Is credibility a guide to possibility? A challenge for toy models in science

Ylwa Sjölin Wirling
University of Gothenburg
University of Manchester
ylwa.wirling@gmail.com

ABSTRACT

Several philosophers of science claim that scientific toy models afford knowledge of possibility, but answers to the question of *why* toy models can be expected to competently play this role are scarce. The main line of reply is that toy models support possibility claims insofar as they are *credible*. I raise a challenge for this credibility-thesis, drawing on a familiar problem for imagination-based modal epistemologies, and argue that it remains unanswered in the current literature. The credibility-thesis has a long way to go if it is to account for the epistemic merits of toy models.

1. Credible toy models

A typical scientific model is an artificial system, composed of a concrete or mathematical structure with an intended interpretation, that scientists use to learn about the world. For example, one can learn about the behaviour of an aircraft by studying a scale model in a wind tunnel; or about meteorological phenomena by manipulating a computer simulation, interpreted in terms of the weather system.

Toy models are very abstract, simple, and highly idealized models. Stock examples include Schelling's checkerboard (1971), Akerlof's 'market for lemons' (1970), Maynard Smith and Price's Hawk-Dove game (1973), and the Lotka-Volterra model of predator-prey dynamics. Toy models of e.g. markets or populations, are very dissimilar to, and exclude factors known to influence, actual markets or populations, and contain idealisations not susceptible to de-idealisation methods.

Some toy models are 'embedded' in an empirically confirmed theory, and their epistemic merit piggybacks on the merits of that theory. But many prominent toy models, including the examples above, are not embedded but *autonomous* (Reutlinger et al. 2018). Scientists evidently find autonomous toy models epistemically useful, despite lack of framework theory, but their epistemic merit presents a puzzle. What, and how, can they teach us about the world?

A number of authors – e.g. Gelfert (2019), Grüne-Yanoff (2009), Reutlinger et al. (2018), Weisberg (2013) – suggest that autonomous toy models give us *modal knowledge*, in particular by supporting objective *possibility claims*.[1] This raises a question: in virtue of what can can toy models help scientists gauge the possible? An influential answer, due to Sugden (2000), and taken on board by several others (e.g. Fumagalli 2016, 437; Grüne-Yanoff 2009, 95; Mäki 2009, 39-40), suggests that they do so insofar as they are *credible*.[2] I refer to this, i.e. the thesis that the credibility of a toy model is evidence of possibility of the model result (appropriately interpreted) as Credibility.

Credibility is a claim in modal epistemology, i.e. about how we learn modal truths. As such, it should be subjected to the same scrutiny as other such claims. I will argue that Credibility faces a challenge analogous to a well-known problem for imagination-based modal epistemologies. This

---

[1] I examine only this claim, and leave open whether they also have other epistemic import.

[2] Sugden additionally claims that we can move from the possibility-conclusion to claims about actual target systems. That claim is irrelevant to my discussion, but see Grüne-Yanoff (2009, 89-91) for criticism.

has hitherto not been acknowledged by philosophers of science. The challenge must be met in order for Credibility to illuminate the epistemic contribution of toy models.

2. Credibility in action: Schelling's checkerboard
Thomas Schelling's much-discussed checkerboard model consists of a two-dimensional grid with two types of individuals, say circular and triangular, initially distributed randomly on the grid. Individuals move on the grid according to one rule: if not at least one-third of individuals on neighbouring squares are of the same type, move to an empty square where that condition is fulfilled. That is, individuals are happy to 'live' in mixed areas, as long as they are not in a strong minority. After relatively few iterations, this generates a 'segregated' grid with triangles and circles in separate clusters.

It is commonly agreed that Schelling's model taught us something about residential segregation in cities. But the model system's high level of abstraction, simplification, and dissimilarity to real cities and populations, makes it hard to see how it could justify conclusions about any actual target. Instead, it is suggested that the model describes a *possible* mechanism behind the actual phenomenon of segregation (it gives a *how-possibly explanation*). In short, Schelling's checkerboard gives us a piece of modal knowledge: it is possible for racial segregation to result despite individual citizens' preference for living in mixed areas. Such modal knowledge can contribute to different projects of interest to science. For instance, Ylikoski and Aydinonat (2014, 30) notes that Schelling's model 'expanded the menu of possible causes', which in turn is important for various scientific explanatory efforts. According to Grüne-Yanoff (2013), Schelling's model was important partly because it helped distinguish the contingent from the necessary – arguably an important part of science. The possibility claim it supported, contradicted the entertained hypothesis that segregation is necessarily a consequence of racist preferences.

According to Credibility, Schelling's model provides evidence for the possibility claim above because it is credible. The toy model's formal structure is interpreted by users as an imaginary world, which, if credible, indicates that the model results is possible.

> [W]e see Schelling's checkerboard cities as *possible cities* (…) We recognize (…) that the model world *could be* real – that it describes a state of affairs that is *credible* (Sugden 2000, 25).

While suggestive, Sugden's notion of 'credibility' needs specification. The best current attempt at elaborating what it means for a model to be credible is due to Grüne-Yanoff (2009) who picks up on Sugden's claim that credibility in models is 'rather like credibility in "realistic" novels' (2000, 25). When confronted with a fiction, one imagines a fictional world, proceeding from the information in the text, but going beyond it by filling gaps, adding detail, drawing out implications. When assessing the fiction for credibility, one assesses this imagined world. Grüne-Yanoff argues that analogously, scientific modellers imagine a model world, proceeding from but going beyond (as above), the model description, and this world can be assessed for credibility.[3]

A fiction can be credible even if its particulars deviate extensively from what the actual world is like. This highlights that the relevant sense of credibility is not believability. What matters is, first, internal coherence: the imagined fictional world must be sufficiently detailed and free of contradiction. Moreover, the development in the fictional world must be plausible *conditional on* the information provided about preferences, environment, etc. In short, what matters is how the whole thing fits together (Grüne-Yanoff 2009, 94–95; Sugden 2000, 26), in some more substantial sense I'll refer to as internal cohesion. The same goes for a model. The individual model assumptions may deviate from what we know about the world, yet the system described by the model may be credible, as long as the assumptions and subsequent development fit together cohesively.

---

[3] Some philosophers argue that models *are* fictions (Godfrey-Smith 2009; Salis forthcoming). Fictionalism goes well with Credibility, but neither Credibility nor my argument below depends on it.

3. The Problem of Modal Epistemic Friction

Credibility attempts to account for the epistemic contribution of toy models. It rests on a claim about an epistemic route to modal truth, namely via the construction and study of *credible* models, and is thus on a par with other theses in the epistemology of modality. I argue that a well-known challenge for imagination-based modal epistemologies arises also for Credibility.

Several modal epistemologists (e.g. Kung 2010; Yablo 1993) have advanced some form of the claim that one can find out whether *p* is possible by attempting to imagine – in some relevant sense – a scenario in which *p* is true. If one has imagined a *p*-scenario, one is (absent defeaters) justified in believing that *p* is possible. A driving assumption behind imagination-based modal epistemology is that we can access what we imagine, and reliably judge whether we have managed to imagine a *p*-scenario. The property of being imaginable acts as an epistemically accessible mediator – as evidence of something we cannot directly judge, namely possibility.

But imagination-based modal epistemologies must deal with the fact that we can imagine impossible scenarios.[4] For instance, it seems we can imagine that silver has a different atomic number than 47. So imagination, or assessment thereof, must be kept in check in order to be a reliable guide to possibility. In a recent paper on this well-known issue, Vaidya and Wallner (2018) call it the problem of modal epistemic friction. It arises for imagination-based modal epistemologies in the form of a question:

> *Under what circumstances*, can I be confident that my ability to imagine a scenario in which I judge *p* to be the case, is evidence that *p* is possible?

A trivial answer is: as long as there are no defeaters. For instance, we have independent reason to think that the imagined *is* impossible in the silver-case, so we won't follow imagination in the wrong direction. But presumably those who appeal to imagination think that in many cases, the imaginative exercise is our best, or only, source of evidence with respect to the prospective possibility. So there must be something which provides epistemic 'friction', making imagination track modal truth.

There are two main avenues of response available. First, one can specify a *mode* of imagination, which *is* plausibly a reliable guide to modal truth (Chalmers 2002). A problem with this strategy is that the new idealised notion of imaginability is often somewhat artificial. Moreover, it might undermine accessibility – a key motivation for appealing to imagination – as it is unclear whether ordinary epistemic subjects can tell a probative kind of imagining from a spurious one (Roca-Royes 2011; Worley 2003). Second, one can require that imagination, or assessment thereof, be constrained by some background knowledge that prevents us from imagining the impossible, or judging that what we have imagined is possible when it isn't. A modal epistemologist taking this avenue must (1) specify what the relevant background knowledge is, and (2) spell out how it is acquired.

Now, Sugden and Grüne-Yanoff take credibility to be a property that both fictions and models can have. Some of Credibility's intuitive pull comes from the assumption that we have at least a tacit grasp on, and an ability to judge, when a fiction is credible. Assessment of toy models is then said to work in the same way. As in the case of imaginability, the property of being credible acts as a mediator – accessible evidence of possibility, something we cannot judge directly.

If we take the analogy with fiction seriously, Credibility faces an analogous epistemic friction challenge. Fictions can describe (physically or metaphysically) impossible worlds. Consider Philip K. Dick's *The Minority Report,* premised on the existence of individuals endowed with precognition, where the police division, with the help of 'precogs', can arrest people about to commit crimes before the act and, sometimes, before the thought has occurred to the would-be-criminal-offender. Or, as Nolan (forthcoming) describes, an early 20th century sci-fi centrally featuring disturbance in the (Lorentzian) ether, caused by a massive generator.

---

[4] See Kung (2016) for several examples.

If credibility just amounts to internal cohesion, there is no obvious reason to deem these fictions incredible. *The Minority Report* neither says nor implies that the form of backwards causation required for precognition is metaphysically impossible. If this is right, credible fictions sometimes describe impossibilities. If fictions and models are credible in the same sense, then some credible toy models may describe impossible systems. So:

> *Under what circumstances* can I be confident that the credibility of a model *is* evidence of the possibility of its result?

Again, something must provide the appropriate modal epistemic friction.

The friend of Credibility also has two avenues of response available. First, she could specify a sense of 'credible' that plausibly tracks possibility. She might deny that the problem of credible but impossible fictions carries over to models, because models are credible in a different sense than fictions, and no credible model describes an impossible system. But if fiction-credibility and model-credibility are different properties, what *is* model-credibility? We were supposed to gain an intuitive grasp via the analogy with fiction, but now that route to the concept is off-limits. Moreover, whatever model-credibility is, why think we are good judges of whether a model world has this property? The analogy with fiction gave us reason to think we have epistemic access to model-credibility too, but once the two notions of credibility are pulled apart, that reason is lost.

The other avenue involves restricting credibility with appropriate background knowledge, and say that our judgement that a toy model is credible is evidence of possibility just in case the credibility-assessment is informed by the right background knowledge. This seems to generate the right result for the fictions described above. If one *does* feel any pull to say that those fictions are incredible, this is likely because e.g. scientific knowledge informs the credibility-assessment. The developments in the fictions are incredible, conditional on that knowledge. In contrast, what reason would a reader of the ether-fiction, alive prior to, or otherwise unaware of, certain scientific progress, have to find the fiction incredible? None, it seems – and the apparent credibility of the story might lead her astray with respect to possibilities.

Grüne-Yanoff writes that the judgements of credibility are 'driven by empathy, understanding, and intuition' (2009, 94–95), but says nothing of whose intuitions, or understanding of what. Appealing to possession of appropriate background knowledge is one way to fill that out. The background knowledge in question can be tacit, but one must possess it in order to rely on it. As with imagination-based modal epistemologies, this avenue brings the further challenges of (1) specifying what the relevant background knowledge is, and (2) spelling out how that knowledge is acquired. They currently remain unaddressed.

4. What background knowledge?
Since the aim is objective and not epistemic possibility, one should not hold fixed everything one knows about the actual world. But what background knowledge needs to inform credibility-assessments? Notably, it cannot be required that one justifiably accepts a theory of the relevant phenomena[5] (e.g. segregation) since Credibility is supposed to account for the epistemic import of *autonomous* toy models that are, by definition, not embedded in theory. Indeed, the fact that toy models are often used to further research of phenomena for which there are no systematic theories, adds an extra wrinkle to the need for background knowledge as friction-provider.

Importantly, what counts as a satisfactory reply to the modal epistemic friction challenge, depends on the relevant notion of modality. There are more and less restricted objective modal notions, and any given *p* may be possible in some senses but not others. Modal epistemologists are often interested in the least restricted notion, metaphysical possibility. For science, a prominent one is physical possibility (possibility given the actual laws of nature), but there are others, e.g. mathematical, biological, and practical possibility. Since at least some of these clearly come apart

---

[5] This rules out adapting a modal epistemology like Fischer's (2017).

in scope, different background knowledge is required to inform credibility-judgements depending on what kind of possibility they are supposed to be a guide to.[6]

As Verreault-Julien (2019) points out, not all models are meant to support possibility claims of the same kind. This is presumably true also of toy models. At one point, Sugden writes that that credibility-judgements are conditioned on 'what we know (…) about the general laws governing events in the real world' (2000, 25). This passage is interpreted by others (e.g. Grüne-Yanoff 2009, 92) as the idea that credibility involves compatibility with the laws of nature. A corresponding response to (1) is that judging a model to be credible is evidence of possibility, just in case that judgement is informed by knowledge of the laws of nature.

Insofar as toy models aim to justify claims of physical possibility, this is a potential answer to (1) – it is plausible enough that credibility-judgements restricted by knowledge of the laws of nature, would be suitable guides to *physical* possibility. But while I am happy to grant that some toy models aim for physical possibility, far from all do. Importantly, the stock examples in the literature don't. For instance, it is implausible that Schelling's model is supposed to show that segregation is *physically* possible despite individuals' preferences for mixed areas. Nor does it seem right to say the Hawk-Dove model supports the claim that it is *physically* possible that a trait like restraint in combat results from individual selection alone. Likely more restricted modalities are of interest here. I submit that this generalises: knowledge of natural laws will not account for very much of the epistemic work that toy models are supposed to be doing, as much of the relevant modelling is from economics, biology, or computational sociology, where physical possibility does not seem to be at issue.

But if so, it is insufficient if these models are credible in the sense informed by nomological knowledge, because that indicates possibility only in a less restricted sense. This comes out especially in relation to rebuttal of necessity hypotheses, one task for toy models. Compare: showing the *metaphysical* possibility of superluminal travel does nothing to rebut the claim that it is *physically* necessary that nothing with mass travels faster than the speed of light. Some other knowledge needs to bear on credibility-assessments in order to support the right possibility claims.

It won't do to suggest that credibility-assessment of these models should be informed by knowledge of more specific laws that govern the relevant domain. The special and social sciences feature remarkably few law-like principles, that also do not hold without exception (Gelfert 2016, 63; Grüne-Yanoff 2009, 92).

This in turn highlights that it is not clear what sense(s) of possibility is relevant in fields like biology, economics, or sociology. Without a grasp of that, we cannot say what knowledge is required to restrict credibility-assessments that allegedly track it. Imagination-based modal epistemologies typically aspire to be guides to metaphysical possibility, which presents challenges of its own. But philosophers of science need to say more about the kind(s) of possibility knowledge toy models afford – especially since possibility in those senses cannot be understood in terms of laws, analogously to physical possibility. So, we add an item to the to-do list for those who appeal to Credibility in elucidating the epistemic merit of toy models: (0) identify the sense(s) of possibility relevant to toy modelling practices.

5. Conclusion

According to some philosophers of science, Credibility elucidates the epistemic contribution of toy models. But Credibility faces an analogue of the epistemic friction challenge for imagination-based modal epistemologies. In order to meet it, and account for the epistemic import of toy models, friends of Credibility must (0) identify the sense(s) of possibility relevant to toy modelling practices;

---

[6] Here is an interesting parallel with credibility in fiction, which is more complex than what Sugden and Grüne-Yanoff makes it out to be. Fictions may be credible in different ways: scientifically (aspired to by 'hard science fiction' writes), psychologically, or morally. Presumably, different beliefs (e.g. moral, scientific) are brought to bear in the credibility-assessment, depending on the relevant sense of credibility.

(1) specify what kind(s) of background knowledge should inform credibility-assessment; and (2) spell out how that knowledge is acquired.[7]

References
Akerlof, G. A. 1970. The Market for "Lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* 84 (3), 488–500.
Chalmers, D. 2002. Does conceivability entail possibility? In *Conceivability and Possibility*, ed. T. S. Gendler, & J. Hawthorne, 145–200. Oxford: Clarendon Press.
Fischer, B. 2017. *Modal Justification via Theories*. Cham, Switzerland: Springer International.
Gelfert, A. 2016. *How to do science with models: A philosophical primer*. Cham, Switzerland: Springer International.
Gelfert, A. 2019. Probing Possibilities: Toy Models, Minimal Models, and Exploratory Models. In *Model-Based Reasoning in Science and Technology*, ed. M. Fontaine, C. Barés-Gómez, F. Salguero-Lamillar, L. Magnani and Á. Nepomuceno-Fernández, 3–19. Cham, Switzerland: Springer International.
Godfrey-Smith, P. 2009. Models and fictions in science. *Philosophical Studies* 143, 101–16.
Grüne-Yanoff, T. 2009. Learning from Minimal Economic Models. *Erkenntnis* 70 (1), 81–99.
Grüne-Yanoff, T. 2013. Appraising Models Nonrepresentationally. *Philosophy of Science* 80 (5), 850–61.
Kung, P. 2010. Imagining as a Guide to Possibility. *Philosophy and Phenomenological Research* 81 (3), 620–63.
Kung, P. 2016. You Really Do Imagine It: Against Error Theories of Imagination. *Noûs* 50 (1), 90–120.
Maynard Smith, J. and G. R. Price. 1973. The Logic of Animal Conflict. *Nature* 246 (5427), 15–18.
Nolan, D. Forthcoming. Impossible Fictions Part I: Lessons for Fiction. *Philosophy Compass*.
Reutlinger, A., D. Hangleiter, and S. Hartmann. 2018. Understanding (with) Toy Models. *British Journal for the Philosophy of Science* 69 (4), 1069–99.
Roca-Royes, S. 2011. Conceivability and de re modal knowledge. *Noûs* 45 (1), 22–49.
Salis, F. Forthcoming. The New Fiction View of Models. *British Journal for the Philosophy of Science*.
Schelling, T. C. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1(2), 143–86.
Sugden, R. 2000. Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology* 7 (1), 1–31.
Vaidya, A. J., & Wallner, M. 2018. The epistemology of modality and the problem of modal epistemic friction. *Synthese,* doi:10.1007/s11229-018-1860-2.
Verreault-Julien, P. 2019. How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science Part A* 73, 22–33.
Weisberg, M. 2013. *Simulation and Similiarity: Using models to understand the world*. Oxford: Oxford University Press.
Worley, S. 2003. Conceivability, possibility and physicalism. *Analysis* 63 (1), 15–23.
Yablo, S. 1993. Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research* 53 (1), 1–42.

---

[7] Thanks to Till Grüne-Yanoff, Frans Svensson, participants at the *Metaphysical Explanation III* workshop in Lund, and two anonymous referees from this journal for insightful comments.

Ylikoski, P. and N. E. Aydinonat. 2014. Understanding with theoretical models. *Journal of Economic Methodology* 21 (1), 19–36.