Synthese (2022) 200:448
https://doi.org/10.1007/s11229-022-03928-z

ORIGINAL RESEARCH

Two epistemological challenges regarding hypothetical modeling

Peter Tan[1]

Received: 30 December 2021 / Accepted: 10 October 2022

Abstract
Sometimes, scientific models are either intended to or plausibly interpreted as representing nonactual but possible targets. Call this "hypothetical modeling". This paper raises two epistemological challenges concerning hypothetical modeling. To begin with, I observe that given common philosophical assumptions about the scope of objective possibility, hypothetical models are fallible with respect to what is objectively possible. There is thus a need to distinguish between accurate and inaccurate hypothetical modeling. The first epistemological challenge is that no account for the epistemology of hypothetical models seems to cohere with the most characteristic function of scientific modeling in general, i.e., surrogative representation. The second epistemological challenge is a version of "reliability challenges" familiar from other areas. There is a challenge to explain how hypothetical models could be a reliable guide to what is possible, given that they are not and cannot be compared against their nonactual targets and updated accordingly. I close with some brief remarks on possible solutions to these challenges.

Keywords Hypothetical modeling · Possibility · Modal epistemology · Reliability challenge

1 Introduction

Sometimes, scientific models are either intended to or plausibly interpreted as representing nonactual but possible targets. For instance, Thomas Schelling writes regarding his widely discussed checkerboard model of segregation that.

✉ Peter Tan
ptan8@fordham.edu

[1] Philosophy Department, Fordham University, 113 W 60th St, Lowenstein Building, Rm 916, New York, NY 10023, USA

Published online: 29 October 2022

> "Sometime in the 1960s, I wanted to teach my classes how people's interactions *could lead to results* that were neither intended nor expected" (2006, p249, emphasis added)

At least judging from this passage, the intended representational target of Schelling's segregation model is something modal in nature: the mere possibility of purely agent-based segregation. And some philosophers of science have noted that "[Schelling's] model system that produces the segregation pattern is not established as an adequate representation of any real-world system" (Grüne-Yanoff 2013, p. 6). Hence, insofar as Schelling's model tells us about the possibility of certain mechanisms that lead to segregation, it appears to tell us about a *nonactual* possibility.

Or consider Maynard Smith and Price's hawk-dove game, a simple game-theoretic model used in evolutionary biology. The game is a toy model of competition over resources; rather than engaging in high-risk, aggressive behavior, players often display restraint in conflict resolution. By the practitioners' own lights, the hawk-dove model is intended to represent something possible: how selection at the individual level alone could give rise to certain behaviors (Maynard Smith & Price, 1973). Philosophers of science tend to agree.[1] Collin Rice and Yasha Rohwer write that "although the model fails to accurately represent the selection dynamics of any real-world population, it does tell us something about how individual selection could lead to the trait in a wide range of possible systems, by investigating a hypothetical scenario" (2013, p341). Just as with Schelling's checkerboard, the hawk-dove model's target appears to be a nonactual possibility.

Or consider the uses of optimality modeling in biology. Optimality models typically attempt to provide an explanation for some phenotypic trait by showing that it is at the optimal trade-off point between some other factors, for instance, the copulation period for dung flies (Parker, 1978; see Rice, 2012, 2015 for commentary) or the number of sexes in a population (Fisher 1930; see also Weisberg, 2013; Rice, 2015). Many optimality models appear to target nonactual possibilities, too. They plausibly fail to represent actual populations; Rice writes that, "the explanations of several optimality models utilize various idealizations (e.g. idealized mathematical curves or assuming infinite population size), which ensure that the models themselves do not accurately represent the causal processes of any real-world biological population," (2012, p700). And regarding the modal dimension, Angela Potochnik comments, "the aim of optimality modeling is merely to represent *possible* selection dynamics… This may be what Seger and Stubblefield have in mind when they say that optimality models are used to evaluate whether specific assumptions about selection and constraints, etc., "could account at least in principle" for the target evolutionary outcome" (2009, p. 189).

Finally, consider the Lotka-Volterra model. The model is quite minimal and idealized: it only considers a predator and prey species' populations as they relate to one another, abstracting entirely from any facts about the ecosystem's carrying capacity, resources, even the predator species' rate of reproduction or appetite, and so on. The Lotka-Volterra model, as a highly minimal so-called "toy" model, is also sometimes regarded as only representing hypothetically or "how-possibly" (Reutlinger et al.,

[1] For a dissenting view, see Verreault-Julien (forthcoming).

2018). It shows that under certain nonactual idealizing conditions, predator–prey populations would display a certain kind of oscillation when paired together. Again, just like the previous examples, there is reason to think that the Lotka-Volterra model represents a nonactual possibility.

There are still other examples of models that follow the general pattern just described: Akerlof's market of lemons (Reutlinger et al., 2018), the Arrow-Debreu theorem of general equilibrium in economics (Grüne-Yanoff & Verreault-Julien, 2021; Hands, 2016; Verreault-Julien, 2017), a variety of models in synthetic biology (Rami, 2017; Knuuttila & Loettgers, 2022), and many others. These models appear to be aimed at, or are otherwise plausibly interpreted as, representing nonactual but possible targets. Following an existing precedent, let's call this set of practices 'hypothetical modeling.'[2] My aim in this paper is to raise two epistemological challenges regarding issues that commonly arise in philosophical discussions of hypothetical modeling. I do not intend to offer a set of "knock-down" challenges; rather, my aim is to flag a pair of issues that have not been adequately flagged and which philosophers of science would do well to remedy.

I first introduce a baseline problem. Discussions of hypothetical modeling frequently allege that hypothetical models provide information about objective possibilities. But I argue that upon inspection of the kinds of idealizations involved in these models or sometimes the content of the model itself, there are at least some cases of hypothetical modeling where what is alleged to be possible isn't objectively possible. In some cases, the targets of the hypothetical models are nomologically impossible, and in some other cases, perhaps even "flat-out" impossible. The general point is that hypothetical modeling seems to be fallible with respect to whether something alleged to be possible *objectively is* possible.

Once it becomes clear that the allegedly possible targets of hypothetical models are sometimes not possible at all, we face an epistemological challenge. Under what conditions can we have evidence that a hypothetical model is *successful* in its modal ambitions, i.e., under what conditions can we have evidence that a hypothetical model really does represent a nonactual possibility? Though there have been some attempts at addressing this issue, notably in terms of the "credibility" of the models in question (Grüne-Yanoff, 2009; Sugden, 2000, 2009), I argue that there is a deeper issue left unresolved. I shall show that evaluating the characteristic *surrogate* function of scientific models is impossible if their targets are nonactual possibilities, and that this leaves the epistemology of modeling objectionably disunified.

Next, I shall show that there is a reliability challenge for hypothetical modeling of the familiar sort that arises in metaethics and in the philosophy of mathematics. Even supposing hypothetical modeling is reliable with respect to what's possible, there

[2] Sometimes, discussions of hypothetical modeling include exploratory models, which I intend to ignore here – my focus is only on the uses of hypothetical models which appear to be focused on *objective* possibilities rather than epistemic possibilities. (See Grüne-Yanoff and Sjölin Wirling forthcoming for some excellent discussion.) Likewise, philosophers sometimes refer to hypothetical models as providing "how-possibly" explanations. But these discussions often assume or contest certain background assumptions regarding the nature of scientific explanation. Again, I wish to sidestep these issues and focus only on the kinds of modeling practices that are either aimed at or plausibly interpreted as representing objective possibilities.

appears to be an in-principle difficulty in explaining *how* these modeling methods could be reliable, given the lack of ability to evaluate their surrogacy.

Finally, in the concluding section of the paper, I offer some preliminary remarks about possible solutions to these challenges. One of these solutions, an interpretation of hypothetical models as representing *actual* targets (Nguyen, 2020, Verreault-Julien forthcoming) is, I think, especially promising.

2 The baseline problem: the fallibility of hypothetical modeling

Let me now turn to the first and, for the purposes of this paper, most fundamental issue regarding hypothetical modeling. The core idea is that in at least some cases, hypothetical models represent as possible some states of affairs which turn out not to be possible at all. To recall, prevailing interpretations hold that many hypothetical models do not represent any actual systems – e.g., general equilibrium models do not represent actual markets; optimality models don't represent actual selection processes – but the models instead tell us about *possible* markets, possible ecosystems, and their behaviors. And the kind of possibilities alleged to be represented are meant to be *objective* or worldly possibilities. The models are not typically of interest because they illuminate *epistemic* possibilities, i.e., propositions that "for all we know" might turn out to be the case. In some cases, such as the hawk-dove game, it is antecedently agreed upon that "the model fails to represent…any real-world population," (Rohwer & Rice, 2013 p. 341). In other cases, even if we assume for the sake of argument that some model fails to stand in "any similarity, isomorphism or resemblance relation to the world" (Grüne-Yanoff, 2009, p. 83), and hence that it does not represent some *epistemic* possibility, the model can nevertheless still be epistemically valuable. Whatever the epistemic value of these kinds of models, in these cases it appears to be in representing nonactual *objective* possibilities: an objective possibility regarding predator–prey behavior, or segregation, or the selection of restraint in aggression.

But what would it mean to say that (e.g.) Schelling's checkerboard represents a possible city, or the Lotka-Volterra model represents an objectively possible predator–prey ecosystem, or that the hawk-dove game represents a possible interaction between agents? If by "objective" possibility, we mean *worldly* possibility, i.e., a way that the world really could have been, there are several limitations on what the scope of possibility really is. States of affairs that violate known laws of nature fail to be objectively possible in at least one sense, namely, nomological possibility. So, are the allegedly possible targets of these hypothetical models even possible to begin with?

Because models like these are highly idealized, at minimum it seems like some hypothetical models represent targets which aren't *nomologically* possible. This point is hardly new; there is widespread agreement that many idealizations are nomologically impossible. Point masses, frictionless planes, and their ilk are all beyond the pale of the actual laws of nature.[3] The same applies to the kinds of idealizations involved in at

[3] Of course, some of these idealizations might "seem" less impossible than others. For instance, the behavior of frictionless planes, while nonactual, is quite easily epistemically accessible. We can simply extend the curve of an observed surface as its friction decreases to the hypothetical limit where there is no friction. However, the fact that these idealizations are epistemically accessible should not make us think that they are

least some of the hypothetical models of interest. To take the Lotka-Volterra model, the model comprises a pair of coupled differential equations. Interpreted literally, if it were objectively possible for there to be a Lotka-Volterra ecosystem, there would be possible ecosystems that contained infinitesimal fractions of organisms (cf. McLoone, 2020). But this is definitely nomologically impossible. Concerning optimality models, Rice writes that "the explanations of several optimality models utilize various idealizations (e.g. idealized mathematical curves or assuming infinite population size), which *ensure that the models themselves do not accurately represent the causal processes of any real-world biological population*" (2012, p. 700, emphasis added). These hypothetical models, too, appear to represent things that are not nomologically possible.

Likewise, it is not nomologically possible for there to be the kinds of markets described in, e.g., Akerlof's market of lemons. For if we interpret literally economic models like Akerlof's where there is an idealized state of perfect information (e.g., sellers but not buyers have perfect information), there would be instantaneous information transfer (including, e.g., instantaneous order completion). This is an idealization whose truth would, if interpreted literally, violate special relativity. Regarding other frequently-mentioned models, Ylwa Sjölin Wirling comments that "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" (2021, p. 476). So in at least some cases, what is alleged to be an objective possibility turns out not to be *nomologically* possible. Their possibility would require different laws of nature, different fundamental constants, and so on. And most importantly, if being nomologically impossible means that something just is impossible in a worldly sense, then that means at least some of these models are incorrect with respect to what is objectively possible.

To be sure, I am happy to agree that plenty of hypothetical models successfully represent worldly possibilities. For instance, other cases of optimality modeling regarding, e.g., the optimal copulation period for certain species of dung flies (see Rice, 2012), don't seem to involve the kinds of impossible idealizations just mentioned. And it may be that other kinds of hypothetical modeling in the social sciences, such as general equilibrium models in economics, also successfully represent objective possibilities that "do not violate any laws of logic or nature" (Hands, 2016, p. 38).

Still, regarding the apparently impossible cases, perhaps there is a different sense in which the content of these problematic hypothetical models are objectively possible. Perhaps they are "metaphysically" possible, where metaphysical possibility is frequently alleged to be the broadest, most "flat-out" kind of possibility (cf. Williamson, 2016, pp. 459–460; Sider, 2003; Szabo Gendler & Hawthorne, 2002; Fine, 2002; among many others). So, are optimality-model-populations, Lotka-Volterra ecosystems, Arrow-Debreu markets, and the like *metaphysically* possible?

Though is notoriously difficult to understand exactly what is and is not metaphysically possible, at least some of the systems alleged to be objectively possible still look impossible. Again, begin by considering the case of the Lotka-Volterra model.

Footnote 3 continued
any less impossible: friction results from intermolecular electromagnetism, so, for there to be no friction, there would need to be different fundamental forces than there actually are – hence the idealization is still nomologically impossible. Thanks to a referee for pressing me on this point.

If it were objectively possible for there to be a Lotka-Volterra ecosystem, there would be possible ecosystems that contained infinitesimal fractions of organisms. But this doesn't seem only *nomologically* impossible; this seems "flat-out" impossible. The sorts of things that organisms *are* could not be the sorts of things that there could possibly be infinitesimal fractions of, following McLoone (2020). And as I have argued in previous work (Tan, 2019), whether one considers these kinds of possibility-claims *de re* or *de dicto*, they still look to be flat-out impossible, i.e., metaphysically impossible. (*De re*: would it be possible for the very sorts of things that actually are organisms to be infinitesimal fractions?—definitely not. *De dicto*: would it be possible for an infinitesimal fraction of an ideal organism to fill the causal roles that predator and prey organisms actually fill?—again, quite plausibly not.) Even if the realm of the objectively possible is expanded to include things that are nomologically impossible but in some sense metaphysically or "flat-out" possible, some hypothetical models still appear to be mistaken about what is objectively possible.

This general point also extends to cases of optimality modeling, although perhaps not with as much force. In cases of optimality modeling, common idealizations include infinite population size and no genetic drift. And as we have seen, a common interpretation of optimality modeling is that optimality models identify *merely possible* causal explanations by which certain traits could have been selected for. However, it's admittedly a little difficult to imagine any sense in which infinite population sizes would be possible: are we considering a possibility in which there is infinite space, as well? It's hard to consider an infinite space of the specific sort where organisms could reside as an objective possibility; likewise, if there isn't infinite space, then it seems hard to imagine how there could be the possibility of an infinitely large population.[4] On either way of filling in the further details about these idealizations in optimality modeling, it seems difficult to agree that we are considering an objective possibility when asked to imagine an infinite population.

Now, some may worry here that I have glossed over mathematical or logical possibility. However, in response, many philosophers agree that the objective, worldly possibilities are not coextensive with mathematical or logical possibility. For instance, plenty of mathematical or logical possibilities are inconsistent with the laws of nature; and again, if what is nomologically impossible *just is* impossible, then so too are some apparent possibilities of math and logic. Additionally, many philosophers have argued that some things which are noncontradictory still fail to be "flat-out" objectively possible. An easy example comes from Plantinga (1974). If S5 is the correct modal logic, then for plenty of empirically observable and noncontradictory propositions P, $\Diamond\Box$P is *also* noncontradictory, but known to be false. (For example: let P be the proposition *that there is an elephant in a puddle in the Louvre*.) In this sort of case, $\Diamond\Box$P appears to be logically possible (not contradictory) but nevertheless turns out to be impossible in the worldly sense (since '$\Diamond\Box$P' is false). So, we should not be so quick to infer that because something is mathematically or logically possible, that it is also possible in a worldly sense.

[4] For those worried about my use of "imagination" as a rough-and-ready source of evidence regarding modality here, I am happy to concede that imagination is an unreliable source of evidence about what is possible: in fact, this only strengthens my overall concerns. See Sect. 4 below.

Back to the main issue. Again, I'm happy to agree that there are examples of hypothetical modeling that do represent objective possibilities. But the overall upshot here is that if we are careful to inspect particular cases, sometimes a model is aimed at representing a nonactual possibility but the target turns out not to be possible at all. In other words, given common philosophical assumptions about the scope of objective possibility (whether nomological or "metaphysical"), there is no guarantee that a hypothetical model really does manage to represent what is objectively possible. This observation is different from what some others in this literature have observed, e.g., that the *kind* of possibility involved in hypothetical modeling is often left underspecified (Sjölin Wirling, 2021).[5] My observation is that even when we do attempt to consider what the scope of possibility involved might be, some hypothetical models turn out to be incorrect with respect to what is possible.

To wrap up the first, baseline challenge, put all this another way: there should be room for *fallibilism* about hypothetical modeling.[6] Just like models targeted at actual target systems can be incorrect or inaccurate with respect to what those systems are like, hypothetical models can also be incorrect or inaccurate with respect to what the space of possibility is like.

3 An epistemological gap: surrogative function

This baseline issue helps introduce the first epistemological challenge. If we begin to appreciate the observation that hypothetical models can be wrong regarding what is possible even when they *appear* to represent things that are possible, there arises an epistemic question. Under what conditions would there be evidence that a particular model really does manage to represent something that is objectively possible?

Of course, I am not the first to address this issue. Till Grüne-Yanoff and Ylwa Sjölin Wirling (2021) identify this question as one of two pressing epistemic questions concerning modally-oriented scientific models. (The other pressing question they identify concerns why it would be interesting or useful to learn about nonactual possibilities.) And among the existing attempts at addressing the issue, there is a well-known strand of thought according to which a model's *credibility* is the primary means to assessing whether it successfully represents a possibility. The next section of the paper will discuss this approach. But first, in this section, I wish to press the issue in a slightly different way than what has been suggested thus far in the literature. Rather than asking what might serve as evidence that *hypothetical* models manage to successfully represent possibilities, the issue can be subsumed under a more general one. How would we evaluate the success or accuracy of a scientific representation in general? The question concerning the success of a hypothetical model is simply a particular instance of a question concerning when, in general, we would have evidence that a model successfully represents its target. My arguments in this section will show that our epistemology of successful hypothetical modeling ends up being disconnected from models' characteristic *surrogate* function.

[5] It is worth noting that van Riel (2015) is overall quite sympathetic to these worries.

[6] Many thanks to a referee for this framing.

To see why I say this, consider the more general issue regarding how we would begin to evaluate or gain evidence that a model represents its target successfully. It is important to note here that I am not asking for some standard or criterion for assessing the accuracy of a representation. As has been noted many times, scientific representation is ultimately a pragmatic enterprise guided by the epistemic values and goals of agents (Giere, 2004, 2010; Suàrez, 2015; Frisch, 2015). So, the standards for assessing a model's accuracy or success are of course going to be highly context- and interest-dependent; no general standards are available (cf. Frigg & Nguyen, 2017, p. 54). There are also many *ways* in which a model might represent its target successfully. But again, because of variations in what 'success' means, there is no generalized *standard* for accurate or successful representation in a model. That's not what I'm asking about. Instead, I'm simply asking: how would we even *begin* to evaluate, in general, the extent to which a model represents its target successfully? The most general answer is clearly that we evaluate it with respect to its *surrogate* function.

Regardless of which particular conception of scientific representation one might endorse (e.g., similarity, isomorphism, inferentialism, etc.), or even whatever particular context of representation one examines, the ineliminable rock-bottom function of scientific models is surrogative (Hughes, 1997; Suárez, 2004). Models are used to stand in for their targets, i.e., to generate predictions about the targets, and to be manipulated, observed, and modified in place of their targets. Roman Frigg and James Nguyen note that "*every* acceptable theory of scientific representation has to account for how reasoning conducted on models can yield claims about their target systems" (2017, p52). The surrogative function of models is so fundamental, indeed, that some have argued (Levy, 2015) that any (purported) model that lacks a concrete target simply isn't a model at all. If the primary function of models is surrogative, then the epistemology of when an instance of modeling has been successful must be unified with evaluating it *as a surrogate*: we'd see if the model's predictions, etc., about the target end up matching up to it in whatever way is contextually and pragmatically relevant.

The primary epistemological challenge is that evaluating the surrogative function of hypothetical models is pretty much impossible. How are we to evaluate whether the predictions generated by any of these models obtain *in their purported target*? Non-actual possibilities certainly can't be observed. Manipulation is right out the window, too: we can't manipulate or intervene on merely-possible ecosystems, or markets, or selection processes. So, right off the bat, if the targets of hypothetical models really are nonactual possibilities, it looks like there is no epistemic access whatsoever to evaluating how well they perform in their function as surrogates for the targets.

One might note that part of the issue here is just the more general problem of the epistemology of modality: how can there be evidence or knowledge regarding nonactual states of affairs? And readers might well have their own antecedent views on how, e.g., conceivability, or whatever else, might feature (or not) into the epistemology of modality. But the issue here is that regardless of how one answers more general questions regarding the epistemology of modality, hypothetical models are supposed to be a kind of scientific model. And scientific models are objects that (following the many others quoted above) already have an epistemic function, i.e., representation through surrogacy. So, regardless of how one answers more general questions regarding the

epistemology of modality, the epistemology of hypothetical modeling comes with some antecedent constraints. The function of models is supposed to be surrogative representation, so an account of models' success-conditions is plausibly based in how well they perform that function; and if so, an epistemology of *when* models have met those success-conditions would be based in evaluating the same. My worry is thus that because it is impossible to evaluate the quality of surrogative function in cases of hypothetical modeling, there would first be a disunity in their epistemology from the epistemology of modeling more generally, but more importantly, a disunity from models' basic epistemic function.

There is some need to tread a little carefully here. You might well observe that there is some sense in which any model, including those of actual target systems, represents its target in modal ways. For instance, according to a highly influential view regarding scientific explanation (Woodward, 2003; Woodward & Hitchcock, 2003), the primary content of any representational device that provides explanatory information—be it a model, an argument, or whatever else—is *counterfactual* information about these actual targets. This account of explanation is readily applied to understand the content of many kinds of scientific models (Woodward, 2003, p. 224; Bokulich, 2011), including some hypothetical models (Reutlinger et al., 2018). On this sort of view, every model represents its targets modally, since all models will impute certain counterfactual structure to their targets: "if variable x in the target were to be different in such-and-such a way, then variable y would differ thus-and-so." With this in mind, one might worry that my complaints about hypothetical models overgeneralize. Isn't modal reasoning in some broad stripes the "aim" or target of *any* model, even models of actual systems?[7] Indeed, many real-world systems' counterfactual structure is beyond our epistemic access via intervention, e.g., financial markets, ecosystems, certain kinds of models involving human subjects, etc. It would be worrying if my arguments overgeneralized to say that our epistemology of these ordinary kinds of models is also problematically disunified with their surrogative function.

In response, it's important to clarify between the different ways in which models might have modal content. I agree that many if not all scientific models provide modal information in some respect. But there is an important difference in hypothetical models' representational *target*. Models of actual systems—say, the ideal gas law as a representation of actual gases—aren't directed at nonactual targets, even if they have counterfactual or modal content. Instead, these models represent actual targets—say, samples of krypton gas—as having certain modal or counterfactual properties: e.g., "if this actual gas sample's pressure were to be increased, then *ceteris paribus* its temperature would increase." As Hirvonen et al. (2021) have noted, this kind of modal reasoning about actual targets is thoroughly epistemologically unproblematic. This general observation about the modal content of models directed at actual systems applies even in those cases where intervention is impossible (cf. Woodward, 2003, §3.5). Newton's law of universal gravitation imputes counterfactual structure to the distance between the earth and moon, even though we cannot intervene on that distance. In these kinds of cases, our epistemology of the models' accuracy is still unified with the models' surrogative function, simply because the models' targets are actual, and

[7] Thanks to an anonymous referee for pressing me on this point.

empirical observation is possible. The quality of the model's surrogative function ends up being evaluated with observation (or perhaps simulation) only, instead of manipulation.[8]

By contrast, the kinds of models I have argued face an epistemological challenge are specifically hypothetical models which are not taken in the first place to represent actual targets. Rather than imputing modal structure to actual targets, with hypothetical models, the models themselves are directed at nonactual possibilities. It is precisely *because* the alleged targets of hypothetical models are, as it were, purely hypothetical, that there is a genuine challenge with respect to evaluating these models' surrogate function. Again, even in cases of modeling actual targets where our models might provide modal information, and where we cannot directly intervene on the target of the model, the models' surrogative function can still *in principle* be evaluated by observation (Hirvonen et al., 2021) or computer simulation (cf. Parker, 2017). My worry about surrogacy is that this is not possible with respect to hypothetical models even in principle.

4 Refining the surrogacy challenge: alternative sources of evidence?

Can other sources of evidence help in recovering an epistemology of hypothetical modeling that is consistent with models' general surrogative function? This section will consider a number of alternatives, including what has been the most prominent suggestion to evaluate whether a hypothetical model really does represent an objective possibility: the models' *credibility*.

First, however, before discussing credibility, let me discuss two other options which might come to mind as sources of evidence for evaluating whether a hypothetical model successfully represents an objective possibility. The first potential option is straightforward: given the mention of computer simulation methods in the previous paragraph, you might wonder whether simulation can serve as evidence that an alleged possibility really is possible. Now, in some kinds of cases, comparison against simulation plausibly can perform the required evidential work. If there is independent empirical or theoretical reason to think that a particular set of laws or a certain part of an accepted theory delineate what is objectively possible, then as long as the simulation both coheres with those laws, the fact that a hypothetical model is consistent with simulation can be evidence that that model's possibility claims are true. (Many cosmological models with modal content seem to work in this way, see e.g. Jacquart, 2020.)[9]

However, in other cases, recourse to computer simulation might not be possible in this confirmatory way. The reason is that many hypothetical models are, as Reutlinger et al. (2018) and Gelfert (2019) have observed, "autonomous" of preexisting theory. Their idealizations and characteristic dynamics are explicitly *not* drawn from any preexisting theory but are instead merely stipulated for the sake of building the model.

[8] See Hirvonen et al., (2021, p. 13831) for an observation-only example regarding Alexander Fleming's eventual development of penicillin.

[9] Thanks to Till Grüne-Yanoff for pushing me to clarify this point. See also Grüne-Yanoff 2013, p857 for an example of the confirmatory value of simulation from historical anthropology.

(Plausible examples include Schelling's checkerboard, the Lotka-Volterra equations, and the hawk-dove game.) In these kinds of cases, my worry is that the epistemic concern about surrogacy cannot be resolved by recourse to computer simulation. For any simulation built to validate an autonomous hypothetical model will presumably have exactly that model's characteristic equations as its dynamical laws. But if the model is already autonomous of theory, what grounds would we have for thinking that a *simulation* with more or less exactly the same assumptions, idealizations, and dynamical equations as the model successfully represents a possible system? That is exactly what is supposed to be at stake! In these kinds of cases where hypothetical models are autonomous of theory, relying on coherence with simulation as a source of evidence for whether a model successfully represents a possibility seems at best to only push the question back.

There is a second option worth considering here. Bob Fischer (2016, 2017) has developed an epistemology of modality that locates the evidence we have for modal claims in theories (usually scientific theories) that underlie them. Where p is any modal claim, Fischer says, you justifiably believe that p if both (a) you justifiably believe a theory that says that p and (b) you believe that p on the basis of that theory (Fischer, 2016, p. 11; 2017, p. 23). I find this suggestion regarding the epistemology of modality quite promising. Indeed, right off the bat there do seem to be examples of hypothetical modeling for which this epistemology seems to correctly locate the source of our evidence for taking the hypothetical model to represent an objective possibility: for instance, many optimality models require coherence with general principles of selection (see Potochnik, 2009) that are antecedently well-confirmed. However, there are a few clear cases of hypothetical modeling, too, where Fischer's theory-based epistemology of modality will *not* sufficiently explain why we are justified in taking the models to represent objective possibilities. Consider just two examples. First, sometimes we are justified in taking a hypothetical model to represent an objective possibility even when its possibility-claims *conflict* with established theory: discussions of general equilibrium modeling in economics (Grüne-Yanoff & Verreault-Julien, 2021; Hands, 2016) note that part of the epistemic value of the Arrow-Debreu theorem was to establish something to be at least mathematically possible when previous theory took it to be impossible. Prima facie, Fischer's epistemology of modality has difficulty accounting for our modal evidence in these kinds of cases because it locates that evidence in theories.[10] Second, as mentioned above, many plausible cases of hypothetical modeling in the sciences involve "autonomous" models whose idealizations and characteristic dynamics are explicitly *not* drawn from any preexisting theory but are instead merely stipulated for the sake of building the model. Again, since Fischer's theory-based epistemology of modality locates the source of our evidence for modal claims in theories,

[10] See also Sjölin Wirling (2021, p. 475, fn5), Sjölin Wirling and Grüne-Yanoff (2021, §6). Although, as Fischer (2017, §2.4) notes, depending on which sorts of things count as theories, there might be more of a way to integrate these kinds of cases with the theory-based epistemology of modality. Maybe the Arrow-Debreu model by itself constitutes a theory. However, it is worth noting that when it comes to scientific theories, Fischer endorses the semantic view of theories, according to which theories are collections of models. And under the semantic view, it still looks like at least at the initial time of publication, the Arrow-Debreu model conflicted with the established theory, and didn't itself constitute a theory (since presumably a single model does not constitute a theory). Still, there may be more to say here.

there is some prima facie difficulty in accounting for what the evidence for the modal success of *autonomous* hypothetical models could be.

One last comment about Fischer's theory-based epistemology of modality. To recall, my primary concern is that if the targets of hypothetical models really are interpreted as nonactual possibilities, then it will be impossible to evaluate their accuracy in terms of models' characteristic surrogative function. The issue is not simply how to locate any source of evidence at all that would help discriminate between successful and unsuccessful hypothetical modeling. With that in mind, note that even if we endorse Fischer's theory-based epistemology of modality and we can find a source of evidence that helps discriminate between hypothetical models that are correct about what is possible and those that are incorrect, it's not immediately obvious that this will address my concern. For even if we have evidence from some antecedent theory that some hypothetical model really does represent a objective possibility, this is clearly a different source of evidence from evaluating the model's accuracy against that merely-possible target *as a surrogate*.

Now let's move to consider a final alternative source of evidence regarding whether a hypothetical model successfully represents a possibility. Consider the idea that a model's *credibility* can be a source of evidence regarding its successfully representing a possibility. The idea originates with Robert Sugden, who writes (regarding Schelling's well-known checkerboard model of segregation): "…we 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, p. 25, emphasis in original). Others, including Uskali Mäki (2009) and Grüne-Yanoff (2009), have taken on board the idea that credibility can be a source of evidence regarding a model's having successfully represented an objective possibility. It is unclear what exactly credibility is; Mäki, for instance, chalks credibility up to a certain kind of conceivability (2009, p. 40). The most fleshed-out account of credibility currently on offer draws on an analogy to fiction and imagination. Sugden (2000, p. 25), writes that "credibility in models is…rather like credibility in 'realistic' novels," and Grüne-Yanoff (2009, p. 94), comments, "Imagination creates *what* could have been, and assessment of this imagination focuses on *whether* it could have been… Models share these two aspects of fiction."

There is at least one serious preliminary concern one might have about credibility as a source of evidence regarding possibility. Sjölin Wirling (2021) persuasively argues that since credibility ends up being a form of imagination, the thesis that credibility can provide evidence of possibility faces the same obstacles that any imagination-based epistemologies of modality face. Kung (2016; Vaidya & Wallner, 2021) all discuss this issue. Sjölin Wirling provides a compelling example, writing,

> If we take the analogy with fiction seriously, Credibility faces an analogous [challenge regarding imagining impossibilities]. Fictions can describe (physically or metaphysically) impossible worlds. Consider Philip K. Dick's *The Minority Report*, premissed on the existence of individuals endowed with precognition… 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 (2021, p. 474)

The biggest problem here is that imagination is straightforwardly unreliable as a source of evidence about what is objectively possible, since we can coherently imagine impossibilities. If credibility does turn out to be a form of imagination, then credibility will be equally unreliable.

However, the reliability or unreliability of credibility aside, the deeper issue is that credibility would not answer my concern about evaluating hypothetical models' surrogative function. To see why I say this, consider again how surrogative reasoning works: one considers a model system as its own object of observation and manipulation, and after determining what happens *in the model system*, those in-the-model facts are interpreted as claims about the target system (Hughes, 1997; Mäki, 2009). Surrogative reasoning is essentially comparative or relational in nature; surrogacy is a relation between a model system and a target system. Hence, evaluating how *accurate* a model is as a surrogate is a matter of comparing or relating the model-system facts to the facts in the target system, not merely evaluating the qualities of the model itself.

By contrast, if we are assessing whether a model is credible, it seems (from the best suggestion on offer) that we examine the model's idealizations and claims and see if those claims are credible, plausible, imaginable, and so on. In other words, if credibility is what is supposed to indicate that a model has successfully represented an objective possibility, we aren't evaluating any kind of representational relationship between the model and the target, in this case, a possibility. We would simply be evaluating the model's own properties and structure, evaluating their credibility, and using that as a judgment for whether it has successfully represented a target.[11] Rather than saying, "Look, here's what obtains in the model system, and here's what obtains in the target system; the model matches up to an acceptable degree," judging the success of representation by a model's credibility amounts to saying, "Look, the model system's properties are such-and-such, it is plausibly, coherently imaginable; so, it represents its target successfully." In other words, by relying on credibility, the judgment regarding whether the model succeeds as a surrogate for some nonactual possibility would not be made on the basis of comparing the model against the target. Of course, the target systems in cases of hypothetical modeling are supposed to be empirically inaccessible possibilities, so one might object that it is hard to come by any such comparison to begin with. But this is exactly what the source of the problem is: regardless of what the targets are, models are supposed to be surrogates. What is needed is some way to assess the model against its purported target, and not simply assess the model in terms of its own consistency or plausibility. Hence, even if we granted that a model's credibility could give us evidence about whether a hypothetical model successfully represents an objective possibility, this still would not recover the essential and characteristic surrogate function of scientific models.

[11] This is not to say that the only factors that help decide whether some model is credible or imaginable are those that are intrinsic to the model. Consistency with existing theory or empirical knowledge can help explain why it is that something seems credible or imaginable (e.g., Kung 2010). Thanks to Till Grüne-Yanoff for this point.

The concerns raised in this section and the previous section can be summarized as follows. Models are meant fundamentally to function as surrogates for their targets; and when a model successfully represents its target, that is simply to say that it performs well in its surrogative capacity. But in cases of hypothetical modeling, evaluating that surrogative capacity is nigh impossible. So, insofar as there is another option, i.e., credibility, the epistemology of whether a case of hypothetical modeling has been successful ends up being disconnected from the epistemology of modeling more generally. To summarize the "surrogacy challenge," it would do well for philosophical discussions of hypothetical modeling to show how there can be an epistemology of hypothetical models' accuracy regarding their nonactual targets that is consistent with scientific models' more general surrogative function.

5 A reliability challenge for hypothetical modeling

It's time to introduce the second epistemological challenge regarding hypothetical modeling. This second challenge proceeds where the first one left off. Here is the simplest way to introduce the issue. Suppose that everything I've said in the previous two sections is correct: it is impossible to evaluate hypothetical models' surrogative function, i.e., comparatively evaluate them against their nonactual targets. Now consider the sorts of cases where some hypothetical model *successfully* represents an objective possibility. As previously mentioned, while there are some cases where an alleged possibility turns out not to be possible (e.g., infinite populations, infinitesimal organisms), there are other cases where there is at least some sense in which a hypothetical model turns out to represent part of the space of possibility successfully. (Examples might include general equilibrium models in economics, and as mentioned above, certain kinds of optimality models.) The second epistemological challenge I wish to raise is the following: given that even the successful hypothetical models *are not* evaluated against their targets, what could explain the fact that some of the models nevertheless manage to "get the targets right"? If there really is a lack of connection between the models and the targets, there should be some account of how the models nevertheless manage to be reliable with respect to what is possible.

Look at the issue a different way. Temporarily suppose for the sake of argument that even the kinds of cases which, I've argued, turn out to incorrectly represent certain states of affairs as possible, really turn out to be successful. That is, temporarily suppose that hypothetical modeling *in general* is reliable as a way of finding out what the space of possibility is like. But even if hypothetical models *were* reliable regarding possibilities, there is still a lack of the usual sort of connection between the models and the targets that would help explain why the models are accurate. As a rough gloss, presumably in "normal" cases of modeling, accurate models are accurate because they are revised and updated after manipulation and observation (consider the increasingly accurate historical trajectory of models of atomic structure). But again, with hypothetical models, this comparative evaluation does not occur. So even if hypothetical models turned out to be reliable in general with respect to what's possible, there would still be the puzzle of what *accounts for* the reliability.

What my concern amounts to is the sort of reliability challenge that is often familiar in the epistemology of other domains.[12] The Benacerraf-Field challenge in the philosophy of mathematics is one well-known reliability challenge:

> We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in those entities … Benacerraf's challenge [is to] explain how our beliefs about these remote entities can so well reflect the facts about them … *If it appears in principle impossible to explain this*, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them (Field 1989, p26, emphasis in original)

Field's spatial metaphor that the objects of mathematical belief are "remote" bears repeating. The source of a reliability challenge is not, as Justin Clarke-Doane notes (2022, §3), a generic convince-the-skeptic challenge where the challenge first requires showing that some beliefs or claims really are true. The source of the challenge is to show how even one assumes if the beliefs or claims *are* true, there could be a *connection* between the sorts of things the beliefs or claims are about and the ways in which the beliefs are formed or claims are made. This is the same sort of worry I have about hypothetical modeling: even if hypothetical models *are* a reliable guide to the possible, what could explain how they manage to be reliable?

Notice that when the epistemological challenge is framed this way, introducing an alternative evidential story regarding hypothetical modeling does not solve the problem at all. Consider again the suggestion that a model's credibility or imaginability could be the source of evidence regarding its objective possibility. However, if we take the targets of hypothetical models to be worldly, objective possibilities (that is, ways that *the world itself could have been* independent of our conceiving or imagining) there still does not appear to be the right kind of *link* between the targets of the belief and the source of evidence, i.e., credibility, to establish reliability. It is not as if when we consider a model to be credible or imaginable we are somehow "in contact" with objective possibilities, unless one endorses a deflationary view where the kind of objective possibility required for hypothetical modeling *just is* coherent imaginability.[13] Bringing in an additional source of evidence regarding possibility, like credibility, would just push the question back one stage; the reliability challenge would re-arise for the additional source of evidence.

Importantly, the challenge of explaining hypothetical models' reliability about what's possible is only part of the issue here. The general force of any reliability

[12] See Korman (2019) for an excellent survey of this literature. Another example besides the one mentioned here is "evolutionary debunking" in moral epistemology. If non-naturalist moral realism is correct, then *even if* all of our ordinary moral beliefs are true, they face a reliability challenge (see Vavova, 2015 for an overview of this extensive literature). For the explanation for why we come to have our moral beliefs would presumably be due to selection factors over the long course of human evolutionary history, which is to say, the non-natural moral properties would not enter at all into the explanation for why we have the moral beliefs we do. And if so, there needs to be an explanation for how those beliefs could manage to coincide so miraculously with the moral facts. Absent such an explanation, one naturally worries that those moral beliefs are unjustified even if they are true.

[13] Importantly, this opens the path to a possible response—more on this later, in the concluding section of the paper.

challenge is that if it appears impossible to explain the reliability of some method for generating beliefs about some "remote" domain (e.g., mathematics, or in the current case, modality), then those beliefs appear to be unjustified even if they turn out to be true.[14] For absent such an explanation, we would have no good reason to think that those methods really *are* reliable. This, too, is exactly part of the issue in the case of hypothetical modeling. One would think that, in general, the explanation for why scientific models are reliable would be tied to their surrogative function: again, other models can be revised and updated after being compared to their targets. But this explanation is exactly what is unavailable in the case of hypothetical modeling. So, it appears genuinely difficult to see what could account for the reliability of hypothetical models; and accordingly, one might begin to worry that the modal beliefs provided by hypothetical models would be unjustified even if they turned out to be true.

To be sure, there might be case-by-case examples of hypothetical models where the reliability challenge can be met, and we can avoid the specter of defeat. For instance, Michela Massimi (2019) gives some cases of hypothetical modeling which aim to show how something might be possible (if not actual) given the actual laws of nature, and which are hence thoroughly consistent with existing laws of nature. In these sorts of cases, one might plausibly explain the reliability of these hypothetical models at identifying worldly possibilities in terms of our general reliability at identifying laws of nature.

However, it's important to note that this sort of strategy will probably not generalize. This is because there appear to be some examples of hypothetical modeling which face the reliability challenge even when their targets appear to be objectively possible (e.g., are consistent with the laws of nature). Consider what D. Wade Hands says about general equilibrium models:

> Notice that an actual perfectly competitive economy is an economy that has never existed nor ever will exist; it is a hypothetical economy with no monopolists, no oligopolies, no production by the government or non-profits, no brand names, free exit and entry into every industry, and a host of other features. And yet it is a *possible world.* It does not violate any laws of logic or nature (Hands, 2016, p83, emphasis in original)

If Hands is right, this is an example of an objective possibility regarding which the reliability of the model would be hard to account for. This is because of his observations in the first sentence of this excerpt. Even if we granted Hands that a perfectly competitive economy is objectively possible, given that a perfectly competitive economy has never existed and never will, what could explain the reliability of hypothetical modeling *methods* with respect to these economies?

One might suggest, as Hands continues on to observe, that certain sectors of real-world markets, e.g., agriculture, approximate the behaviors of the merely-possible markets (*ibid*). But observing that this degree of approximation obtains is not an explanation of the reliability of the modeling methods with respect to their hypothetical

[14] I am not the first person to observe that a reliability challenge arises for knowledge of modality. The "integration challenge" (Peacocke 1999) in modal epistemology often presented as a version of a reliability challenge as well; for discussion of both integration and reliability challenges, see Thomasson (2018), Wang (2018), and Sjölin Wirling (2021), among many others.

targets. Instead, it is some defeasible confirmation that they *are* reliable. An explanation of where the reliability came from is still needed. Of course, if hypothetical models were constructed *by way of* how closely they approximated sectors of real-world markets, then of course the reliability challenge would be easy to answer.[15] But to recall an earlier point, the methods by which many hypothetical models are constructed are stipulatory and entirely "autonomous" from preexisting theory and observation (Reutlinger et al., 2018). So, the worry remains: *because* the construction methods of hypothetical models often don't have any connection either to real-world targets, nor to their merely-possible targets, an explanation for the reliability is hard to locate. How could such "autonomous," stipulative methods have been reliable concerning worldly possibilities? Even if we had some defeasible evidence that a hypothetical model *is* reliable regarding worldly possibilities, in some cases there will still be no story for how the *methods* by which the model was constructed could have achieved such reliability, and the issue will rear its head again.

All the preceding can be summarized as follows. Because the quality of the putative surrogative or representational relationship between hypothetical models and their target nonactual possibilities cannot be evaluated, hypothetical models face a reliability challenge. In ordinary instances of modeling, we can modify or update the models in view of what the targets' behavior seems to be. However, unlike ordinary modeling, hypothetical models' content appears to necessarily be totally independent of anything we've "observed" regarding their targets. If they managed to a reliable guide to those objective possibilities nevertheless, that would be a strange coincidence. So we need an explanation of how hypothetical models could manage to be a reliable guide regarding what is possible. Introducing an alternative source of evidence regarding modality, such as credibility, likewise brings up the reliability challenge. There would need to be some explanation of how hypothetical models or their corresponding credibility judgments could manage to be a reliable guide to the properties of merely possible, stipulated systems, some of which appear to be nomologically impossible or at least never-to-be-physically-realized. Given all of these concerns, it is simply a bit difficult to see how such an explanation for the modal reliability of hypothetical models could be given. This challenge seems worth exploring more in future work on the epistemology of hypothetical modeling.

6 Conclusion (and Some Possible Solutions)

I have presented two epistemological challenges regarding hypothetical modeling in the sciences. First, as a baseline issue, it seems that there are cases where what a hypothetical model represents as possible turns out not to be objectively possible. In other words, given commonplace philosophical assumptions about the scope of objective possibility, hypothetical modeling is fallibilistic when it comes to what is possible.

[15] And if there are such cases, the "surrogacy challenge" raised in the previous sections will also be easy to answer.

Because of this, there is a basic epistemological issue in identifying the kinds of evidence there could be that a hypothetical model has been successful in representing a possibility. To be sure, there are sources of evidence that have been suggested—e.g., a model's credibility, or perhaps a model's coherence with established theory. I have argued that none of these options is satisfactorily consistent with the idea that an epistemology of modeling should be based in models' characteristic surrogate function.

Second, I have argued that hypothetical modeling faces a reliability challenge. Given that there can be no evaluation of the model-to-target relation, if the targets of hypothetical models really are nonactual possibilities, there is a challenge in explaining *how* hypothetical models turn out to be reliable with respect to those targets. Again, alternative options such as credibility seem to have little purchase on the challenge, since in this case, introducing an alternative source of evidence simply reintroduces the reliability challenge.

I mentioned near the beginning of the paper that my aim in presenting these two challenges is not to present what I take to be definitive objections to prevailing philosophical interpretations of hypothetical modeling. All I intended was to flag a set of issues which future work on this topic might fruitfully benefit from addressing. Indeed, I think that there are some nascent strategies which may prove promising as a response to one or both challenges I've raised.

Let me discuss one option which I think is somewhat less promising than the next one I'll mention. As we have seen throughout the paper, the baseline issue is that some targets of hypothetical models which are alleged to be possible turn out not to be objectively possible, according to widely held views about the scope of objective possibility. According to these widely held views about objective or worldly possibility, not every conceptual or logical possibility turns out to be an objective or worldly possibility. This issue briefly reoccurs in the reliability challenge for hypothetical modeling, where I noted that it is difficult to account for the reliability of hypothetical modeling unless one thinks the kind of possibility involved in hypothetical modeling *just is* coherent imaginability.

So, a first possible response worth noting would be to respond that the sense of possibility relevant to hypothetical modeling just is something like conceptual, logical, or mathematical possibility. This would be a bit deflationary. It would mean that there is at least sense that hypothetical models aren't genuinely about the targets they appear to be about. If there are conceptual or mathematical possibilities that can't ever be physically realized, but ecosystems, markets, selection processes (and so on) are physical, it starts to look a little dubious as to how the mathematical possibilities can help us learn *about* those systems. Indeed, for some kinds of hypothetical models, there is criticism along exactly these lines. Rosenberg (1992) and Blaug (2003) complain that if the sense in which general equilibrium models represent possibilities is mathematical possibility, the models are "empty" and of no epistemic value to economics.[16] Still, in some other cases of hypothetical modeling, Gelfert (2019) notes that there is epistemic value in simply having a "proof of concept," which is a sense of possibility that might not require more than logical or mathematical possibility. At any rate, this first avenue

[16] Thanks to Grüne-Yanoff and Verreault-Julien (2021, p. 119) for bringing these references to my attention.

of response certainly seems to be an available one; but to my ears it brings up some additional and unneeded complications.

There is a second avenue of response which I think is more promising. An issue regarding hypothetical modeling that is even *more* basic than what I have dubbed the "baseline" issue is that hypothetical models appear to be targeted at nonactual possibilities or have been interpreted as being so directed. It is *because* their targets are taken to be nonactual possibilities that all these issues arise. Both the surrogacy challenge and the reliability challenge can be traced to the idea that the targets of hypothetical models are nonactual. So, the simplest and most effective way to respond to these epistemological challenges is to view hypothetical models as instead representing actual targets in some sense or another. While this may seem to go against prevailing philosophical wisdom, and indeed sometimes scientists' own views of these models, there has been nascent precedent defending this view.

Nguyen (2020) has argued that some toy models typically taken to be hypothetical models in fact *do* represent actual systems. Namely, by Nguyen's lights, they represent broadly dispositional properties these systems instantiate: say, in Akerlof's market of lemons, an increase in the asymmetry of information in a market contributes *ceteris paribus* to equilibrium failure. Nguyen's arguments are aimed at the observation that the kinds of possible systems which appear to be the contents of these models are often highly idealized in *ceteris paribus* ways that in principle admit of interpretation in dispositional terms.

Additionally, Philippe Verreault-Julien (forthcoming) has argued that hypothetical models in fact represent actual systems. His arguments turn on the following observation, which seems to me fundamentally correct. The observation is that even if the *contents* of the models appear to be nonactual possibilities, the models are still used to reason about actual systems. Verreault Julien writes, "if the model has no real-world target, then we shouldn't be able to use it to make true inferences about actual phenomena. And if we can use is to draw such inferences, then the model must have a real-world, actual, target. To reason about actual phenomena with a model implies the model has an actual target" (forthcoming, pp5-6). This seems right. General equilibrium models and other equilibrium models in the economic sciences are generally *used* to reason about real-world targets; they serve a role as an ideal benchmark that can help diagnose what might be causing inefficiencies in real markets (Verreault-Julien forthcoming, §3.3; also Jhun, 2018). Likewise, optimality models typically are used to show how an *actual* phenotypic trait could have arisen in ideal conditions; closeness to an optimality model's predictions can be used to confirm hypotheses about how selection actually occurred. And in some cases of optimality modeling, such as Fisher's models of n-sexed populations, the hypothetical system is used to explain why something we observe (e.g., only two sexes) *in fact* occurs or is likely to occur (cf. Weisberg, 2013, §7.4). So, hypothetical models most frequently seem to help us learn about actual systems. I take all this to be good reason to support the nascent interpretation that the targets of hypothetical models are *not* nonactual possibilities, but instead are actual systems.

In any case, what I have aimed to provide in this paper is a pair of difficulties concerning the epistemology of hypothetical modeling in the sciences. I have indicated that these issues give at least preliminary support to a new philosophical interpretation

of hypothetical modeling. There may be other solutions yet worth supporting; that will be a worthwhile topic for future discussion.

Acknowledgements A predecessor of this paper was presented at Till Grüne-Yanoff and Ylwa Sjölin Wirling's workshop "Modal Modeling in the Sciences," as well as at the University of Birmingham and Fordham University. Thanks to audiences at all these venues for comments on the paper, including Otavio Bueno, Mike Hicks, Brian McLoone, and others. Thanks also to Jonathan Barker, Harjit Bhogal, and Jenn Wang for discussion of this paper. And finally, special thanks to two referees, Philippe Verreault-Julien, and to both Till Grüne-Yanoff and Ylwa Sjölin Wirling for their detailed comments on this paper, and for organizing a great workshop.

Funding This research was not supported by any external funding.

Declarations

Conflict of interest I have no conflicts of interest to report.

References

Bokulich, A. (2011). How scientific models can explain. *Synthese, 180*(1), 33–45. https://doi.org/10.1007/s11229-009-9565-1

Clarke-Doane, J. (2022). *Mathematics and metaphilosophy*. Cambridge University Press.

Fine, K. (2002). Varieties of necessity. In T. S. Gendler & J. Hawthorne (eds.), *Conceivability and possibility*. pp. 253–281.

Fischer, B. (2016). A theory-based epistemology of modality. *Canadian Journal of Philosophy, 46*(2), 228–247.

Fischer, B. (2017). *Modal justification via theories*. Springer.

Frigg, R. & Nguyen, J. (2017). Models and representation. In L. Magnani & T. Bertolotti (eds.), *Springer handbook of model-based science*. pp. 49–102.

Frisch, M. (2015). Users, structures, and representation. *British Journal for the Philosophy of Science, 66*(2), 285–306.

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*. Springer.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese, 172*(2), 269–281.

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), 850–861. https://doi.org/10.1086/673893

Grüne-Yanoff, T., & Verreault-Julien, P. (2021). How-possibly explanations in economics: Anything goes? *Journal of Economic Methodology, 28*(1), 114–123.

Hands, D. W. (2016). Derivational robustness, credible substitute systems and mathematical economic models: The case of stability analysis in Walrasian general equilibrium theory. *European Journal for Philosophy of Science, 6*(1), 31–53.

Hirvonen, I., Koskinen, R., & Pättiniemi, I. (2021). Modal inferences in science: A tale of two epistemologies. *Synthese, 199*(5–6), 13823–13843.

Hughes, R. (1997). Models and representation. *Philosophy of Science, 64*(4), 336.

Jacquart, M. (2020). Observations, simulations, and reasoning in astrophysics. *Philosophy of Science, 87*(5), 1209–1220.

Jhun, J. S. (2018). What's the point of ceteris paribus? or, how to understand supply and demand curves. *Philosophy of Science, 85*(2), 271–292.

Knuuttila, T. and Loettgers, A. (2022). (Un)Easily possible synthetic biology. *Philosophy of Science*.

Korman, D. Z. (2019). Debunking arguments. *Compass, 14*(12), 89.

Kung, P. (2016). You really do imagine it: Against error theories of imagination. *Noûs, 50*(1), 90–120.

Levy, A. (2015). Modeling without models. *Philosophical Studies, 172*(3), 781–798.
Mäki, U. (2009). Missing the world: Models as isolations and credible surrogate systems. *Erkenntnis, 70*(1), 29–43.
Massimi, M. (2019). Two kinds of exploratory models. *Philosophy of Science, 86*(5), 869–881.
Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(5427), 15–18.
McLoone, B. (2020). Calculus and counterpossibles in science. *Synthese, 198*(12), 12153–12174.
Nguyen, J. (2020). It's not a game: Accurate representation with toy models. *British Journal for the Philosophy of Science, 71*(3), 1013–1041.
Parker, W. S. (2017). Computer simulation measurement and data assimilation. *The British Journal for the Philosophy of Science, 68*(1), 273–304. https://doi.org/10.1093/bjps/axv037
Plantinga, A. (1974). *The nature of necessity*. Clarendon Press.
Potochnik, A. (2009). Optimality modeling in a suboptimal world. *Biology and Philosophy, 24*(2), 183–197.
Rami, K. (2017). Synthetic biology and the search for alternative genetic systems: Taking how-possibly models seriously. *European Journal for Philosophy of Science, 7*(3), 493–506.
Reutlinger, A., Hangleiter, D., & Hartmann, S. (2018). Understanding (with) toy models. *British Journal for the Philosophy of Science, 69*(4), 1069–1099.
Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs, 49*(3), 589–615.
Rice, C. C. (2012). Optimality explanations: A plea for an alternative approach. *Biology and Philosophy, 27*(5), 685–703.
Rohwer, Y., & Rice, C. (2013). Hypothetical pattern idealization and explanatory models. *Philosophy of Science, 80*(3), 334–355.
Sider, T. (2003). Reductive theories of modality. In M. J. Loux & D. W. Zimmerman (Eds.), *The oxford handbook of metaphysics* (pp. 180–208). Oxford University Press.
Sjölin-Wirling, Y. (2021). Is credibility a guide to possibility? A challenge for toy models in science. *Analysis, 81*(3), 470–478.
Sjölin-Wirling, Y., & Grüne-Yanoff, T. (2021). The epistemology of modal modeling. *Philosophy Compass, 16*(10), 12775.
Sjölin Wirling, Y & Grüne-Yanoff, T. (forthcoming). Epistemic and objective possibility in science. *British Journal for the Philosophy of Science*.
Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*(5), 767–779.
Suárez, M. (2015). Deflationary representation inference and practice. *Studies in History and Philosophy of Science Part A ,49*, 36–47. https://doi.org/10.1016/j.shpsa.2014.11.001
Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology, 7*(1), 1–31.
Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis, 70*(1), 3–27.
Szabo Gendler, T. & Hawthorne, J. (eds.) (2002). *Conceivability and possibility*. Oxford University Press.
Tan, P. (2019). Counterpossible non-vacuity in scientific practice. *Journal of Philosophy, 116*(1), 32–60.
Thomasson, A. L. (2018). How can we come to know metaphysical modal truths? *Synthese, 198*(Suppl 8), 2077–2106.
Vaidya, A. J., & Wallner, M. (2021). The epistemology of modality and the problem of modal epistemic friction. *Synthese, 198*(Suppl 8), 1909–1935.
van Riel, R. (2015). The content of model-based information. *Synthese, 192*(12), 3839–3858.
Vavova, K. (2015). Evolutionary debunking of moral realism. *Philosophy Compass, 10*(2), 104–116.
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, Philippe (forthcoming). Representing non-actual targets? *Philosophy of Science*:1–10.
Wang, J. (2018). The epistemological objection to modal primitivism. *Synthese, 198*(Suppl 8), 1887–1898.
Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford University Press.
Williamson, T. (2016). Modal science. *Canadian Journal of Philosophy, 46*(4–5), 453–492.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part I: A counterfactual account. *Noûs, 37*(1), 1–24.