Through the Prism of Modal Epistemology: Perspectives on Modal Modelling

Ylwa Sjölin Wirling (University of Gothenburg) & Till Grüne-Yanoff (KTH Royal Institute of Technology)

Several philosophers of science have recently taken care to highlight different modelling practices where scientific models primarily contribute *modal* information. While examples of this now abound, comparatively little attention is being paid to question of *under what conditions*, and *in virtue of* what, models can perform this modal epistemic function. But there are some preliminary attempts. In this paper, we review these attempts to spell out and explain the success conditions of modal modelling, through the lens of some dominant themes in modal epistemology. The aim is to more clearly expose the respective justificatory strategies of these accounts, and secondly, to identify lacunae where further work is needed.

1. Introduction

Philosophers of science recently have begun to investigate modelling practices from a modal perspective, for at least two reasons. First, many scientists explicitly describe their modelling practices with modal terms like possibilities, necessities or dispositions. Secondly, even where modelers do not use explicit modal language, philosophers of science have sometimes offered a modal interpretation, in order to address philosophical issues that have been left unanswered by prior accounts of scientific modelling. The picture is further complicated by the different roles that modal claims have been identified to play in scientific modelling - ranging from the exploration of possible scenarios of actual targets through the rejection of necessity claims to the investigation of counterfactual objective possibilities for their own sake (for recent discussion of some of these roles, see Gelfert 2019).

Despite their diversity, all these accounts face questions about the epistemology of modal claims. If scientific modelling practices deal with modals, or if they are philosophically reconstructed to trade in modal claims, then *in virtue of what* can models perform this function, and what are the conditions for their success? This epistemic question for modal modelling has received comparatively little attention in the philosophy of science literature, and answers remain fragmented and tied to the documentation and analysis of the specific practices or reconstructions mentioned above.

Our purpose in this paper is to provide an improved and more systematic grasp on that epistemic question for modal modelling, by analyzing three existing accounts of modal modelling from the philosophy of science. The paper is structured as follows. Section 2 summarizes various modal aspects in scientific modelling, discussed in the recent philosophy of science literature. Section 3 specifies the epistemic question for modal modelling that is the central concern of this paper. Section 4 introduces three key themes in modal epistemology, the study of how and in virtue of what one can come to know modal claims. Section 5 analyzes three modal modelling accounts – Sugden's (2000) credibility account, Massimi's (2019a) physical conceivability account, and Batterman and Rice's (2014) universality account – using the modal epistemology themes as interpretative tools. Section 6 concludes.

2. What is modal modelling?

With "modal modelling" we mean modelling practices that aim at delivering *modal* information, in particular about what is possible, what would be the case under counterfactual circumstances. This is in contrast to modelling that aims at delivering information about what actually is, was, or will be the case. Of course, in practice, there is no sharp separation between these two aims. The reason scientists are interested in acquiring modal information will typically consist in its contributing to our understanding of the actual world. Nevertheless, one can conceptually distinguish modelling practices by the immediate aim they are employed for, even if the thus acquired results are then employed to achieve some further objective. Modal modelling practices thus are identified by aiming at delivering modal information as their immediate results.

Under the "modal modelling" flag we also include modelling practices that do not explicitly aim at modal information, but of which philosophers of science nevertheless offer such a modal interpretation, in order to address philosophical issues or conceptual challenges that are left unanswered when standard modelling accounts are applied to the practice in question.

The most prominent example of modal modelling is perhaps connected with *how-possibly explanations* (HPEs). There is no general consensus on how to characterize HPE practices, but most philosophers seem to agree at least that (i) they involve modal claims, and (ii) models play a crucial role in supporting HPEs (see e.g. Bokulich 2014; Grüne-Yanoff 2009; 2013; Reutlinger et al. 2018; Rohwer and Rice 2013; Verreault-Julien 2017; 2019; Weisberg 2013, chapter 7; Ylikoski and Aydinonat 2014). As we see it, this is a clear case of modal modelling: scientists draw the conclusion that such-and-such is possible (or should be re-interpreted as doing so), on the basis of modelling. There are several different kinds of modelling practices used to support HPEs, and different purposes for which the resulting HPEs are employed in a broader scientific context, that have been highlighted by philosophers of science in recent years.[1]

For instance, *exploratory* or *hypothetical* modelling practices are important in situations where a putative target lacks theoretical descriptions consisting of shared and widely accepted principles and concepts. Such situations might arise for different reasons – for example, investigations of the putative phenomenon began only recently, empirical evidence is scarce, or the subject matter in question integrates widely divergent features (e.g. social and physical properties in many engineering tasks). These models can presumably have different epistemic functions in such a context, one of which is providing HPEs that refute necessity or impossibility claims (Grüne-Yanoff 2013), that serve as proofs of principle (Gelfert 2018), or that otherwise help delineate the space of what is possible and impossible regarding a putative or actual phenomenon (Massimi 2019). In some cases, the modelling focuses on phenomena that are known to not be (currently) actual. Some are "explanations in search of observations" (Sugden 2011) offering representations of possible properties of possible explanada, for the purpose of understanding such phenomena in case such explananda should become actual. Other disciplines that study unactualized possibilities include synthetic biology, where models are used to represent minimal cells and alternative genetic systems, even though such targets might turn out to be only partially realizable or prove outright impossible (Knuuttila and Koskinen 2020).

Modelling that supports HPEs can also play other roles, some of which are more common in areas of inquiry where the state of knowledge is more advanced. For instance, model-supported HPE can contribute to a deeper scientific understanding of the studied phenomena. They do so by enabling their users to draw correct counterfactual inferences, thus providing a core ingredient for successful scientific explanation. This is roughly how Ylikoski and Aydinonat (2014) describe the epistemic contribution of Schelling's (e.g. 1971) checkerboard model, and Verrault-Julien (2017) also suggests that certain economic models work this way. Batterman and Rice (2014) argue that what they call minimal models in e.g. physics and biology are "holistically idealised": they severly distort their targets, where this distortion is crucial to the explanatory contribution of the model, in terms of counterfactual information about the target. In many (although not all) cases, these modelling practices are not overtly modal, but philosophers of science argue that they should be re-interpreted in modal terms in order to account properly for the epistemic contribution of the models in question. This is typically because e.g. minimal models or toy models apparently do not satisfy standard criteria of representational accuracy. If those models cannot be shown to represent actual targets accurately, these authors ask, what kind of epistemic functions can such models then

[1] The lines between these modelling practices and what they (aim to) establish are not sharp ones, and we are not trying to offer a neat taxonomy.

play? Their answer: these models represent how the target possibly is, or how the target could possibly have been produced.

Finally, not all modal modelling need be understood in terms of HPE. For instance, Nguyen (forthcoming) explicitly contest the idea that toy models like Schelling's checkerboard model support possibility claims in the form of HPEs. Nevertheless, Nguyen's interpretation of these modelling practices is also modal in nature. He suggests that facts about the toy models should be translated into claims about the real world that are a.) less specific than the model facts (e.g. to qualitative trends from real values) and b.) ascribe a *capacity* or *susceptibility* to the target. For example, in Akerlof's (1970) "market for lemons" model asymmetric information *prevents car trades from occurring*, despite the fact that at any given price there are sellers willing to sell their car and buyers willing to buy it. When properly interpreted in terms of a given target, the claim supported by the model is something like: an asymmetric information state in this (particular, real world) market *increases the market's susceptibility to fail* to reach Pareto-efficient equilibrium. This is a supposedly true claim about an actual target. But, we would like to point out, it is a modal claim in the sense that it ascribes a *de re* modal property – a susceptibility, which is a form of disposition – to a particular entity.

In sum, several authors recognize that models are employed to perform the epistemic function of supporting objective modal claims in various forms, and there are several detailed case studies describing such modelling practices – we refer the reader to the papers referenced above for these. These accounts of modal modelling are naturally assessed along the following two dimensions: (i) for *accuracy in describing* the roles that these models play in scientists' practices, and (ii) for *conceptual improvement* on the reconstruction of modelling practices that present challenges for standard modelling accounts. But if it is right that models *are* used in these ways to support or draw modal conclusions, this also raises the question of *in virtue of what* models can perform this function. This third dimension of assessment for modal modelling accounts is what we call *the epistemic question* for modal modelling, and it is what we focus on in this paper.

3. The Epistemic Question for Modal Modelling

We start from the assumption that with modal claims, just as with non-modal claims, there is generally a fact of the matter as to whether they are true or false. Exactly in virtue of what modal claims are true or false is a question that keeps exercising philosophers who do metaphysics of modality, and on which we will not take a stand here. The important point is that if the truth-value of modal claims is not a matter of opinion, then one can be right, or one can be wrong, when making a modal claim. Thus, in order to be *justified* in claiming that e.g. such-and-such is possible, one needs to give reasons for such claims, e.g. by citing relevant evidence or describing pertinent inference procedures.

Philosophers of science, as we saw in the previous section, have highlighted that scientists often appeal to *models* when making modal claims. For instance, economists claim that the invisible hand hypothesis describes a possible scenario, with reference to the Arrow-Debreu model (Arrow and Debreu, 1954). That is, the fact that an invisible hand scenario has been successfully modelled provides reasons for the claim that such a scenario is possible (Verreault-Julien, 2017). Assuming that there is a fact of the matter regarding the possibility (or not) of what the invisible hand hypothesis describes, and that economists are justified when they are claiming, with reference to the Arrow-Debreu model, that it *is* possible, there must be something about the Arrow-Debreu model in virtue of which it *does* provide good reasons for this claim. What is that something? More generally, what must models or modelling practices be like in order to constitute good reasons for modal claims?

Many philosophers of science hold that the content of science must, in order to be successful in e.g. providing explanation, be (approximately) true (Kitcher, 1981, p. 519; Strevens, 2008, p. 71; Woodward, 2003, p. 203). This places constraints on models, insofar as they are to contribute to science's uncovering of information about the world. They must satisfy some "standards of accuracy" (Frigg and Nguyen, 2016a), or be "faithful epistemic representations" (Contessa, 2007). There must be some conditions for when a model is a *reliable* scientific tool. As we saw in the previous section, the suggestion that a certain modelling practice is to be interpreted modally is often motivated by the idea that the models in question do *not* meet the standard accounts of such accuracy, e.g. model-target similarity, or isolation of relevant causal factors.

What we emphasize is that modal modelers are not off the hook: models, if they are to contribute to science by teaching us about *modal* truths in some form or other, must reasonably be taken to satisfy some modal counterpart of such requirements. In many cases, especially when the aim of the account is reconstruction of problematic cases, these requirements will presumably be different from the standard ones. But there must be *some* conditions in virtue of which the result of a modelling exercise provides reasons that justify asserting a modal truth. This *epistemic question* arises equally for all modal modelling, whether the modal claim supported by the model is a possibility claim (in the form of an HPE, or otherwise), an impossibility claim, a dispositional claim, or a counterfactual conditional.

4. Themes from the epistemology of modality

Modal epistemology is the philosophical field investigating how we can come by knowledge of modal truths – e.g. possibilities, necessities, counterfactual claims. Much of what modal epistemologists do consist in presenting putative justifications for modal claims, and examine them with respect to something like the epistemic question stated above. For instance, some think that intuition provides justification for modal claims (e.g. Bealer 2002). The epistemic question is then in virtue of what, and under what conditions, intuition can play this cognitive role. Spelling that out is a crucial part of bolstering the claim that intuitions can justify modal claims.

Modal epistemology has not yet investigated scientific modelling as a source of modal justification. But few modal modelling accounts have had something preliminary to say in response to the epistemic constraint, without explicitly engaging the modal epistemology literature. We conjecture that viewing modal modelling accounts through the lens of existing work in modal epistemology will be beneficial to the continued development of these accounts. To this end, we will presently introduce three key *themes* from modal epistemology: imagination, background theory, and similarity judgements. These themes are non-exclusive strategies for understanding modal justification: as we will see later, philosophers of science sometimes have appealed to more than one theme simultaneously in order to account for modal modelling. They are broadly conceived and encompass many, although not every single, main strand in the epistemology of modality literature. Notably, the themes cut across the distinction between rationalist and non-rationalist modal epistemologies, in the sense that both rationalists and non-rationalists have used them when constructing their accounts of modal knowledge – but they differ on whether they take e.g. the relevant strategy to be *a priori* or not. We recognize that philosophers of science will presumably be interested exclusively in the non-rationalist versions of these modal epistemologies, but insofar as the central themes go, rationalist modal epistemologies may have insights to offer too.

Many philosophers have assumed that the *imagination* – or the ability to conceive – is centrally involved in the way we actually, and justifiably, arrive at modal beliefs. This idea is both historically salient – going back to the writings of Descartes and Hume – and widely thought to be true to the phenomenology of (much) actual modalizing: what we often do when we consider whether something would be possible – say, whether the new couch can possibly fit through the doorway – is to try and imagine it. Several modal epistemologies that rely centrally on imagination suggest that it primarily supports claims about what it possible: if one can conceive of, or imagine, a scenario in which *p* is the case, one is justified in holding that *p* is possible (see e.g. Yablo 1993; Chalmers 2002; Kung 2010). For instance, if one is interested in whether it is possible for there to be talking donkeys, one would attempt to conceive of a scenario in which it is true that there are talking donkeys. If one succeeds in this, one would (under the right circumstances) be justified in believing that talking donkeys are possible. Other accounts, most notably Williamson (2007), connect imagination with the evaluation of counterfactuals, i.e. claims of the form If *a* were the case, then *b* would be the case. According to Williamson, knowledge of possibility is downstream from counterfactual knowledge, and imagination is importantly involved in acquiring the latter. For instance, what would happen to a rock sliding down a slope if the bush that actually stopped it halfway down had not been there? For an answer, one supposes the antecedent – that there is no bush on the slope – and develops the scenario from there on, in the imagination – e.g. the rock sliding down the slope, past the place where in reality there is a bush, further down the slope, ending up in the lake below. One thereby comes to know that if the bush had not been there, the rock would have ended up in the lake. Possibility knowledge is a result of such counterfactual development (in imagination): "we assert ◇A when our counterfactual development of A does not robustly yield a contradiction" (Williamson 2007, p. 163). In this case, we can conclude that it is possible that the rock could have ended up in the lake.

Any modal epistemology that assigns a central role to the imagination have to deal somehow with the pressing worry that imagination itself appears too liberal. Basically, the problem is that we can imagine things we know are in fact impossible – see Kung (2016) for a host of examples – so unchecked imagination cannot be a reliable source of modal justification.[2] There needs to be a way of avoiding widespread modal error, and also an explanation of why imagination *can* (under the right circumstances) be trusted as a guide to modal knowledge, in the face of the fact that imagination easily reaches also into the impossible.

This brings us to the second theme we wish to highlight: *background knowledge.* In response to the problem of restricting imagination, it is common to suppose that the imagination must be somehow constrained by some appropriate background knowledge which prevents us from imagining the impossible, or from judging that what we have imagined is possible when in fact it isn't.[3] But background knowledge also features in accounts that make no mention of the imagination, so the two themes are independent. What kind of background knowledge is needed in order for one to make justified claims about possibility?

That depends on the relevant sense of "possible", but it has generally been assumed that one needs to have knowledge of that which determines or restricts the relevant notion of possibility (or necessity).[4] Modal epistemologists have mainly been concerned with how we come to know about what is necessary in the strongest objective sense, and what is possible in the least restricted objective sense. This is commonly referred to as metaphysical necessity and possibility, respectively. Another familiar form of modality is *natural* or *nomological* modality, which is typically defined in terms of the actual laws of nature. Philosophers disagree over whether natural and metaphysical

---

[2] An initial reaction to this problem was to say that we cannot really imagine impossibilities – at most it *seems to us* that we imagining something impossible, while we are in fact imagining a possible situation that we mistake for and misdescribe as the impossible one (Kripke 1980). See Kung (2016) for a convincing case against both such error-theoretic approaches and Yablo's (1993; 2006) related claim that we cannot imagine something we antecedently know to be impossible.

[3] An alternative strategy is to specify a more particular species of imagination which *is* plausibly a reliable guide to modal truth. A common complaint about this approach is that many of the more specific senses of "imaginable" or "conceivable" that philosophers try to construe as reliable guides to objective possibility are artificial, in the sense that it is highly doubtful whether human cognizers can normally conceive in the required sense, and even if they can it is doubtful that they can distinguish the mode of conceiving or imagining that is conducive to modal truth from the one(s) that is not. For an interesting instance of this approach that seeks to avoid such criticism, see Kung (2010). Rather than distinguishing species of imagination, he notes that all imaginings are made up of two kinds of content: qualitative content, which is what we phenomenologically experience (e.g. "see") in the imagining that we conjure up, and stipulative content which is a form of "labelling" of the qualitative content. As far as imagination as an independent source of justification is concerned, only the qualitative content of an imagining can do any work. If one imagines a scenario *s* in which they judge it to be that case that *p*, one is only justified in believing, on the basis of imagining *s*, that *p* is possible if it is the qualitative content of *s* that makes it intuitive to one that *p* is the case in *s*.

[4] As we argue in Sjölin Wirling and Grüne-Yanoff (ms), it matters greatly, as far as the epistemic question is concerned, whether the possibility in question is objective or epistemic. Modal epistemology is almost exclusively concerned with objective modality of various kinds, but in science, both epistemic and objective possibility tend to be in play.

modality come apart or not, i.e. over whether the laws of nature are necessary in the strongest sense or not, and thus over whether there are some things that are possible while not compatible with the laws of nature. We do not have strong views on this, but although we recognize that necessitarianism about the laws of nature is controversial, we will continue our discussion as if it is correct. The reason for this is that we are concerned here with modal truths of interest to science, and it is reasonable that scientists are primarily – if perhaps not entirely exclusively – interested in that which is compatible with the actual laws of nature. Although the modal epistemology claims we discuss in what follows have often been presented under the assumption that necessitarianism is false, what they suggest is workable also in the context of necessitarianism (or so we believe). In any case, the answers in the modal epistemology literature concerning the relevant background knowledge form three main groups.

First, a popular view in recent years is that knowledge of what is possible depends on having access to essentialist, or *constitutive knowledge* – that is, knowledge of what is constitutive of being a certain (kind of) object or property. Roca-Royes (2011a, 2011b), Vaidya & Wallner (2018) and Tahko (2012) all argue that both conceivability/imagination based modal epistemologies, and Williamson's imaginary evaluation of counterfactuals, presuppose that the epistemic subject has access to constitutive knowledge (but that these modal epistemologies fail to elucidate how that knowledge is acquired). For some modal epistemologies explicitly based on constitutive knowledge, see e.g. Lowe (2012), Mallozzi, (2018), Jago (2018).

Next, it is natural to think that *knowledge of laws* have an important role to play in modal epistemology. For instance, physical possibility is naturally defined (at least partly) in terms of compatibility with the actual laws of nature. It thus plausible that knowledge of the laws of nature is very helpful to drawing justified conclusions about what is physically possible. One way (although not the only one) to utilize such knowledge in modal epistemology is to require such knowledge to be held fixed when one attempts to conceive of a scenario, or develop a counterfactual in the imagination. The corresponding view for non-necessitarians is that metaphysical modality is (partly) determined by the "laws of metaphysics", and that knowledge of metaphysical modal truth presupposes knowledge of the laws of metaphysics. See Kment (2014) for talk of metaphysical laws in relation to modality, epistemology, and e.g. Schaffer (2017) for discussion of metaphysical laws more generally.

Finally, some have argued that one needs to have a *theory* concerning the relevant phenomena or entities that one is seeking modal knowledge of. For instance, Fischer (2017) proposes that a modal claim is justified if its truth is implied by a scientific theory that is itself justified. For instance, I am justified in claiming that it is possible for cells to evolve on the basis of something other than RNA/DNA, just in case I am justified in accepting a scientific theory which implies that it is possible for cells to evolve on the basis of something other than RNA/DNA, and I base my claim on that theory. This makes modal knowledge downstream from scientific knowledge more generally. Similarly, Bueno and Shalkowski (2014) argue that we arrive at modal knowledge by investigating the relevant properties and objects in question by *both* scientific and common-sense means. Through common-sense observations, we learn that everyday objects have certain properties only contingently, because we have observed them to lose these properties under changing conditions. In scientific practices, we deepen such observations by systematically varying interventions on and background conditions of these objects. Furthermore, we also seek to connect objects' macro-properties to various micro-level properties, thus expanding the application of theoretical knowledge across properties of different levels. One is justified in concluding that *p* is possible, on this view, on the basis that nothing in the body of relevant theoretical knowledge suggests that *p* is not possible.

Notably, knowledge of laws, constitutive knowledge, and (scientific) theory need not be mutual exclusives. There are different views on how the three relate, and how they relate to the relevant notions of possibility and necessity. Presumably many justified scientific theories may incorporate constitutive knowledge, or knowledge of laws, but they need not. Conversely, one may have some knowledge of laws, or of some constitutive facts, but not have a fully developed and justified theory. Moreover, some necessitarians – e.g. dispositional essentialists like Bird (2007) and Ellis (2001) – think there is a very close connection between essences and the laws of nature, but again this is not obligatory.

The third and final theme in modal epistemology that we wish to highlight is the idea that *similarity-judgements* can justify possibility claims, a claim most prominently advanced by Roca-Royes (2017). In a nutshell, she suggests that one can draw justified conclusions about what is possible for some individual entity *e* on the basis of knowledge of what is actually the case with other, distinct entities that are relevantly similar to *e*. For instance, I know that it is possible for this table to break because I have seen other tables, relevantly similar to this table, actually break. Background knowledge plays a part here too, but the similarity-account originates in an explicit effort to present an epistemology of possibility, which does *not* presuppose that the epistemic subject has access to constitutive knowledge or a full-blown, justified theory of the relevant entities. Nor does it rely on imagination itself as conducive to modal truth, so it warrants separate mention. All that is required for knowledge of nonactual possibility here, is knowledge of actual token events involving actual entities, and the ability to reliably judge that certain entities (the targets of the prospective possibility judgements) are relevantly similar to certain other entities. At bottom, this is just an application of a weaker form of induction. Knowing that this key unlocked my office door on each visit last week, I legitimately conclude it will unlock it today, too. Similarly, I conclude that it is possible for two particular chimpanzees to mate – even though they might never actually mate – on basis of knowing e.g. that other relevantly similar mammals have actually been able to mate.

5. Attempts to answer the question

The epistemic question for modal modelling is: *in virtue of* what can models perform the epistemic function of supporting modal claims, and what are the conditions for their success? A few of the philosophers of science who have documented and analyzed modal modelling practices (sketched in section 2) have offered answers to this epistemic question. In this section, we examine three such answers, specifically through the lens of the three central themes in modal epistemology presented in section 4. In the process, we also point out questions yet unanswered and indicate ways how the respective approaches can be advanced with respect to these lacunae.

5.1. Credible worlds

Robert Sugden (2000) suggests that we can learn from some models – in particular, toy models in economics – in virtue of the fact that these models describe *credible worlds*. More exactly, economic toy models are formal structures that are interpreted by their users as imaginary worlds or scenarios. If the model world is credible, that is a reason to think that the model result (or some equivalent thereof) is possible.[5, 6] That idea has been taken on board as plausible by a number of other authors. For instance, Grüne-Yanoff (2009, 95) writes that "The credibility of a minimal model establishes that it depicts a possible world"; Mäki (2009, 39-40) makes a similar point; and Fumagalli (2016, p.

[5] Sugden's own view is that we can *also* draw general, inductive conclusions about how things are in the actual world, on basis of these credible models. We set the latter part of his view aside.

[6] The model scenarios *themselves* are often clearly impossible in the sense they could never occur in anything like the actual world, since they involve assumptions that are not just false but impossible (e.g. infinite population size and perfect correlation between winning the resource and reproductive success, in the Hawk-Dove model; limitless supplies of food for prey and infinite appetite of predators in the Lotka-Volterra model). At the very least this indicates that one cannot equate the credible model world with a possible world. That a certain model world is credible doesn't mean that model world is possible, but at most that some part of it is possible, or that something else is possible. One way to deal with this problem is to apply accounts of model idealization to modal modelling. From such perspectives, the model is an idealizing representation of a credible scenario. The idealizing assumptions are motivated by tractability considerations, but they are not themselves judged to be credible. What the model user then needs to show is that these assumptions do not substantially influence the scenarios – they could be replaced without affecting the relevant parts of the scenario – and that they are excluded from conclusions about possibility.

437) concedes that "Considerations of models that are credible (…) may enable scientific modellers to acquire epistemically informative insights about the *possible worlds* posited by these models".

While suggestive, the notion of 'credibility' in Sugden's original paper is in dire need of unpacking. At one point, Sugden writes that credibility in models is "rather like credibility in 'realistic' novels" (2000, p. 25), and Grüne-Yanoff (2009) picks up on this analogy with fiction, in his attempt to elaborate on what is means for a model to be credible. This emphasizes the role of the imagination, which plays a central role in most accounts of fiction. The idea here is that when we read a novel, we imagine a fictional world, proceeding from the fictional text but going beyond it by adding detail, drawing out implications and filling in gaps. It is this imagined world that we assess when we consider whether the novel presented a credible story or not. Analogously, scientists employing a model imagine a model world, proceeding from but going beyond (as above), the model description (Frigg & Nguyen 2016b). This imagined model world is then assessed for credibility. So imagination plays an important role in *generating* the system to be assessed for credibility.

Unsurprisingly, the credibility account of modal modelling therefore seems to face the challenge of reining in imagination. Clearly, the mere fact that one succeeds in imagining a model world can be imagined does not guarantee that the results deriving from it are possible, since we can easily imagine the impossible – a familiar fact from the epistemology of modality. So the justificatory power, on this account, lies with the credibility judgement.

But when do we judge a fiction, or a model, to be credible? Grüne-Yanoff stresses that many particular features of the imagined model/fictional world can deviate extensively from what the actual world is like, yet the imagined world can be judged credible. What matters is, instead, on the one hand, internal coherence. That is, the model/fictional world, imagined on the basis of the model description/fictional text, must be sufficiently detailed and free of incoherent or contradictory assumptions and implications. On the other hand, the development in that imagined model/fictional world must be judged to be plausible *conditional on* the background information provided about e.g. preferences, environment, and so on. These conditional judgements are, he writes, "driven by empathy, understanding, and intuition" (2009, pp. 94-95).

Now, if the judgement that a model world is credible is to justify one in taking the model result to be possible, one must presumably be a *reliable* judge of when an imagined model world *is* credible. That is, one must not issue a lot of erroneous judgements, taking for credible models that in fact have impossible results. Grüne-Yanoff's claim that credibility-judgements are guided partly by "understanding", indicates that only the assessment of a *competent user* of the model will do. That is, someone with the appropriate background knowledge.[7] With this addition, the full credibility thesis can be stated as follows: That a certain scenario is successfully modelled is a good reason to think that the scenario is possible, just in case the model is credible. A model is credible just in case the model world is internally coherent and a competent user of the model would judge the development in the model world to be intuitively plausible, conditional on the model setup.

As in modal epistemology, two questions arise here: what is the appropriate background knowledge, and is it likely that users of the model typically possess it?[8] Presumably, competence with credibility judgements is relative to discipline, e.g. different background knowledge is required to assess an economic model and a biological modal for credibility, respectively. But it is reasonable to suspect that in general, the candidates will be same as those floated in general modal epistemology: knowledge of laws, constitutive knowledge, or justified (scientific) theory.

Interestingly, to the extent that issue has been discussed, answers have conformed to this prediction. Sugden at one point suggests that credibility requires compatibility with the "general laws governing events in the real world" (2000, 25).[9] Mäki suggests that credibility requires

---

[7] This is also reminiscent of solving the runaway imagination problem by appealing to "ideal conceivers", which is Chalmers' (2002) strategy.

[8] For a more elaborate discussion of this challenge to the credibility account, see Sjölin Wirling (ms).

[9] See Grüne-Yanoff (2009) for some problems with this proposal in the context of economic models.

compatibility with what he calls a "*www constraint*" (Way the World Works" (2009, p. 39). It is not clear from Mäki's brief discussion exactly what that involves, but the basic idea invokes phrasings reminiscent of appeal to constitutive knowledge. Models that violate this constraint are to be rejected because they violate *the very nature of the (kind of) system* (e.g. a market) it sets out to describe, or the nature of "the sorts of things that populate the (…) system" (2009, p. 40), and so does not even represent a *possible* version of such a system (Mäki 2001, p. 383; 385).[10]

5.2 Physical Conceivability

Michaela Massimi (2019a) argues that the epistemic import of certain exploratory modelling practices – "targetless" and "hypothetical" modelling – is that they deliver knowledge of what is possible. They do so by involving what Massimi calls "physical conceivability", which is a form of imagining. The key idea is that if a scientist can physically conceive of *p*, she is justified in believing that *p* is possible, and certain forms of exploratory modelling involve this particular form of imagining.

Again, imagination is assigned a central role, which brings up the issue of how imagination is to be properly constrained. But in this case, the answer appears to be built into Massimi's preliminary definition of physical conceivability:

> *p* is physically conceivable for an epistemic subject S (or an epistemic community C) if S's (or C's) imagining that *p* not only complies with the state of knowledge and conceptual resources of S (or C) but it *is also consistent with the laws of nature known by S (or C)* (Massimi 2019, 872, our emphasis).

In other words, in attempting to conceive of *p* in the relevant sense, the subject needs to hold fixed her knowledge of the laws of nature. Massimi illustrates this with the hypothetical modelling of SUSY (super symmetrical) particles in physics. Particle physics has theorized a hypothetical entity known as the SUSY particle, in order to account for certain gaps in the Standard Model. However, scientists have not been able to confirm whether there actually are any SUSYs in nature, but presumably that is ultimately what they want to do. As an important step towards this, they investigate the different ways in which it is physically possible that a SUSY particle exists, so that they can then go on by trying to rule out some of these scenarios as non-actual by using experimental evidence coming from proton-proton collisions. To this end, scientists have developed a modelling technique – the pMSSM-19 – which produces different "model points", (roughly: fictional model systems) that each portray SUSY particles as having mutually inconsistent properties and value assignments (e.g. a given mass value, a given decay mode, etc.) for 19 parameters, *and are consistent with certain nomological constraints* (e.g. R parity conservation, and consistent electroweak symmetry breaking). Specifically, what the pMSSM-19 is expected to do is to "trim down the bewildering space of 500 million conceivable model points to a more manageable size" (Massimi 2019a, p. 876).[11] For a given model point produced by pMSSM-19, it is concluded that a hypothetical target, corresponding to the model point, is objectively (physically) possible. That is: it is possible that a particle with such-and-such properties exists. This appears to hold some promise as a method for justifying claims of physical possibility – provided that the epistemic subject *does* possess the relevant knowledge of laws (and that this is successfully implemented in the pMSSM-19)

Massimi's physical conceivability account raises some interesting issues, in particular concerning the respective roles of imagination and background knowledge that we briefly touched upon in the previous section. This comes out especially in her attempt to distinguish between *law-bounded* (LB) and *law-driven* (LD) physical conceivability. The SUSY modelling just described is an

---

[10] See also Gelfert's comment that minimal models "shine a spotlight on the essential character of a phenomenon" (2019, p. 10-11),

[11] Presumably "conceivable" here refers to something more colloquial than the technical notion of physical conceivability.

instance of LB-conceivability, and is comparatively straightforward. As in the credibility account, imagination appears to play a – perhaps crucial – generating role. But – again, as on the credibility account – the justificatory power of the modelling exercise vis-à-vis the relevant modal claims, depends on the imaginary activity's being successfully constrained (or assessed, in the credibility account) by certain background knowledge. This follows what appears to be the standard in much contemporary modal epistemology: imagination is allowed to play the role of an important cognitive tool for exploration and creativity, but the justification is all down to the background knowledge which presumably already has modal content (although not necessarily overtly so), and must have been acquired in some other way. It is an interesting issue for future research – both in modal epistemology more generally, and for philosophers of science, given the current surge of interest in fiction and imagination as tools for explanation and justification, rather than "just" discovery – to what extent this generalizes.

Interestingly, Massimi takes the respective roles of knowledge of laws, and the imagination, to be more intertwined in LD-conceivability. Here, knowledge of the laws of nature allegedly *drives* analogical reasoning with concrete models from other fields, guiding the construction of a model indicating what is causally possible for the target system(s) of interest. Massimi's example here is Maxwell's construction of the molecular vortex model to derive the equations describing how electric and magnetic fields are generated by charges, currents, and changes of the fields. The molecular vortex was imagined in analogy with better-understood systems in other fields, specifically hydrodynamics. Maxwell drew on Faraday's law of electromagnetic induction, but also on Helmholtz' equations for fluid dynamics in imagining the system from which he could infer what possibly caused electromagnetic induction. Thus, knowledge of laws "drive" the imagining (rather than merely constrain it). It is unclear to us, however, how justification comes into the picture here – whether it is a result of background knowledge, the imaginary exercise, or a mix of the two, and in either case exactly how the justificatory route goes.

5.3. Universality

As noted above, Robert Batterman and Collin Rice have argued that minimal models enable scientists to draw true counterfactual conclusions about targets of interest. But they also attempt to address the epistemic question that arises for this modal modelling claim. As Rice writes,

> "[I]t seems somewhat mysterious how holistically distorted models can provide *true* counterfactual information about their target systems (…) I will try to offer one possible solution to this problem by appealing to *universality*." (2018, p. 2812, italics in original)

In a nutshell, the idea is this: suppose one knew that a given model system, despite being highly dissimilar to the target system of interest, were disposed to behave in largely the same way as the target system. In that case, they argue, one could justifiably use the model in order to learn about what would happen to the target system under such-and-such circumstances. Batterman and Rice's key claim is, in essence, that we *can* have this knowledge of sameness of behavior between idealized models and target systems. In their terminology, some idealized models and many target systems that interest scientists, are *in the same universality class*, and scientists can systematically discover that they are. They give several examples of strategies for such discovery, ranging from the Lattice Gas Automaton model vs. real fluid flow, Fisher's linear substitution cost model vs. sex ratios in various animal populations (both in Batterman and Rice 2014), to an optimal foraging model in an infinite population vs. Eider duck foraging behavior (Rice 2018).

Universality, Rice writes, is a "convenient feature of our universe", which in its most general form is just "the fact that (perhaps extremely) different physical systems will display similar macrobehaviors that are largely independent of the details of their physical components" (2018, p. 2812). Whether some systems are in the same universality class, is an empirical question –

something to be *discovered* (Rice 2018, p. 2813; 2019, p. 200). This discovery is what justifies scientists in using a certain model system to elicit modal information relevant to explaining the target phenomenon.

Interestingly, the universality account has important affinities with similarity-based modal epistemology. In Roca-Royes' similarity-based modal epistemology, what underwrites reasoning from the fact that *a* is F to the conclusion that *b* could possibility be F is the fact that *a* and *b* are similar in the sense that they share some relevant feature(s). Batterman and Rice denies that model and target need to be similar in this sense – systems in the same universality class need not share any features. Indeed, that's part of what motivates their account. However, what underwrites reasoning from model to modal conclusion about target is *similarity of token behavior* – to be in the same universality class just is to exhibit the same macrobehavior.

A looming question for similarity-based modal epistemology, as for standard accounts of scientific modelling that appeal to similarity between model and target features, is that of just what is a *relevant* similarity. As the conceptual problems of a binary similarity relation are well known, *what* the similarity comparison should include is typically made dependent on the purpose of the modelling exercise (Giere 1988). Proponents of the universality account seem to agree that the same holds for similarity of behavior. Rice for example notes that "the universality class required to justify a particular instance of idealized modeling will depend on the details of the modeling context; e.g. the target explanandum" (2018, p. 2816). That is, *what* behavioral similarity is needed depends on the counterfactual information one is after. But this admission casts into bold relief the fact that the universality account, just like similarity-based modal epistemology, faces the non-trivial question of how we come by the knowledge, or ability to reliably judge, what similarities (or dissimilarities) are relevant to a prospective possibility. It seems intuitively obvious that not any accidentally identified behavioral similarity between model and target would license modal inference. The question also seems somewhat more complicated for the universality account. On Roca-Royes' similarity-based modal epistemology, the assumption is that individuals with the same properties have the same causal profile, and so just coming to know that things are similar justifies concluding that they are disposed to behave similarly (although the question of *relevance* remains). But on the universality account, one needs to establish similarity of behavior *independent of* the instantiation of certain properties.

Batterman and Rice use several different cases to illustrate how it can be established that a model system and a target are in the same universality class. They have little to say about what these have in common, but as far as we can tell, at the most general level it is a form of robustness analysis that back up such claims which, in turn, justify relying on the model to draw counterfactual conclusions about the target. On closer inspection however, this seems more plausible in some cases than in others.

Their case of the Lattice Gas Automaton (LGA) model (Batterman and Rice 2014) stands out because of the many constraining assumptions it is based on. It relies on the renormalization group strategy, illustrated at the hand of the Kadanoff block spin transformation. The purpose of such investigations is to identify the physical systems, consisting of many interacting entities, that share the same scale-invariant macrobehavior, e.g. transitioning to an orderly state below a certain transition temperature. The basic idea of the Kadanoff transformation is to start with a space of possible systems, in this case constrained by assumptions about its entities and interactions. Some members of this possibility space might be real fluids, others include the LGA of a certain dimensionality and scale. Each of them exhibits some macrobehavior, and some of them might exhibit the same. A rescaling procedure applied to each possible system aggregates entities into a group and determines the macrobehavior of this rescaled system. Systems that continue to exhibit the same macrobehavior under such transformations are identified to belong to the same universality group, thus justifying the use of LGA to explain the behavior of fluids belonging to the same universality class.

Here, the procedure plausibly identifies a possibility space, determined by assumptions of what entities systems consist of and how they interact. It also offers a plausible rescaling procedure, and an easy way to infer the macrobehavior of these scaled-up entities. Arguably, all these constraints are supported by some kind of background knowledge, which might well contain implicit modal information.

But in other cases discussed by Batterman and Rice, the robustness analyses supporting the claims that models and targets are in the universality class are nowhere near as constrained or thorough. For instance, in the Eider foraging case, no attempt was made to systematically describe the possibility space – the authors only compared three alternative models that incorporated different constraints on optimization. Nor was there any attempt to re-scale the models and compare their predictions – the authors only compare the three model predictions with actual Eider behavior, concluding that one fits the data better than the other two. It is far from clear that mere comparative similarity of model prediction and empirical data licenses counterfactual and modal conclusions.

In sum, while the epistemic question for modal modelling may well be answered (for some cases) in terms of universality classes, it is a highly non-trivial matter to establish that there *is* relevant similarity of behavior, given the counterfactual conditionals one is interested in with respect to a target. First, establishing similarity at all, in the relevant sense, is less straight-forward than on e.g. Roca-Royes' similarity-based modal epistemology (where it is a matter of property sharing). Second, the universality account inherits the difficult question of relevance, in particular regarding which behavioral similarities are relevant to justify the counterfactual conclusions of interest.

6. Conclusions

Philosophers of science have documented scientists explicitly engaging in modal modelling practices. They also proposed to re-interpret scientific modelling practices as modal modelling, where these practices do not meet the standard criteria for epistemically good representation of actual phenomena. In both cases, the claim that scientists engage in modal modelling raises *the epistemic question of modal modelling*: in virtue of what do models indicate modal truths, and provide reasons for modelers believing them?

In this paper, we reviewed some attempts by philosophers of science to answer this question. We scrutinized these proposals through the interpretative lens of themes from general modal epistemology that do not specifically address the role of modelling. We found that these themes – relying on either imagination, background knowledge or similarity to the actual world, or some combination thereof – are very much present also in philosophers of science's attempts to answer the epistemological question for modal modelling.

We also found that the modal modelling accounts, in virtue of relying the same themes, face many of the same challenges and unresolved questions that arise in modal epistemology more generally. For one, while appeal to both imagination and similarity seem to implicitly rely on certain background knowledge (that might well contain modal information), it remains unclear what kind of knowledge is actually required for justifying modal claims. Before that is made clear, it is hard to assess the justificatory status of the relevant modal modelling practices, since it depends partly on whether scientists plausibly possess that knowledge. For another, the role of imagination – which according to many philosophers of science is very centrally involved in scientific modelling – is also somewhat unclear. Is it, after all, a "mere" exploratory tool, or does it somehow contribute to the justificatory power of models vis-à-vis possibility claims? Both of these are interesting avenues for future research on modal modelling, and we suspect that seeking further exchanges between work on modal modelling and on modal epistemology more generally will prove fruitful on both these and other issues.

References

Akerlof, G. A. (1970). The Market for "Lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, 84(3), 488-500.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, 22, 265–290.

Batterman, R. W., & Rice, C. C. (2014). Minimal Model Explanations. *Philosophy of Science, 81*(3), 349-376.

Bealer, G. (2002). Modal Epistemology and the Rationalist Renaissance. In T. S. Gendler, & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 71-125). Oxford: Clarendon Press

Bird, A. (2007). *Nature's Metaphysics*. Oxford: Oxford University Press

Cartwright, N. (2009). "If No Capacities Then No Credible Worlds: But Can Models Reveal Capacities?" *Erkenntnis 70* (1), 45-58

Chalmers, D. (2002). Does conceivability entail possibility? In T. S. Gendler, & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145-200). Oxford: Clarendon Press.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science, 74*, 48-68.

Ellis, B. (2001). *Scientific Essentialism*. Cambridge: Cambridge University Press

Fischer, B. (2017). *Modal Justification via Theories*. Cham: Springer International.

Frigg, R., & Nguyen, J. (2016a). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2020 edition)*.

Frigg, R., & Nguyen, J. (2016b). The fiction view of models reloaded. *The Monist*, *99*(3), 225-242.

Fumagalli, R. (2016). Why We Cannot Learn from Minimal Models. *Erkenntnis, 81*(3), 433-455.

Gelfert, A. (2018). Models in Search of Targets: Exploratory Modelling and the Case of Turing Patterns. In A. Christian, D. Hommen, N. Retzlaff, & G. Schurz (Eds.), *Philosophy of Science. European Studies in Philosophy of Science, vol 9.* (pp. 245-269): Springer International Publishing.

Gelfert, A. (2019). Probing Possibilities: Toy Models, Minimal Models, and Exploratory Models. In M. Fontaine, C. Barés-Gómez, F. Salguero-Lamillar, L. Magnani, & Á. Nepomuceno-Fernández (Eds.), *Model-Based Reasoning in Science and Technology*. Cham: Springer International.

Giere, R. N. (1988) *Explaining Science: A cognitive approach* Chicago: University of Chicago Press

Grüne-Yanoff, T. (2009). Learning from Minimal Economic Models. *Erkenntnis, 70*(1), 81-99.

Grüne-Yanoff, T. (2011). Isolation is not characteristic of models. *International Studies in the Philosophy of Science*, *25*(2), 119-137.

Grüne-Yanoff, T. (2013). Appraising Models Nonrepresentationally. *Philosophy of Science, 80*(5), 850-861.

Jago, M. (2018). Knowing how things might have been. *Synthese,* doi:10.1007/s11229-018-1869-6.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*(4), 507-531.

Kment, B. (2014) *Modality and Explanatory Reasoning* Oxford: Oxford University Press

Kung, P. (2010). Imagining as a Guide to Possibility. *Philosophy and Phenomenological Research, 81*(3), 620-663.

Kung, P. (2016). You Really Do Imagine It: Against Error Theories of Imagination. *Noûs, 50*(1), 90-120.

Lowe, E. J. (2012) What is the source of our knowledge of modal truths? *Mind*, *121*, 915-950.

Mallozzi, A. (2018). Putting modal metaphysics first. *Synthese,* doi:10.1007/s11229-018-1828-2.

Massimi, M. (2019). Two kinds of exploratory models. *Philosophy of Science, 86*(5), 869-881.

Mäki, U. (2009). MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis, 70*(1), 29-43.

Nguyen, J. (forthcoming). It's Not a Game: Accurate Representation with Toy Models. *British Journal for the Philosophy of Science*, doi:10.1093/bjps/axz010.

Reutlinger, A., Hangleiter, D., & Hartmann, S. (2018). Understanding (with) Toy Models. *British Journal for the Philosophy of Science, 69*(4), 1069-1099.

Rice, C. (2018). Idealized models, holistic distortions, and universality. *Synthese, 195*(6), 2795-2819.
Rice, C. (2019). Models Don't Decompose That Way: A Holistic View of Idealized Models. *British Journal for the Philosophy of Science, 70*(1), 179-208.
Roca-Royes, S. (2017). Similarity and Possibility: An epistemology of de re possibility for concrete entities. In B. Fischer, & F. Leon (Eds.), *Modal Epistemology After Rationalism* (pp. 221-245). Cham: Springer International.
Rohwer, Y., & Rice, C. (2013). Hypothetical Pattern Idealization and Explanatory Models. *Philosophy of Science, 80*(3), 334-355.
Schaffer, J. (2017). Laws for metaphysical explanation. *Philosophical Issues, 27* (1), 302-321
Shelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(2), 143-186.
Sjölin Wirling, Y. (ms) Is credibility a guide to possibility? A challenge for toy models in science.
Sjölin Wirling, Y. & Grüne-Yanoff, T. (ms). Epistemic and Objective Possibility in Science.
Strevens, M. (2008). *Depth*. Cambridge, Mass.: Harvard University Press.
Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology, 7*(1), 1-31.
Verreault-Julien, P. (2017). Non-causal understanding with economic models: the case of general equilibrium. *Journal of Economic Methodology, 24*(3), 297-317.
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.
Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.
Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research, 53*(1), 1-42.
Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology, 21*(1), 19-36.