On a paradox of truth, or
how not to obscure the issue of whether explanatory models can be true

Uskali Mäki

To appear in the *Journal of Economic Methodology*

"I have not found anyone writing on economic models who has explicitly challenged logic (though their writings sometimes suggest otherwise)." (Reiss 2012, 49)

Introduction

Julian Reiss's "The explanation paradox" has a charming title and an attractive macro structure, but the care of its argumentative performance leaves a lot to be desired. I will show how the resolution of the paradox of falsehood in explanation presupposes resolving another paradox, a paradox of truth in modelling. My own previous work has pursued such a resolution, and Reiss's article sets out to refute it. I will point out several serious flaws in his reasoning, including important ambiguities and inconsistencies as well as the failure to adequately address the views he discusses.

What's the paradox?

The presumed paradox is a set of three claims (this is my reformulation that rephrases one of them and reorders the three claims in what seems an obvious way):[1]

---

[1] Reiss's formulation of the paradox is this (49):

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

The reason for suggesting changing the order of the three claims should be clear. The changed order brings out something like a paradox in a way that the original ordering does not seem to. The reasons for reformulating one of the claims by replacing 'accounts' by 'models' might also be obvious. Reiss does not say what he means by 'account'. If he were to mean 'explanation' then the claim would be a rather uninformative 'Only true explanations can explain' and it would not be closely linked to the other two that are put in terms of 'model'.

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

Calling the presumed paradox "the explanation paradox" is not maximally informative, since explanation is not its sole focus; it rather deals with the connection between truth, explanation, and models. Other slogans – such as 'the paradox of false explanatory models' – would compromise on simplicity but would be more accurate.

Formulations (1)-(3) are not specific enough to provide us with a clear idea of what exactly the paradox is. Each of the three elements requires explication. Their present formulations are too ambiguous to allow for a careful inquiry into the grounds on which one might or might not accept any one of these claims.

Claim (1) expresses a philosophical principle, also endorsed by many practicing scientists. Warrantedly accepting or rejecting it requires a philosophical argument. But the contents of (1) and the respective arguments depend on what we mean by models explaining and models being true. So its precise contents is parasitic upon those of the other two claims.

The formulation of claim (3) is in terms of being explanatory. This can mean a number of different things. These ambiguities can be put by means of a few dimensions and distinctions. First, a model can be explanatory in the weak sense of contributing to explanations, by playing a useful or perhaps indispensable role in the business of explaining phenomena; or a model can itself directly explain. Second, a model can be explanatory by being a candidate for either of the above roles, for possibly explaining and for possibly providing explanatory contributions; or it may actually accomplish the tasks. Third, a model may be explanatory; or it may just feel explanatory by its users (somewhat puzzlingly, Reiss says that "perhaps more importantly", models *feel* explanatory (48-49)). Fourth, 'explanatory' means as many different things as one takes 'to explain' to mean, such as tracing ontic dependencies in the world; answering contrastive why-questions; performing inferences using arguments with certain required structures; and so on. Given that Reiss does not specify his favourite meaning(s) of 'explanatory', his reasoning about (3) and the rest of the paradox remains excessively vague.

Claim (2) is in the indicative mode, appearing to suggest a contingent fact. But what fact exactly? The relevant quantifier is not spelled out. Are all economic models false, or many of them, or just some? Perhaps models built by economists thus far are all false. This could be a conclusion from an empirical investigation of all actual economic models. The possible reasons for the contingent falsehood of economic models are numerous, such as underdeveloped tools available to economists, mistaken theories held by them, poor education, systematic ideological bias, flawed incentives in the institutional structure of the discipline, the cognitive difficulty of the tasks of modeling, and so on. Alternatively, one could claim that not just economic models actually built in the past, but any economic models, including those to be built in the future, are false. This latter idea might be difficult to justify without resorting to a modal formulation. Indeed, this is another possibility: economic models – perhaps because of the general nature of models – cannot be true. Whatever models economists might build, it would be impossible for them to be true, regardless of their contents. It seems Reiss is ambiguous between all these versions – and yet others, as we shall see.

Reiss correctly identifies my account as disputing something like (2) in the paradox. But he fails to address the version of it that I seek to rebut as well as the arguments I have proposed. My account can be characterized by saying that it aspires to resolve another paradox by suggesting that possible truths hide behind apparent falsehoods.

A paradox of truth

The paradox I have in mind might be formulated as follows (the precise formulation does not matter much for the purposes here): (a) Models violate nothing but the truth. (b) Models violate the whole truth. (c) Nevertheless, models might be true.

I have looked for ways of resolving this paradox. If found, they would also dissolve the paradox of false explanation. My solution is based on the usual recommendation when dealing with paradoxes in general: look more closely! Economic models only appear to be false but upon closer inspection, you will see that the reasons you thought are sufficient for the falsehood of models aren't sufficient after all. What you will discover is that even if isolating just small fragments of the world (thereby violating "the whole truth") and involving false assumptions (thereby violating "nothing but the truth) models can in principle be true about what they are about.

In articulating this strategy and its implementation I've actually used formulations that invoke a paradox, such as "the truth of false idealizations" (2011), and I've often said that "there is more truth in economic models than easily meets the eye". The point is that models and idealizing assumptions appear to be – mostly, always or necessarily - false, but when appropriately understood, they may be given the chance of being true. In particular, the features of models and their assumptions usually considered as sufficient reasons for their falsehood are not sufficient. Models can be true in spite of such apparent falsehoods or even by virtue of them.

Note the way I am putting the idea. I am defending the *possibility* of economic models to be true, arguing that this possibility is not undermined by the appearance of falsehood. This is not captured by the phrasing of (2) in the paradox of false explanations, nor does Reiss explicitly address this formulation elsewhere.

My suggested resolution of the truth paradox, if correct, would immediately also resolve the paradox of false explanations. In order for Reiss to block this line of reasoning, he would have to refute my suggested resolution of the truth paradox. He does no such thing, as I will show in the following sections.

On a strategic failure: missing the target

In section 4.1 Reiss starts out saying I am "the main advocate" (50) of the view that models can be true, but then he proceeds without providing criticisms of the details of my arguments. Indeed, he does not really discuss my views. There is a chance that my arguments are flawed, and therefore I would be the first to welcome a critical assessment of my account. Since Reiss does not offer one, his criticisms do not get off the ground and so his conclusions a few pages later (53) remain unwarranted.

Yet it is possible that Reiss has offered some sort of indirect or implicit criticism of the view that economic models can be true even if he does not explicitly address the arguments that have been proposed in support of this idea. Let us see what he says.

On truth bearers

Reiss starts section 4.1 with a disclaimer: "I do not think that models have true [sic] values. Whatever models are … it is most certainly not the case that models are sentences. But its [sic] sentences that are true or false." (49) This is odd for at least three reasons.

First, philosophers have proposed and defended several sorts of candidates for truth bearers (things that can be true or false), including sentence types and sentence tokens (Reiss does not say which of these is his favourite), but also beliefs, utterances, statements, propositions, thoughts, and more. Reiss says nothing to motivate or justify his picking out sentences from this set of options. He does not let the reader know that he is picking. He simply asserts that sentences are the appropriate truth bearers, without mentioning other possibilities. But there are other options. That models are not sentences therefore does not imply anything for the idea that models might be true or false. Models may fail to be sentences and yet may qualify as appropriate truth bearers. Suggesting otherwise would require a little bit of an argument. Mere assertion is not enough.

Second, my defense of the possibility of models being true is partly based on rejecting the argument (held by Ronald Giere) that models cannot be true because they are not linguistic entities like sentences are. Indeed, I have explicitly aspired to reject the doctrine that only sentences can be truth bearers (e.g. 2001). In my (2009a) and (2011a) I outline the beginnings of an elaboration of the claim that models (as imagined objects) are the sorts of thing that qualify as truth bearers. Reiss is silent about this suggestion. This means his paper does not even touch my account, let alone criticize it.

Third, Reiss contradicts himself. On the one hand, he asserts that models cannot be truth-valued – true or false – because they are not sentences. On the other hand, the whole argument of his paper presupposes that models are truth-valued – they are false. He formulates his paradox in terms of the premise that "economic models are false" and the question he sets out to answer in his paper is built on the presupposition that models themselves are truth-valued after all: "Do false models explain?" Likewise, he characterizes Hotelling's model as "a paradigmatic example for a false explanatory model" (44). But if Reiss believes that only sentences can be truth-valued and that models are not sentences, the key question of his paper would seem to make little sense.

Reiss might defend his case by appealing to an idea that appears just in passing in his paper (but is not used by him to do any systematic work): "… when we say that a model is true or false, we speak elliptically" meaning to say that "(s)tatements are true or false *of*" models (49). One obvious reading of this is that it proposes the *translation rule* according to which 'model M is true or is false' means 'claim C about model M is true or is false'. But this too would lead to odd consequences. First, we can make a number of different kinds of claim about any given model, and surely some of those claims might be true. But it would be strange to say that one and the same model is true whenever a claim about it happens to be true, and false if a claim about it happens to be false. Second, Reiss holds that no model is, or can be, true. By the translation rule, he would come to hold that no claim about any model can be true. So some qualifications would be needed.

As we next look more closely at what Reiss might mean by 'false' and 'true' we discover that he uses these terms in a number of mixed and unconventional ways that seemingly allow him to say that models are false and therefore truth-valued after all (while denying this elsewhere). This contributes to a conceptual mess.

Falsehood as "misrepresentation"

Reiss sets out to clarify the setup by suggesting: "Thus, when we say colloquially 'all models are false' what we mean is 'all models misrepresent their targets in one way or another'. (49) This is uninformative and clearly does not clarify. Replacing 'false' with 'misrepresent' adds nothing to the clarity of the statement, on the contrary. 'Misrepresent' has no single well-understood and standardized meaning that would be helpful for the present purposes. Consider two options. If one takes 'misrepresent' to mean 'fail to represent', it seems to be a misnomer itself: not only does this suggestion fail to bring about clarification, it is internally incoherent. This is because in order for something to be false (or true), it must represent, so nothing that "misrepresents" can be false. Any falsity is a false representation. If, secondly, one takes 'misrepresent' to mean 'failing to correctly represent', one will add nothing to clarify the notions of falsehood and truth. This only puts forward the same challenge in other words – in words that are not more but less lucid. 'Correct' does not help when we try to understand what we mean by 'true' and 'false'.

The same idea appears in his claim that instead of saying that *models can be true or can contain truths* (as I would) it would be "more accurate to say (for instance) that a *theoretical hypothesis stating that the model correctly represents* a target system's causal power or mechanism can be true …" (50; italics added). So instead of models themselves, theoretical hypotheses *about models* can be true; using his phrases, the latter would be a non-elliptical, more accurate way of putting the idea. But this adds to the confusion by not explaining what "correctly represents" means and why this is not a matter of true representation.

Types of falsehood / idealisation / unrealisticness

In his mixed reasoning strategy, Reiss also appeals to William Wimsatt's paper "False models as means to truer theories" (included e.g. in Wimsatt 2007, 94-132). It seems this only makes things worse. The first confusion emerges when Reiss calls Wimsatt's typology one of "different kinds of idealisations" (46) while Wimsatt himself says to have suggested “ways in which a model can be false” (Wimsatt 2007, 101). The confusion is reinforced when Reiss reproduces Wimsatt’s type [2] that Wimsatt himself calls an *idealisation* in contrast to the other types that are not so called by him (46). However, Reiss appears to be undecided about this since a page later he turns to Wimsatt’s parlance by beginning to talk about models being “false” in senses [3] and [5], and Hotelling’s model being “false in all relevant senses from [1] to [5] from Wimsatt’s list” (48).

This is after all a minor confusion compared to the major one that arises from the use of Wimsatt’s typology for dealing with the issue of truth in explanation. Reiss concludes that Hotelling’s model is “false” in Wimsatt’s senses [1]-[5]. “And yet, it is considered explanatory.” (48) However, Wimsatt’s suggested typology of ways of being “false” that Reiss part of the time renames a typology of “idealisations” is neither of these. Even though Wimsatt uses the terminology of falsehood, he has not formulated a typology of falsehoods (Mäki 2009a) – and so the typology cannot serve in Reiss’s argument.

Wimsatt’s is rather an incomplete typology of kinds of *unrealisticness*. Unrealisticness is a mixed family of properties of representations. The reason for bringing them together under a shared umbrella term is nothing more profound than people’s unreflected inclination of characterizing models and their assumptions as “unrealistic” while picking out very different properties. Falsehood proper is one of those properties - but it is just one among many others. ‘Falsehood’ is not the umbrella term that should be used for them all. Nor is ‘idealization’.

The confusion is repeated in Reiss’s comments on Sugden’s ideas. Reiss states that Sugden subscribes to claim (2) of the paradox, namely that economic models are false. But the evidence Reiss cites does not support this, since Sugden only admits that models “appear absurdly unrealistic” (55).

On truth and functional decomposition

I have called my account of models (and the locus of their truth) a functional decomposition account (2009, 2011b). It is a *decomposition account* since it relies on splitting models into bits and pieces rather than dealing with them as undifferentiated wholes. It is a *functional account* in that it is based on attributing distinct functions to those bits and pieces. Different parts of the modelling exercise typically serve different functions, and one cannot understand the point and possible achievements of any given modelling exercise without considering them in relation to such functions and the overall goals of the particular exercise. Some of those functions are such that there is no truth claim made about the world when using the respective parts of the exercise.

Reiss believes to have given representative examples of falsehood in models when reminding us that money "is not wet as the water in the Phillips machine" and that banks "are not plastic tanks filled with water" (49). This is a misunderstanding. Not all properties of models are (equally) relevant for representing their targets – or do not serve equally relevant representational functions. In the case of the Phillips machine, properties such as the quantity of water and the velocity of its flow are among the relevant ones, while the wetness of water and the plastic material of the tanks are not. The irrelevant properties are not to be submitted and assessed for their similarity with the properties of the targets. Since they are not candidates for such assessment – they are not truth nominees (see Mäki 2011b) – they are not to be treated as truth valued (or if so treated for some irrelevant purpose, their falsehood should be of no relevance).

Compare this to Reiss's favourite truth bearers, sentences – let's say sentence tokens. I presume he would not want to say that a given sentence token, say, the one written by him in his diary using pink ink, confessing he has lost sympathy with a colleague, is false because his feelings are not pink nor made of ink. I would suggest that these ingredients are not among the relevantly truth-valuable parts of his favourite truth bearer.

On isolation by idealization

It has been part of my account of models that models isolate by idealization. Building and manipulating theoretical models is structurally similar to laboratory experimentation. While in the lab "other things" are controlled for – eliminating or stabilizing their influence -- by way of causal manipulation, in theoretical modelling they are controlled for by making and changing idealizing assumptions. In this way, important factors and connections of interest are isolated from the rest. I have suggested that much of economic modeling uses this strategy. Reiss disagrees. He argues that "the models of economics … are by and large very much unlike" isolations by idealization (or as he puts it, "Galilean thought experiments") (51).[2]

How does Reiss argue for his rejection of the idea that economic models isolate by idealization? Using Hotelling's model as an example, he argues that few of its assumptions "aim to eliminate disturbing causal factors" (51). He gives three examples:

> "Assuming businesses move along a line with no breadth or thickness is not assuming away the influence of geography…
> Assuming that transportation costs are linear in distances is not assuming away the influence of transportation costs…
> Assuming that demand is perfectly inelastic is not assuming away the influence of demand…" (51)

I do not see the point of saying these things. They are trivial, and I presume no one ever thought otherwise. More importantly, they seem to be beside the point, based on a misunderstanding. They are like saying "assuming perfect competition is not assuming away the influence of competition" while the relevant thing to say is that this assumption indeed can be used to remove from the model a number of things,

[2] For a different criticism, see Till Grüne-Yanoff 2011.

such as price making and entry restrictions[3] (on assuming perfect information, see Mäki 2009, 31). In a similar vein, the assumption of businesses moving along a line removes the influence of anything along the other two dimensions. These can be interpreted as geographical dimensions, and also as dimensions of the space of product characteristics over which consumer preferences are defined (such as the sourness of cider as in one of Hotelling's examples of one-dimensionality, removing all other characteristics of cider). The assumption that transportation costs are a linear function of distance removes non-linearities by ignoring, say, costs of loading and uploading, variations in road (or whatever) congestion, and generally of factors other than distance. The assumption of perfectly inelastic demand removes the impact of price on demand, for reasons such as the lack of substitutes. And so on.

The general challenge is to be careful with what exactly an assumption says, or can be used to express. Any given assumption can be formulated in a number of ways and used for expressing a number of ideas, or making a variety of claims. For example, the assumption of a one-dimensional geography can be turned into assuming that the impact of other dimensions is negligible given the purpose for which the model is used, or that it is bracketed for the time being and will be incorporated later (see Mäki 2011b). It takes some extra reflection by a modeller or an analyst of a model to determine the appropriate versions in relation to the use of the model for any given purpose. The outcome of this reflection can be reported in what I have called the *commentary* of the model. Without this, the issue of unrealisticness and the paradox of truth also remain unresolved.

On "Galilean" and "non-Galilean" assumptions

Reiss continues by suggesting a distinction between Galilean and non-Galilean assumptions and by listing three differences between them (51-52). I must confess I was not able to understand this passage, nor did I see its relevance. Reiss does not say what he means by "non-Galilean assumption" and the presumed differences may not be terribly striking given that he describes them in terms of 'normally' and 'usually'.

[3] And, incidentally in this case, the assumption does remove certain ways of competing, often called rivalry.

I also fail to see why he claims the one-dimensional geography of "a line with no thickness" (along which firms move in Hotelling's model world) to be of a kind different from "point masses" such that only the latter would be quantitative and have a natural zero (51). I would have thought one-dimensional geographies and point masses to be in the same family.

On idealisations as giving applicability conditions

Reiss writes: "A model may be an *idealisation* whose conditions of applicability are never found in nature (e.g. point masses, the uses of continuous variables for population sizes, etc.)…" (46) This sentence is hard to decipher, but if this is taken to mean that point masses and continuous population sizes are examples of the (never-met) conditions of applicability of the respective models, I would suggest this is mostly wrong. Such idealisations are not usually intended as describing applicability conditions; if they were, the respective models would never apply, and so could never explain. However, a model about the solar system depicting the planets as point masses can explain certain facts without the planets really being point masses – and the model would not improve on its explanatoriness even if the planets actually were to be reduced in volume, with the consequence that the assumption would become a closer approximation to the facts. The assumption of mass concentrated in a point does not provide an applicability condition.

On truth and recognizing what is true

Reiss writes that "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. … therefore we do not know where to look for 'truth in the model' …" (52) This switches the focus of Reiss's argument. He is here talking about difficulties of knowing or of recognizing the truth, of assuring ourselves that we have discovered it. His concerns here are epistemic.

But recall the formulation of Reiss's paradox. One of the claims was, "Economic models are false". As I pointed out in the beginning, this is ambiguous. It is also not clear what claim Reiss has set out to reject in section 4.1 dealing with my account. As I explained, I would not defend claims such as "All economic models are true" or "Most economic models are true"; instead, I only defend the idea that models might be true, or that they have a chance of being true, or that the usual appearances of falsehood in models do not make them false.

The ambiguities are now fortified in that the claim he is criticizing here seems to be, "Models can be known to be true". This new focus comes as an unreflected surprise to the reader. This claim (that models can be known to be true) is not part of the original formulation of his paradox – nor of that of mine. Moreover, as we have seen, Reiss has no need to attack this new claim, given that he has already asserted that models are the sorts of thing that cannot be true (or false). This assertion should first be withdrawn for the new issue of epistemic recognition to make sense.

To support his new claim, Reiss argues that because robustness cannot be ascertained, one cannot identify, in the whole set of assumptions of a model, those that "drive the result" and so the "model result depends on the entire array of assumptions" (53). Moreover, "if these assumptions are false", the mechanism depicted by the model cannot be believed to function in the target, and so the model cannot explain (53). It is not clear whether by "these assumptions" he means the entire set of assumptions, but yet the conclusions he draws are radical. Do any models in any discipline ever possibly manage to depict real mechanisms in their targets?

It is obvious that economics is an extremely difficult subject just because deep epistemic uncertainties are so hard to remove. But this alone implies nothing about whether models could be true, or could contain truths. The latter has been my primary concern, one that has to be settled before addressing the epistemic concern. I have examined the issue whether models are the sorts of thing that in principle could be true, and I have traced a path to a positive answer. My concerns here are semantic (dealing with how models relate to their targets) rather than epistemic (dealing with how cognizers relate to the truth-values of models). The semantic concern has to be settled before addressing the epistemic concern. So one suspects Reiss's argument begins from the wrong end and misses its target.

But let us play along and check how Reiss's reasoning proceeds. Within an epistemic framework, he sets out to block the argument that economists try to reduce the uncertainties of modeling by conducting robustness tests (see Kuorikoski, Lehtinen, Marchionni 2010). But, says Reiss, "robustness tests are not possible, and if possible and performed, their result is negative" (52). There are a number of possible claims that Reiss might want to make with these statements. He might claim that robustness imposes no constraints on the choice of models and their assumptions. If so, robustness would be useless even if found. Or he might claim that there is no way ever to establish robustness with even a slightest degree of assurance. If so, the search for robustness would be useless. These would be very radical claims, and far stronger arguments would be needed to support them.

Let me return to my first point, namely that the semantic concern (with the possibility of models being true) has to be kept separate from the epistemic concern (with our chances of recognizing the extent to which truths are actually attained), and that the latter has no implications for the former. Now one might entertain the thought that no epistemic progress whatsoever is possible in economic modelling – that regardless of how much and whatever kind of testing is done, all possible models remain equally good (or equally bad) candidates for truth. Economists do not and cannot have any capacities of telling good from bad, or better from worse models in this respect. On such a scenario, we might then want to draw the conclusion that, even if analytically appropriate perhaps, the distinction between semantic and epistemic concerns has zero relevance and therefore had better be discarded. Having an idea of what it would be for a model to be true is entirely useless because economists can never say anything whatsoever about whether one model is better than any other in terms of its truth content. I wonder if this is what Reiss would like to conclude. I would not, and he has not given me any reason to revise my view.

On non-explanatory functions of models

In section 4.2 Reiss briefly discusses some views of models that stress their non-explanatory functions, such as conceptual exploration, serving as open formulae for constructing hypotheses, and establishing modal hypotheses (54-55). These are supposed to represent one line of resolution of the paradox, based on denying claim (3) that models are explanatory. I fail to see how viewing models as having any of the above functions would imply the denial of their explanatory functions. It seems Reiss has not presented (and rejected) an argument in support of rejecting claim (3) of the paradox.

Section 4.2 manifests another weakness in Reiss's reasoning. Nowhere does he give an articulate idea of what he means by 'explanatory' – it is, after all a key notion in the paradox. Consider the idea of how-possibly explanation that he in this section seems to depreciate as not really explanatory: "Possibility hypotheses, as much as they might teach us about the world, do not explain economic phenomena." (54) Now it all depends on what one takes to be a mark of being explanatory. If being explanatory involves providing explanatorily relevant information about the ontic structure of the world, then no doubt correct how-possibly accounts are explanatory (and false ones are not!). If, by way of empirical inquiry, they can be turned into how-actually explanations (perhaps of singular phenomena), they start providing different kinds of explanatory information – rather than start being explanatory per se (as it seems Reiss would think).

On credibility

The pragmatic notion of credibility has no established meaning in the philosophy of science, therefore anyone using the term in an important role has the duty of specifying its meaning. Several ideas of credibility appear in Reiss's section 4.3, but he does not clearly distinguish them from one another as distinct notions.[4]

[4] On top of those to be discussed below, there are of course other possible notions of credibility, such as persuasiveness and credit-worthiness (Mäki 2009, 39-40).

This is formulated so as to look like Reiss's official definition: "There is something that characterizes good economic models in virtue of which they are acceptable by the economics community. Let us call that their credibility." (56) On this suggestion, credibility becomes defined in terms of goodness and acceptability (or just acceptability in case credibility and goodness are not intended as distinct attributes). Whatever contributes to acceptability is among the characteristics that constitute credibility. This seems to give us a rather uninformative dummy notion of credibility.

There are other, and different, notions of credibility that appear in passing in the section. One is put in epistemic terms that make appeal to ontic matters. This is in a passage that Reiss quotes from Sugden where the latter says that a model is credible in the sense that "it is compatible with what we know, or think we know, about the general laws governing events in the real world" (55). This is roughly what I've called the *www constraint* on acceptability (Mäki 2001) and that I've suggested is a matter of *conceivability*. This appears to be Sugden's concept of credibility, but Reiss does not use it when dealing with Sugden's overall account.

A third notion of credibility is implied by the statement, "economists' subjective judgements of plausibility or credibility are strongly influenced by their theoretical preferences …" (56) This suggests a version stronger than conceivability. To be *plausible* is is to be believed actually to be the case with some sufficiently high likelihood, while to be conceivable is to be considered merely possible.[5]

Finally, there is the idea that a "credible world is … a parallel or counterfactual world that … *resembles* aspects of our own world" (55; italics added). This characterizes credibility as a special relationship between a model world and the real world, so does not capture a pragmatic concept. Credibility so defined occupies the same conceptual space with truth. Indeed, my account turns resemblance (that is symmetrical) into truth (that is asymmetrical) in virtue of the modeller using the model as a

[5] One might note that given that plausibility and credibility themselves are constituted by subjective judgements, 'subjective judgements' in "subjective judgements of plausibility and credibility" is redundant if not used to refer to higher-order attitudes.

representation of some real target system (Mäki 2009, 2011).[6] But I consider resemblance and truth to be different from credibility.

On explanation and unification

Reiss discusses Robert Sugden's views describing them as rejecting premise (1) of the paradox and as advocating the idea that false models can be explanatory and that explanatory models must be credible. For some reason Reiss ascribes to Sugden the view that credibility as such guarantees explanatoriness (56). Reiss makes the obvious statement that credibility is not sufficient and then sets out to consider a supplementary property, that of unifying power. Provided unification and explanation go together, credible models can explain.

What exactly is the argument? Reiss asks: "What if economists regard models that are unifying as particularly credible?" (57) It is evident that something like this is indeed the case: given the disciplinary conventions of economics, unification does contribute to credibility (see Mäki 2001; Mäki and Marchionni 2009). However, it does not follow from this that "A credible model is one that is explanatory *because* it is unifying." (57) One can at most conclude that a model is *credible* because it is unifying. The connection between unification and explanation is still missing.

So Reiss asks the right question -- why a more unifying account is more explanatory -- and answers: "Because to no small extent it is the business of science to achieve cognitive economy … the idea that accounts are explanatory to the extent that they are unifying is defensible" (57-58). I am afraid this does not answer the question. It seems this rather provides us with a tautology. How does the pursuit of cognitive economy justify the doctrine that to unify is to explain, and to unify more is to be more explanatory? It doesn't, for a simple reason. The pursuit of unification just *is* (or is part of) the pursuit of cognitive economy, and therefore the appeal to the latter – the

---

[6] This notion of credibility brings this property of models in close contact with another possible property, namely their explanatoriness, provided explanation be understood as a matter of tracing ontic dependencies and locating explanantia in the web of such dependencies. But Reiss chooses not to trace this line. This is another manifestation of his neglect of the details of my account.

pursuit of cognitive economy – simply cannot justify the connection between unification and explanation.

This has implications. Reiss makes the familiar observation that economics's argument patterns are not terribly stringent and therefore standard economic theory is not very unifying in the Kitcherian sense (58-59). From this he concludes that the accounts economists give "are not explanatory qua the unifying power of the argument patterns from which they are derived" (59). But this has little relevance to the issue at hand: given that Reiss has not established the connection between explanation and unification he cannot conclude that failure to unify implies failure to explain.

Yet another unacknowledged implication of introducing unifying power as a criterion of explanatoriness is, again, for the scope of Reiss's argument. Claim (3) of the paradox now becomes restricted to models that unify, in place of any models that explain. This rules out, among others, special-purpose models that might explain.

Conclusion

I conclude Reiss has failed to refute attempts to resolve the paradox of false explanatory models. He has failed to provide an articulate conception of what exactly the paradox is, wavering between different versions of it, drowning them in uncontrolled ambiguities. And he has failed to adequately address the accounts of economic models that might contribute to its resolution.

This is not to deny that there are many open and important issues around models, explanation, and truth in economics. The issues are very difficult, calling for an exercise of rigour and modesty by anyone having the courage to address them. By mobilizing their best skills in addressing them economic methodologists will have a bright full-employment future.

Acknowledgements

Thanks go to Caterina Marchionni for helpful comments on an earlier version. Work on this paper has been sponsored by the Academy of Finland Centre of Excellence in the Philosophy of the Social Sciences.

References

Grüne-Yanoff, Till (2011) "Isolation is not characteristic of models", *International Studies in the Philosophy of Science* 25, 119-137.

Kuorikoski, Jaakko, Aki Lehtinen and Caterina Marchionni (2010) "Economic modelling as robustness analysis", *British Journal for the Philosophy of Science* 61, 541-567.

Mäki, Uskali (2001) "Explanatory unification: Double and doubtful", *Philosophy of the Social Sciences* 31, 488-506.

Mäki, Uskali (2004) "Some truths about truth for economists, their critics and clients", in *Economic Policy-Making under Uncertainty: The Role of Truth and Accountability in Policy Advice*, ed. P. Mooslechner, H. Schuberth, and M. Schurtz. Cheltenham: Edward Elgar. Pp. 9-39.

Mäki, Uskali (2009a) "Models and truth. The functional decomposition approach" in *EPSA Epistemology and Methodology of Science. Launch of the European Philosophy of Science Association*, ed. M. Suarez, M. Dorato and M. Redei. Dordrecht: Springer. Pp. 177-187.

Mäki, Uskali (2009b) "MISSing the world: Models as isolations and credible surrogate systems", *Erkenntnis* 70, 29-43.

Mäki, Uskali (2011a) "Models and the locus of their truth", *Synthese* 180, 47-63.

Mäki, Uskali (2011b) "The truth of false idealizations in modelling" in *Models, Simulations, and Representation*, edited by Paul Humphreys and Cyrille Imbert. London: Routledge. Pp. 216-233.

Mäki, Uskali and Caterina Marchionni (2009) "On the structure of explanatory unification: The case of geographical economics", *Studies in History and Philosophy of Science* 40, 185–195.

Reiss, Julian (2012) "The explanation paradox", *Journal of Economic Methodology* 19, 43-62.

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