract result, and those with a concrete result.

Thus, how-actually explanation requires the identification of true (sufficient parts of) causes that brought about the explanandum. This requires true models. How-possibly explanations, in contrast, identify elements of *possible* causes for an explanandum. Models can represent such possible causes – and hence contribute to how-possible explanations – without representing real-world targets. How-possibly explanations, in Dray's, Reiner's, Persson's or Forber's sense, give a purpose to models that do not accurately represent real-world targets (for more details on this, see Grüne-Yanoff, in press).

Crucially, this purpose is distinct from that of how-actually explanations and makes different requirements on models employed for these ends. This allows me to qualify Reiss' claim (2). While some models indeed are explanatory in the sense of intended for how-actually explanations, many economic models instead are intended for how-possibly explanation.

Adjusting Reiss' original claims in the light of the above arguments yields the following:

(1^a) Many economic models are false, but some are true (in the abstract).
(2^a) Some economic models are intended for how-actually explanations, but others are intended for how-possibly explanation.
(3) Only true accounts can explain.

Claim (3) need not be changed, but should be interpreted as 'Only true accounts can provide how-actually explanations.' Then those models that are true offer how-actually explanations, and those models that are false might offer how-possibly explanations. Of course, there will be some false models mistakenly intended for how-actually explanations, but that is part of normal scientific practice. Hence, (1^a), (2^a) and (3) are mutually consistent. The genuineness of Reiss' paradox is resolved, bafflement subsides and no non-standard account of explanation is necessary.

Notes

1. For example, Alexandrova (2008, p. 393), with reference to Plott (1997), identifies a stable causal fact of multiple-round auctions: the length of intervals between stages does not affect the price-generating properties of auctions.
2. Typically by setting the coefficient of this factor to 0. But of course such salience considerations might vary across contexts.
3. The time interval arguably has a natural zero, 0. But that cannot be the natural zero of the causal factor: a multi-stage auction with interval 0 is not a multi-stage auction.
4. I find it difficult to understand what it means that a subset of model assumptions 'is driving the result'. Is this a florid way of saying that some assumptions are relevant for the model result, while others are not? Or is it meant to express inferential robustness? Or that some assumptions' influence on the model result is of a different quality than others' (despite the fact that all assumptions have some influence)? Until the underlying intuition of this claim is made clearer, I am afraid that the Galilean/non-Galilean distinction will remain somewhat obscure.
5. Hence, I agree with Sugden's analysis when he writes 'Banerjee is trying to convince us that we should take seriously the *possibility* that his explanation of real-world herding is correct' (Sugden, 2009, p. 10, his emphasis).

References

Alexandrova, A. (2008). Making models count. *Philosophy of Science*, *75*, 383–404.
Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, *107*, 797–817.

Dray, W. (1957). *Laws and explanations in history*. Oxford: Oxford University Press.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.
Forber, P. (2010). Confirmation and explaining how possible. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*, 32–40.
Grüne-Yanoff, T. (2009). Learning from Minimal Economic Models. *Erkenntnis*, *70*, 81–99.
Grüne-Yanoff, T. (2011). Isolation is not characteristic of models. *International Studies in the Philosophy of Science*, *25*, 1–19.
Grüne-Yanoff, T. (in press). Appraising models non-representationally. *Philosophy of Science*.
Persson, J. (2011). Three conceptions of explaining how possibly – And one reductive account. In H. de Regt, S. Okasha & S. Hartmann (Eds.), *The European philosophy of science association proceedings* (Vol. 1, pp. 275–286). Dordrecht: Springer.
Plott, C. R. (1997). Laboratory experimental testbeds: Application to the PCS auction. *Journal of Economics & Management Strategy*, *6*, 605–638.
Reiner, R. (1993). Necessary conditions and explaining how-possibly. *Philosophical Quarterly*, *44*, 58–69.
Reiss, J. (2012). The explanatory paradox. *Journal of Economic Methodology*, *19*, 43–62.
Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186.
Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, *70*, 3–27.