Chapter 10
Assessing the Credibility of Conceptual Models

Axel Gelfert

Abstract Whether or not the results of a computer simulation are credible depends to a large extent on the credibility (or lack thereof) of the underlying conceptual model. If a model has been developed explicitly with the goal of running a computer simulation in mind, the two types of credibility may seem deeply intertwined. Yet, often enough, conceptual models predate the subsequent development of simulation techniques, or were first developed outside the context of computer simulation. In such a situation, the specific contribution that a conceptual model makes to the credibility of a simulation requires considerable analysis. How, then, should we assess the credibility of a conceptual model, and which factors ought to play a role in judging whether simulation results derived on its basis are trustworthy? In order to answer these questions, the present chapter begins with the premise that models are never by themselves credible *simpliciter*, but acquire credibility within a given context of inquiry, which itself depends on the cognitive interests of the inquirer. Judgments concerning the credibility of a conceptual model thus need to be based partly on a characterization of the intrinsic features of the model, partly on the cognitive goals and interests of its users. This realization helps explain why credible models have been variously understood as (pragmatically and empirically) adequate representations of real-world target systems, as constructions of "credible worlds" that display internal coherence, and as exploratory tools that may aid our understanding even before a well-developed underlying theory takes shape.

Keywords Credibility · Scientific modeling · Conceptual model · Empirical fit · Causal understanding

---

A. Gelfert (✉)
Technische Universität Berlin, Straße des 17. Juni 135,
FG Theoretische Philosophie H72, 10623 Berlin, Germany
e-mail: a.gelfert@tu-berlin.de

C. Beisbart and N. J. Saam (eds.), *Computer Simulation Validation*,
Simulation Foundations, Methods and Applications,
https://doi.org/10.1007/978-3-319-70766-2_10

10.1 Introduction

Contemporary science would not be what it is, were it not for the emergence of computer simulation techniques from the mid-twentieth century onwards. First pioneered in such disciplines as nuclear physics and meteorology, numerical methods for implementing computer simulations have since spread to a wide range of scientific disciplines, including astrophysics, high-energy physics, materials science, quantum chemistry, biochemistry, molecular biology, ecology, climate science, economics, sociodynamics, and many others. This applies equally to fundamental and applied research, and also extends to neighboring disciplines such as design and engineering. As things stand, much of our best—and instrumentally most important—scientific knowledge is best thought of as being *simulation-based*.

Typically, discussions of computer simulation go hand in hand with discussions of scientific models, and for good reason: modeling and computer simulation both are modes of inquiry that scientists engage in when the behavior of a target system cannot easily be derived from an underlying theory—either because this task is too complex (and, for example, does not allow for an analytical solution) or because no underlying theory can be unequivocally specified (e.g., since the phenomenon in question is the result of a heterogeneous mix of factors). In order to construct a model and implement any simulation, decisions need to be made—however implicitly—about how to represent the target system in the real world that is the subject of the proposed simulation study. Such decisions concern, amongst others, the level of detail, overall structure, relevant factors, and purported causal mechanisms of the target system that is to be represented and simulated. Constructing a model of this sort will not only guide future inquiry, but will also shape how we should interpret any subsequent results; in this sense, specifying a model determines—at least in broad, qualitative terms—the very content of a simulation study.

This suggests that whether or not the results of a simulation study are credible also hinges on the credibility, or lack thereof, of the underlying model, which is often called the conceptual model. Crudely speaking, and notwithstanding the specific insights that only computer simulation methods can afford, a simulation can only ever be as good as the conceptual model on which it is based. It seems legitimate, then, to spend some time reflecting on what makes models credible in the first place, and it is precisely this question that the present chapter tackles. Yet models are never by themselves credible *simpliciter*, but acquire credibility within a given context of inquiry, which itself depends on the cognitive interests on the inquirer. To the extent that models are credible, then, they are credible *for an inquirer* in a particular problem situation, or so it will be argued. Combining these two perspectives—a focus on the particular model at hand, and a recognition of their significance to the model user—thus leads to the thesis that judgments concerning the credibility of a conceptual model need to be based on a characterization of the intrinsic features of the model as well as of the cognitive goals and interests of its users. This also helps explain why credible models have been variously understood as (pragmatically and empirically) adequate representations of real-world targets, as constructions of "credible worlds" that display internal coherence (and may also serve as an "intuitionpump"), or as exploratory tools that may aid our understanding even before a well-developed theoretical account of a phenomenon has become available.

The remainder of this chapter is organized as follows: Sect. 10.2 reviews the twin notions of verification and validation in connection with computer simulation and relates them to the multiplicity of functions of scientific models more generally. Section 10.3 takes our commonsense understanding of "credibility" (in connection with human interlocutors) as the starting point for a discussion of the qualities required for attributions of credibility to nonhuman agents and entities; Sect. 10.4 then applies these ideas to examples from scientific practice. Sections 10.5 and 10.6, respectively, reflect on specific—sometimes competing—criteria for model credibility: empirical fit and furthering causal understanding on the one hand (Sect. 10.5), and the construction of "credible worlds" along with exploratory uses of models on the other hand (Sect. 10.6). The chapter concludes with a brief reflection on how credibility is jointly constituted by features of the model itself and the overall goals and cognitive interests of its users (Sect. 10.7).

10.2 Simulation-Based Knowledge, Verification, Validity, and the Function of Models

Whether a particular scientific investigation begins with data collection or theoretical analysis, once we turn to computational methods to derive simulation-based knowledge, conceptual models are never very far off. If studying a problem using computer simulation is the explicit goal of a given process of inquiry, then deriving a conceptual model may be understood as an important first step in preparing the problem for the deployment of computer simulation methods. Yet, arguably, many conceptual models in science were developed independently of concerns with computer simulation—either because they predate the development of powerful simulation techniques or because they initially served independent illustrative, descriptive, or explanatory purposes and were only later found to lend themselves to computer simulation. Tackling a scientific problem using computer modeling is, rightly, often characterized as a multi-step process, with decisions concerning the problem domain preceding the process of designing a simulation model. This model can then be implemented on a given type of computer system. Thus, even before a conceptual model is proposed, "a description of the problem situation and the system in which the problem situation resides" (Robinson 2011, p. 1432) must be given. Where an underlying fundamental theory can reasonably be assumed to exist (e.g., in the case of planetary motion), this may involve specifying the relevant background assumptions (e.g., considering only objects on closed orbits and looking into their relative position to one another); where a fundamental theory is absent (e.g., in the case of modeling vehicular traffic flow), this may involve selecting well-delineated research questions ("How do spontaneous traffic jams occur?") and specifying the variables that are thought to best describe the phenomenon in question (say, measurable changes in average vehicular speed).

In recent years there has been a growing recognition of the heterogeneous character of model building in science, where models are constructed as representations to investigate a target system, not merely conceived of as realizations of theoretical relationships that are posited as true. On this view, not all models are "derived" from fundamental theory—not only because many research questions require models to be "made up from a mixture of elements, including those from outside the domain of investigation" (Morrison and Morgan 1999, p. 23), but also because models often represent phenomena that have yet to be subsumed under anything resembling a self-contained underlying theory. In the absence of a theory of the domain of investigation, models may thus serve an exploratory function (Gelfert 2016): it is by constructing models that scientists attempt to find out whether a purported phenomenon really does survive closer scrutiny and try to devise "proto-theories" whose relation to more fundamental theories then needs to be analyzed further. It is only once a decision has been made that a particular problem situation can, at least tentatively, be characterized using a certain set of representational tools that a *conceptual model* can be developed. Or, to put it another way, any conceptual model that is being advanced—be it in the form of a set of mathematical equations with corresponding physical interpretations, or as a less formalized set of assumptions, simplifications, and explanatory mechanisms—will already implicitly include *some* assumptions about which sorts of problem situations and approaches are appropriate. It is important to keep this in mind since, as we shall see, a conceptual model's *credibility* partly depends on it.

If we were only interested in constructing a model to represent a target system or phenomenon, arriving at a viable conceptual model might be considered a fitting conclusion to the process of scientific modeling. And indeed, a fair amount of philosophical writing on scientific models takes the construction of models to be the goal and outcome of scientific inquiry—even as it acknowledges that scientific models serve a variety of epistemic and non-epistemic functions. In this sense, what simulationists call the "conceptual model" is generally referred to by philosophers of science as "scientific model" *simpliciter*. Yet, the work of the simulationist does not stop with devising a conceptual model; instead, it is taken as a starting point for the further processes of designing a numerical model in computer code, and implementing the numerical model on a computer model. From the simulationist's viewpoint, the conceptual model is "a non-software specific description of the computer simulation model (that will be, is or has been developed), describing the objectives, inputs, outputs, content, assumptions and simplifications of the model" (Robinson 2008, p. 283). Seen in this light, the construction of a conceptual model by a simulationist is teleologically oriented toward the creation of a piece of software, the computer simulation model, in order to answer specific hypotheses about the target system by running a simulation. It would be hasty, however, to infer from this that the conceptual model play s a merely auxiliary role. Simulationists are keenly aware that many conceptual models enjoy independent support and justification, and that the success of implementing a simulation model partly depends on adequately translating the conceptual model into computer code.

The relationship between a (numerically implemented) simulation model and the (underlying) conceptual model is a nontrivial one, as becomes evident once we consider what we can infer—about the simulation and the conceptual model, respectively—from the empirical success (or lack thereof) of the numerical output thus generated. Even before we compare the numerical output against our observations or measurements, we may ask whether the simulation model does, indeed, adequately reflect the conceptual model; that is, we may engage in *verification* (see Chap. 11 by Rider and Chap. 12 by Roache in this volume). On the one hand, this involves "determining that a simulation computer program performs as intended, i.e., debugging the computer program" (Law and Kelton 1991, p. 299); in a departure from established usage in philosophy of science, the term "verification," thus understood, does not refer to the process of generating observable predictions and testing them empirically. The guiding idea behind verification in this sense is not *empirical testing*, but *formal demonstration* in the spirit of logic and mathematics: "Purely formal structures are verifiable because they can be proved by symbolic manipulation, and the meaning of these symbols is fixed and not contingent on empirically based input parameters" (Oreskes et al. 1994, p. 641). What is being verified, then, is not that the model is a successful representation of the target system, nor that the numerical output matches empirical observations, but rather that an already constructed conceptual model is correctly solved in the software code that constitutes the (implemented) simulation model. In this sense, verification (also called "technical validation") aims to demonstrate internal consistency, sometimes by way of benchmarking numerical results against analytical solutions (where available). On the other hand, this usage of "verification" involves the problematic assumption that formally verifying that a computer simulation approximately solves a set of underlying mathematical equations answers all relevant questions that may arise in the course of computationally implementing a conceptual model. For example, there is considerable latitude in how one should discretize the underlying equations in a simulation model, and each such choice may have advantages and drawbacks, yet these are not a matter of meeting (or failing) certain formal standards. It is, therefore, important to realize that verification is itself typically part of an iterative process of implementing and subsequently tweaking a simulation model and its implementation. As Eric Winsberg puts it, "there can be no justification of the final [conceptual] model that is independent of its discretized implementation, and there can be no justification of the implementation that is independent of the model" (Winsberg 2018, p. 158).

Turning to the second element in the often jointly used phrase "verification and validation," *validation* aims to ascertain the simulation model's performance across a range of empirical contexts, for example by simulating a real data source and comparing the calculated outcomes with real-world observations. (See Chap. 4 by Murray-Smith in this volume.) This can be a formidable task, given that "directly making the validity assessments requires technical expertise and full access to the model and external data" (Caro et al. 2014, p. 178). The ability to predict, or otherwise reproduce, empirical aspects of the behavior of the target system is key to a simulation model's *external validity*, where this refers to the generalizability of the findings of a simulation to the intended class of real-world cases. Different strategies can be pursued: at minimum, an implemented computer simulation should be able to adequately reproduce data sources that went into the creation of the model in the first place ("dependent validation"), though generally, it will be preferable to test a simulation's performance with respect to independent data sets—that is, data that is of a type that the simulation should be able to account for, but which was not actively utilized in the process of simulation design. As in the case of testing theories, the predictive capacity of simulation models, too—that is, its ability to predict empirical results before they have been measured or observed—is often seen as carrying great weight ("predictive validation"). At the same time, drawing too close a parallel between validation of a simulation model and theory-testing, can be misleading. Scientists themselves have occasionally emphasized that "[v]alidation is not a procedure for testing scientific theory or for certifying the "truth" of current scientific understanding, nor is it a required activity of every modelling project" (Rykiel 1996, p. 299), and have lamented the widespread belief "that validation establishes the veracity of the model" (Oreskes et al. 1994, p. 642).

To be sure, textbook definitions of "validation" in computer simulation studies often equate it with "determining whether a simulation model (as opposed to the computer program) is an accurate representation of the system" (Law and Kelton 2000, p. 265), sometimes with the caveat that a model should be "an accurate representation of the real world *from the perspective of the intended uses of the model*" (ITT, cited after Zacharias et al. 2008, p. 302). Critics of such definitions have argued that "[t]he implication is that validated models tell us how the world really is", when we should always keep in mind that any agreement between observed measurements and simulation results "in no way demonstrates that the model that produced the output is an accurate representation of the real system" (Oreskes et al. 1994, p. 642). It seems fair to assume, however, that the authors quoted above intended their textbook definitions to be understood elliptically, in that they would not deny that further assumptions and inferences are required to warrant moving from empirical validation of the simulation results to representational success of the underlying model. Perhaps, then, validation is best understood as both relative to the context of inquiry and the goal of the inquirer:

> Validation is a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the *intended application of the model.* (Rykiel 1996, p. 233; italics original.)

What constitutes a "satisfactory range" depends, at least in part, on the applicable standards of empirical performance which, again, vary between users, depending on their goals, and across different contexts: "That is, a model is declared validated within a specific context which is an integral part of the certification. If the context changes, the model must be re-validated." (ibid.)

The recognition that context matters stems from the realization that models and simulation may serve a wide range of purposes, from promoting epistemic goals (e.g., affording insight into the causal basis of a particular phenomenon) to non-epistemic objectives (e.g., serving as the basis for policy decisions). It also reflects the fact—well-known to scientific practitioners, but perhaps underappreciated in philosophy of science—that the process of modeling and simulation does not, in practice, divide up neatly into distinct "phases" of (1) constructing a conceptual model, (2) translating it into computer code and verifying that the code correctly implements it, and (3) validating the numerical output by comparison with empirical observations. Instead, what one finds—often, not always—is that the various phases overlap and are deeply intertwined. For example, it is not uncommon for approximations that were initially employed during the stage of implementation and verification to become an essential part of the simulation model as a whole, and for them to be credited with ensuring its overall empirical success (and, thus, its validity).[1]

Few practitioners of computer simulation studies would consider any given validation of a simulation model to be sacrosanct. Not only is it the case that, as mentioned earlier, models that have been validated within one context for a given purpose may need to be re-validated for a different use in another context. Like models and simulations in general, the practice of validation, too, serves different (if often complementary) purposes. Sometimes, when the mutual relationships between the numerical output, the simulation model, the conceptual model, and the underlying theory are well-understood, the successful validation of a computer simulation may indeed be interpreted as (good, but defeasible) evidence that the underlying model adequately represents the way the world is. Often, however, the route from a scientific problem to a computer simulation model is less clear-cut, and the various relationships between models, theories, and simulations are contested. In such a situation, validation may also function as a tool "for building model credibility in the user community" (Rykiel 1996, p. 230). This is not to say that validation is being carried out strategically, let alone with manipulative intent to persuade others; rather, validation ensures that a computer simulation adheres to shared standards of accuracy in the given context of inquiry.

An excessive focus on verification and validation may, on occasion, obscure other functions of models and simulations, beyond their ability to reproduce observed phenomena. As the ecologist Edward Rykiel puts it, "modelling and the benefits to be gained from it can also be stifled by an overemphasis on model validation" (1996, p. 240). Such benefits include, but are not limited to, exploratory uses of models in the absence of a fully formed theoretical framework, which have recently begun to receive philosophical attention.[2] Indeed, "[e]xploration of model behavior without validation testing is a legitimate, reportable activity" (Rykiel 1996, p. 241). This—along with the realization that, even where validation of a simulation model is possible, not much can be inferred with certainty about whether or not "a model accurately represents the 'actual processes occurring in a real system'" (Oreskes et al. 1994, p. 642)—strongly suggests that, in relying on conceptual models in our simulation design, we implicitly presuppose that those models enjoy independent justification. For this reason, and because models and simulations enjoy considerable autonomy from one another, the rest of this chapter will focus on the diverse sources of credibility of the conceptual model.

[1] For a fascinating case study, see Lenhard (2007).

[2] For a detailed argument that exploration is a core function of scientific modeling, see Gelfert (2016).

10.3 Taking the Notion of "Credibility" Seriously

If we are to gain a deeper understanding—beyond technical measures of fit, distance, fractional variance, etc., (which, in any case, can only be determined post hoc)—of what leads scientists to trust some conceptual models more than others, then it may be best to begin by taking the notion of *credibility* seriously. At the risk of appearing overly literal-minded, in the present section I wish to discuss some of the relevant connotations of "credibility" as a general concept; in the next two sections, we shall then encounter a range of examples of how scientists tend to arrive at judgments of credibility in relation to scientific models.

On the face of it, it may seem puzzling why, in the seemingly neutral context of model evaluation, one should invoke the term "credibility" at all, given its ethical and interpersonal overtones. Why not stick with more objective criteria such as "reliability" or "verisimilitude"? The concept of *credibility* has its natural place in human communication and is of a piece with—albeit slightly less emotionally tinged than—the concept of *trustworthiness*. (Cf. Chap. 17 by Saam in this volume for the role of humans in validation.) Yet it is not by chance that scientists should turn to the notion of credibility in their interactions with scientific models and simulations, or so I wish to suggest. In interpersonal communication, credibility is usually considered to be a function of both the trustworthiness and competence of an agent. To be sure, we also speak of isolated claims as being "credible" if we think they merit belief. However, when it comes to the credibility of models, the closer analogy is with epistemic agents, not with the level of individual propositions. To put it another way, when we believe a model-based prediction, we typically do so because we consider the model to be credible, not because we have independent reason to think that the specific prediction in question is somehow *ex ante* more likely to be true than any of its close competitors. A credible model, then, is one that we can turn to, with some confidence, for answers on a suitably wide range of relevant research questions. Once we consider a model credible, and resolve to work with it for the purposes of inquiry, we begin to trust its results—not blindly, of course, but in a way that grants its results some measure of (defeasible) default justification.

In human interactions, once we trust someone, we depend, at least in part, on their goodwill. When I ask you for directions to the train station and trust your answer, the success of my subsequent actions depends, among other things, on your having chosen not to play a prank on me and send me in the wrong direction. What I end up believing depends, in part, on your mental processes. This is why, for someone to be deemed credible, they must not only be deemed to be competent with respect to the subject matter in question, but must also be trustworthy (that is, honest and sincere). (See e.g., McGinnies and Ward 1980.) In the case of models, while there is no analogy to the involvement of another mind in our own process of belief formation, we likewise depend on factors internal to the model which cannot readily be inspected. Competence and trustworthiness are not categories that can be directly applied to scientific models, but it is not difficult to identify desiderata that are structurally similar. Like a competent interlocutor, a good model should be able to provide reliable information regarding a broad range of thematically related questions; similarly, in much the same way that a trustworthy interlocutor would not suddenly start offering wildly misleading claims, a good model should not exhibit sudden discontinuities in the quality of information we can extract from it.

It is perhaps no surprise that the very term "inquiry" is ambiguous, inasmuch as it refers both to objective empirical investigation and to interpersonal requests for information. Indeed, in the Baconian tradition of experimental natural philosophy that was at the heart of the Scientific Revolution, scientific inquiry was often likened to the process of *interrogation*—sometimes by violent means, as critics of the metaphor have pointed out. (See Merchant 1980.) Scientific experimentation itself was seen as a method of bringing about conditions that allowed for the extraction of truth—a way of "putting Nature to the question" (where this, of course, was a common euphemism for torturing someone at the rack). What matters for present purposes is not the problematic character of Bacon's imagery, but rather the transactional conception of inquiry as *interrogation*. But surely, one might wonder, there is a difference between experimenting on nature by bringing about material conditions that may elicit novel observable phenomena, and "interrogating" a conceptual model by various means of analysis? Indeed, there is; yet, as Joseph Pitt notes (in a passage concerned with scientific theories, but in a way that naturally extends to models), there are also striking parallels:

> When a scientist works with a theory to derive some results, she is in some sort of communication with it. She knows that if she does this she will get, or at least, ought to get this result. It is in her being able to anticipate the response of the theory to her manipulations that she is communicating with it. (Pitt 2007, p. 55)

Pitt intends this to be more than just a useful metaphor for understanding how we engage with theories. The key notion is "manipulation"—which seems even more apt in the case of models which, as mentioned earlier, are often made up of a heterogeneous mixture of elements, arranged precisely in a way to enable inferences about the target system (and, thereby, make it possible to extract information about the world). As Pitt notes, "to the extent that we manipulate theories [*read:* models—A.G.] we communicate with them": "The key here is in knowing how to communicate and with what kinds of things we communicate." (Pitt 2007, p. 55).

There is clearly a metaphorical element to likening scientific inquiry to verbal interrogation—but this is no more problematic in the case of scientific models than with respect to scientific experimentation, or so I wish to suggest. Scientists continuously labor with scientific models—often the same ones, with only minor variations of the same underlying equations or formalisms—and, over time, come to see them as "mediators, contributors, and enablers of scientific knowledge" (Gelfert 2016, p. 127). When they judge a conceptual model to be *credible*, this is more than merely an interim assessment of a model's utility "here and now," but expresses a commitment to its future use and expected fruitfulness. Judgments of credibility, then, play an important part in the evaluation of conceptual models, and it will be insightful, in the next section, to discuss the standards and criteria deployed by scientists in their assessment of whether a conceptual model deserves our trust.

10.4 The Credibility of Models: Lessons from Scientific Practice

Scientists, in assessing the models they are using, are typically less concerned with reporting overall levels of credibility concerning specific models, let alone subjective judgments of their trustworthiness, but instead—rightly so—tend to acknowledge the hybrid nature of model credibility:

> The credibility of a modeling analysis should be assessed at several levels: validation, design, data, analyses, reporting, interpretation, and conflicts of interests. Validation assesses how well the model accords with reality. The design should follow accepted standards for conceptualizing and framing the model. The data used in building model should be suitable for the purpose, properly analyzed, and incorporated in the model. Analyses should provide the information required to support decision maker. (Caro et al. 2014, p. 178)

"Reporting" and "interpretation", which are being acknowledged by the authors as "not specifically pertaining to a model's credibility" (ibid.), nonetheless are central to the credibility of modeling *as a process* and its application in practical contexts. The final point—conflicts of interest—demonstrates how the overall credibility of a modeling analysis depends both on the credibility of the model (according to the criteria specified in the previous section) and on the trustworthiness of the modeler who, after all, has decided to deploy one (type of) model rather than another for a particular purpose.

If, for the time being, we take actual usage at face value, we find that "credibility" for scientists has an—in the eyes of philosophers perhaps surprisingly—instrumentalist character, with strong social connotations. Earlier, we quoted the ecologist Rykiel as arguing that "validation is not an essential activity for evaluating research models, but is important for building model credibility *in the user community*" (Rykiel 1996, p. 230; italics added). Credibility, in turn, is best defined in terms of the demand that, as Stewart Robinson puts it, a model "[b]e believed by the clients" (Robinson 2011, p. 1433). And Caro et al. (2014), the same group of health scientists who gave the pithy characterization of the hybrid character of model credibility quoted above, have also drawn up a list of questions that may guide assessments of "the relevance and credibility of a modeling study". In addition to obvious concerns regarding verification and validation, these include questions such as the following:

> Does the model have sufficient face validity to make its results credible for your decision? Is the design of the model adequate for your decision problem? Are the data used in populating the model suitable for your decision problem? (Caro et al. 2014, p. 176; format adapted from table)

"Face validity," in particular, is thought to be a first criterion by which to screen out straightforwardly implausible proposals, inasmuch as a model should not contain unrealistic and implausible assumption about core elements that a model is intended to get right.[3] Whether a model meets this desideratum is thought to be "the easiest

[3]This should not be understood as contradicting the frequent—and entirely correct—observation that, as William Wimsatt puts it, false models may function "as means to truer theories" (Wimsatt 2007, p. 94).

aspect of credibility for a user to check because it does not require in-depth technical knowhow"; at the same time, if "parts of the model fail face validity, the effect on credibility depends on the user's judgment about whether the questionable parts are so unrealistic or inappropriate that they affect the accuracy of the results" (Caro et al. 2014, p. 179).

As scientists themselves are keen to point out, mere success at reproducing empirical results does not suffice to render a model credible:

> Agreement between model and data does not imply that the modeling assumptions accurately describe the processes producing the observed climate system behavior; it merely indicates that the model is one (of maybe several) that is plausible, meaning that it is empirically adequate. (Knutti 2008, p. 4652)

Just as we do not consider an interlocutor credible merely in virtue of having made a series of truthful assertions, we do not place trust in a model just because it happens to have successfully reproduced some amount of data. Attributions of credibility derive from the warranted presupposition that a source of information is systematically reliable across a range of relevant contexts and questions. Background knowledge, thus, is a key to assessments of the credibility of models: "The model results we trust most are those that we can understand the best, and relate them to simpler models, conceptual or theoretical frameworks". (Knutti 2008, p. 4656)

If there is one near-universal feature of how scientists talk about the credibility of their models, then it would have to be their recognition that models serve a variety of purposes, such that attributions of credibility depend on the goals of the modeling process. This is not to suggest that attributions of model credibility are subjective, or cannot be challenged, but simply to acknowledge that, for such an attribution to be meaningful (and potentially intersubjectively compelling), the goals and contexts of the modeling process need to be specified. This is precisely why Caro et al. (2014), in their proposed list of diagnostic questions concerning the relevance or credibility of a modeling study, ask researchers to consider what external validation, internal validation, face validity, design aspects, etc., of a model-based study have to contribute toward "mak[ing] its results credible *for your decision*?" (178, italics added) Other scientists, in a similar spirit, note that models are to be assessed by their ability to "[p]roduce sufficiently accurate results for the purpose" that a modeler has chosen (Robinson 2011, p. 1433), and in view of how "acceptable for pragmatic purposes" (Rykiel 1996, p. 230) its results are. This reflects, once more, the largely pragmatic-instrumentalist attitude of practicing scientists regarding the utility of models in specific problem-oriented contexts of inquiry.

Wendy Parker, in the context of discussing precisely what is being confirmed when models—climate models in particular—are found to fit with observations and past data, resists the thought (to be discussed in the next section) that "as we accumulate instances of fit between observational data and output from a climate model, we are accumulating evidence of the truth of the hypothesis embodied by the model" (Parker 2009, p. 234). Such a view regards attributions of credibility as "divorced from any particular use or application of the models"; instead, Parker argues, we should recognize a growing need, especially in the case of complex modeling analyses such as in climate science, “to try to discern, in a principled and careful way, what a […] model’s performance” in specific contexts “indicates about its adequacy (or inadequacy) for various predictive and explanatory purposes.” (Parker 2009, p. 243) Like the scientists quoted above, Parker believes that such *adequacy-for-purpose*, for many modeling contexts, is a more important desideratum than truth or wholesale empirical adequacy. As Parker puts it,

> adequacy-for-purpose does not work like truth and empirical adequacy[…:] from the assumption that a model is adequate for an explanatory or predictive purpose, information about how the model is likely to perform in various other respects, or information about what other properties the model is likely to possess, does not simply follow as a matter of course. (Parker 2009, p. 238)

This suggests that assessments of the credibility of models will, by necessity, always have to be tentative and context-dependent—even if, on rare occasions, a model may turn out to be successful and credible across a wide range of questions and applications.

None of this should come as a surprise: after all, scientists use models in situations of incomplete knowledge—for example, because an underlying fundamental theory cannot be directly applied in any straightforward way to the case at hand, or because it is not even available in principle, or because the available data suggests, but does not entail, a particular interpretation of an empirical phenomenon. Models are also being employed in contexts where a “full” description or derivation may simply be too costly, perhaps because it would require too much time, computational resources, or the like. In all of these situations, it is natural to expect that modelers will face trade-offs—e.g., between completeness (of a model description and derivation) and timeliness (of results and predictions), or between the generality, realism, and precision of one’s models (Levins 1966). Given that trade-offs prevent us from maximizing all desiderata simultaneously, and given that different purposes call for different desiderata to be maximized, we simply should not expect to find that the same model is the most adequate across all contexts. For the same reason, it would be largely uninformative to simply extend the same set of “standard criteria for evaluating the adequacy of a theory” (Kuhn 1977, p. 322) —that is, accuracy, consistency, scope, simplicity, fruitfulness, and other theoretical virtues—to the case of models, without specifying in more detail how these—no doubt worthy—desiderata can be achieved and ascertained in the case of modeling. (See also Chap. 40 by Hirsch Hadorn and Baumberger in this volume.) In short, if one’s goal is to achieve adequacy-for-purpose, a more fine-grained approach to assessing the credibility of one’s conceptual model should be favored. Achieving model credibility—especially as the basis for future simulations—is a complex process and, given “the extent to which this process focuses on elements *external* to anything we would reasonably include as part of theory, it would be unrealistic to interpret this warranting process as being about the relationship of the results to some formal model” of a theory (Winsberg 2001, p. S450).

10.5 Empirical Fit and Causal Understanding

A major tension in the notion of model credibility arises from the question of whether models should always strive to fit the actual world, or whether they can serve explanatory (and other legitimate) functions by imagining plausible (yet counterfactual) scenarios. In this section and the next, we will discuss each of these aspects in detail, keeping in mind that they point to a difference in emphasis, rather than to any fundamental inconsistency. Often, a model that is deemed a successful representation of real-world findings will also successfully predict what would happen if conditions were different, for example, because it has correctly identified an underlying causal mechanism. If we are lucky enough to have a well-confirmed fundamental theory at our disposal, and if the particular problem situation at hand poses no special obstacles to the theory's application, we may even be able to derive a model that "inherits," so to speak, the underlying theory's strengths. This is how philosophers of science used to think about scientific models *in general*, before realizing the significantly greater heterogeneity, diversity, and tentativeness of models in actual scientific practice. Yet, wherever it is, in fact, possible to embed a model in a theory, the credibility of the underlying theory can legitimately rub off on the model as well. At the same time, a highly simplified model—perhaps even one that could not, in principle, be realized—may also be a source of insight about why things are the way they are, for example, because it showcases why certain alternative scenarios could not play out in the world as we know it. Still, there exists more than just a difference in emphasis between, on the one hand, treating models primarily as accurate representations of real-world phenomena, and on the other hand, treating them as "credible worlds" (Sugden 2000) in their own right, which allow us to explore relationships which may, or may not, obtain in (or shed light on) the actual world.

The tension between those who regard models as a way of accounting for empirical data and those who are willing to grant models greater independence from empirical phenomena, is largely due to the fact that models occupy a middle ground between theory and data. Some philosophers of scientific models have even gone so far as to claim that the primary function of models is to serve as "mediators" between the empirical world and the realm of theoretical hypotheses: models, on this account, "are *not* situated in the middle of an hierarchical structure between theory and the world", but operate outside the hierarchical "theory-world axis" (Morrison and Morgan 1999, pp. 17–18). On rare occasions, we may be able to describe the modeling process as an instance of *applying* a fundamental theory to a specific case at hand; but more often than not, such a description would be wildly inaccurate, since modeling often involves interpolating between different realms, making multiple (sometimes inconsistent) assumptions, engaging in different rounds of idealization and de-idealization—all of which render scientific models typically "a *mixture* of elements, including those from outside the domain of investigation" (Morrison and Morgan 1999, p. 23). This echoes a sentiment by Nancy Cartwright, who has long held that theories "do not generally represent what happens in the world—only models represent in this way" (Cartwright 1999, p. 180). While such a view of scientific models opens up a wider range of considerations that modelers can hope to draw on in their quest for model credibility, and while it is generally agreed that models are "inherently intended for specific phenomena" (Suárez 1999, p. 75), when the rubber hits the road—viz., in any given actual instance of model-based inquiry—this view is no clearer than its predecessors on when a model should be deemed credible.

Scientists often place great store by a model's ability to reproduce empirical data; yet, given the inevitable simplifications that go into the design of serviceable scientific models, they are at the same time well aware of the fact that perfect empirical fit, even across a range of situations, may be a fluke—or simply the lucky result of errors from different sources canceling out. One way to guard against mistaking such accidental successes of a model for a sign of its overall credibility is to systematically broaden the range of situations being considered and test a model's performance against the corresponding data sets. If a model performs well with respect to a wide range of independent empirical situations and data sets, it becomes progressively unlikely that its successful performance is entirely a matter of chance. This is why, in addition to predictive successes, retrodiction is likewise valued, since it affords an alternative way of comparing a model against empirical reality—provided the past data in question was not itself used in the construction of the model: "A model demonstrates empirical fit to the extent that its logical implications are observed in data; the data may be historical or not yet observed" (Yuengert 2006, p. 87). When viewed from this angle, ascertaining that model *A* has a better empirical fit than model *B* becomes a matter of demonstrating that *A* entails more empirical consequences found in the data than *B*.

While continued empirical success is a good, if fallible, indicator of a model's "being on to something", it is clear that it cannot be the final word on what makes a model credible. As an example from the special sciences, consider economic models of addictive behavior, in particular, rational addiction models and time inconsistency addiction models. (The discussion in this paragraph follows Yuengert 2006.) Rational addiction models depict consumers as forward-looking and seeking to maximize utility over their life cycle, all the while taking into account the future consequences of their (current) choices. Addictive reinforcement, on this model, merely reflects the assumption that an increase in the addictive stock increases the marginal utility of current consumption. This contrasts sharply with time inconsistency models, according to which consumers have self-control problems and cannot trust themselves to enact their consumption plans, even when in possession of full information. Time-inconsistent consumers, at any moment, pursue immediate gratification more than they would have professed to prefer at any previous point in time. One might expect such radically different assumptions to make an observable difference, once the two types of models are put to the test. Yet, interestingly, both types of models "are nearly indistinguishable by conventional econometric methods" (Yuengert 2006, p. 78), and the case rational addiction and time inconsistency models is sometimes regarded as a case of underdetermination by data (see Goldfarb et al. 2001).

In situations where two models are empirically equivalent, yet a decision needs to be made as to which model should be adopted (e.g., because, due to lack of time and resources, it is not possible to pursue both modeling strategies simultaneously), one evidently needs to appeal to selection criteria other than empirical fit. This may take the form of privileging certain theoretical desiderata—notably, simplicity and parsimony—or may be guided by background assumptions about the causal basis of the phenomenon in question. Thus, in the case of addiction models, it has been argued that ensuring that models of habit information be formulated in terms of rational decision-making leads to models that are "formally equivalent to models without habit formation" (Spinnewyn 1981, p. 92), but only by redefining wealth and the cost of current consumption in unwieldy ways; in other words, the rationality assumption "leads to unnecessary complications" (Chaloupka et al. 2000, p. 115). Yet such a conclusion may be unacceptable to proponents of rational choice theory, for whom the rationality assumption is a nonnegotiable core element of their paradigm. In addition to background commitments and general theoretical virtues, however, there is a third set of considerations that stem from realism about the purported causal basis of the phenomenon being modeled. Thus, in the case at hand, those who take independent findings, including on issues unrelated to addiction, to establish beyond doubt the realism and causal significance of time inconsistency for behavior in general, will likely consider an explanation of addiction as being caused by time-inconsistent preferences to be superior.

In recent years, predictive modeling, not least on the basis of AI algorithms and machine learning, has taken hold across a wide range of activities, in the data-centric sciences as well as in business and the corporate world. To be sure, the goal of prediction has been an integral part of science (and, by extension, of scientific modeling and simulation) from early on; only in recent years, however, has it become a realistic prospect to sift through vast amounts of data in the search for correlations and to "train" neural networks on training data in the hope that they will successfully predict new samples (or recognize relevant features in incoming data). Such data-driven approaches bring their own challenges. Without a firm (and independently justifiable) set of prior assumptions, the only justification of such models consists in their continued predictive successes. Common dangers include committing so-called "Type III" errors (i.e., developing a model that answers the wrong question), over-fitting of models to the data (e.g., when a model reflects the structure of a given data set—including its noise—so well that its predictions do not generalize to new data sets *of the same kind*), and ignoring systematic changes in the environment (such that past data fails to be a guide to the future). In such a situation, model-immanent approaches can only do so much to alleviate any shortcomings a model has acquired by way of how it was developed—even when a model appears to be empirically successful, e.g., in relation to a given data set (as in the case of overfitting). Great care must, therefore, go into the very construction of data-driven models, e.g., by deploying more sophisticated sampling techniques.

Empirical fit, then, is only one consideration among several that researchers draw on when they attempt to determine how much credibility a model merits. Theoretical virtues such as parsimony and unification may aid in resolving situations where multiple models are empirically equivalent; appeals to the realism of assumptions likewise have a role to play, in particular in the following two cases: "if realistic assumptions can be expected to result in better empirical fit eventually, or if realistic assumptions promote worthy goals other than empirical fit" (Yuengert 2006, p. 87). To the latter possibility—that empirical fit may be outweighed by, or may at least trade off against, other worthy goals of inquiry—we shall now turn.

10.6 Models and the Exploration of Credible Worlds

Empirical fit and numerical accuracy may indicate that a model stands in the right sort of relation to its target system, and to the world at large, but, for the reasons already outlined, they can at best constitute defeasible evidence that a model satisfies the kind of "world-linking conditions" (Grüne-Yanoff 2009, p. 81) that would merit trust in its future performance and overall credibility. Turning from one set of desiderata (relating to a model's accuracy and empirical fit) to the other cluster of desiderata identified earlier, viz. a model's aptness for exploring possible scenarios and generating "modal knowledge about what might be *possible* about the target system" (Massimi 2018, p. 339), as well as its fruitfulness in generating truth-conducive lines of inquiry, an analogous point can still be made. After all, though it may be difficult to quantify and compare fit and accuracy—let alone infer on their basis how credible a model is *overall*—we may expect assessments of explanatory success and exploratory potential to be even more controversial. Nonetheless, in what follows, I shall sketch two approaches that tackle these more qualitative criteria of how much insight into the world a given model affords us. The first such approach conceives of models as a way of constructing *credible worlds*; the second regards them as *exploratory tools*.

The term "credible world" as a characterization of the way models operate is due to the economist Sugden (2000). It emphasizes that, for a model to afford insight to its user, it need not always be derived from an idealization of an actual target system. Sometimes, such idealizations may be possible: when modeling mechanical motion, friction is often treated as negligible, and the resulting mechanical models may be considered as idealized representations of actual bodies in motion (where friction is inevitable), which nonetheless display wide applicability. Yet, in many areas of science—not least those that deal with complex systems (such as the social sciences, including economics)—it is often difficult, if not impossible, to determine in advance which factors can, or cannot, be neglected. If idealization is understood as the process of starting from real-world target systems and then proceeding by isolating causally important factors from those of minor significance, then this methodology may face serious limitations when it comes to complex systems.[4] By contrast, Sugden's notion of "credible world," in line with other recent accounts of model-based science, acknowledges the central role of model *construction*. Modeling, in essence, amounts to the construction of credible worlds using representational tools (such as mathematics); whether these successfully "latch on" to the actual world is a question

[4]This, it should be noted, is not the only that one may interpret the procedure of theoretically "isolating" relevant factors; for an alternative view, see (Mäki 2009).

which can only be successfully tackled once a model has been specified. Whereas in the case of gradual idealization from real target system the satisfaction of world-linking conditions can be assumed (if, perhaps, only initially), in the case of models as credible worlds, such a linkage needs to be established subsequently—e.g., by relying on such criteria as similarity, induction, and explanatory success. As Chao (2014, p. 591) puts it rather succinctly,

> if a prototype theoretical model can be applied to a set of particular models across time, space, and context, and each particular model is regarded as satisfactorily explaining a particular […] real-world phenomen[on], then it can be inductively concluded that this prototype theoretical model is credible.

Naturally, on this account, an important criterion of the credibility of models is their *coherence*—both internally and with known external (e.g., causal) constraints: "If a model lacks coherence, its results cannot be seen to follow naturally from a clear conception of how the world might be" (Sugden 2000, p. 26). Yet, coherence among the model's assumptions does not itself suffice: "For a model to have credibility, it is not enough that its assumptions cohere with one another; they must also cohere with what is known about causal processes in the real world". (ibid.) However, in order to conclude, with confidence, that a given credible-world model represents the way the world *actually* is, a further step will typically be required: viz., the abductive inference that, "[i]f a result R is caused by a set of causal factors F in the model world M, and R occurs in the real world W, then we have reason to believe that F operates in W" (Chao 2014, p. 592). Failure to establish a model's real-world connection need not, however, disqualify the model from further study: while such a model could not function as a *surrogate* for its real-world target, it may nonetheless *substitute for* real-world inquiry.[5]

A similar conclusion regarding the sources of a model's credibility may be reached from a perspective that acknowledges that models may serve a variety of functions. In addition to representing actual target systems and deriving specific results, predictions, and explanations about them, models also help *explore* further avenues of inquiry. From this perspective, whether or not a model is credible may not be solely a matter of how faithfully it represents a given target system, and how closely its results mirror the latter, but may also depend on how fruitful it is—for example, in the generation of potential explanations, or when it comes to establishing in-principle possibilities (or, as the case may be, impossibility theorems—which may play an important role in guiding future inquiry). On the one hand, this acknowledges that, in order to make headway in our attempts to model reality, we must sometimes introduce falsehoods—not merely as an unavoidable side effect of idealization and abstraction, but as a direct and deliberate consequence of making (sometimes heavy-handed) model assumptions; on the other hand, it broadens the range of legitimate uses of models to also include those instances of modeling that precede the full theoretical articulation of a phenomenon (or class of phenomena). Not all legitimate uses of a model should, of course, be thought of as bolstering its *credibility*. Yet, if

[5] I am here drawing on Uskali Mäki's distinction between "surrogate models" and "substitute models" (Mäki 2009, pp. 35–37).

we take seriously the earlier idea that one can draw a parallel between the credibility of models and the trustworthiness of interlocutors (see Sect. 10.2), it is not at all far-fetched to insist that the credibility of a model goes beyond the brute empirical reliability of the individual claims it makes about the world. Just as one person may be deemed a "more credible choice" (as a candidate for political office, say) than another, some models may be considered more credible than others—not because of any decisive difference in their past track record, but because background considerations suggest that they hold more promise than their competitors. Judgments of credibility, we might say, are forward-looking—if positive, they engender trust in future performance—in a way that is not captured by looking at brute empirical track record alone. This line of argument is entirely compatible with, but does not presuppose, the idea that some models may be best thought of as "credible worlds." Yet it acknowledges more explicitly that "exploratory modeling often serves the purpose of developing a grasp of (as yet theoretically inaccessible) phenomena" (Gelfert 2016, p. 95)—a situation that scientists encounter all the time. It also serves as a reminder that assessments of a model's credibility depend importantly on their role and function: treating a model as a credible "proof of principle" (Gelfert 2016, pp. 85–86), say, is a quite different matter from relying on it as a credible source of precise numerical predictions.

As an example of such exploratory models, consider certain types of "toy models"—viz., models that are strongly idealized and simplified, so much so that they may border on being minimal, "stylized" accounts of a single aspect of a target phenomenon. Some such models may be derived from empirically well-confirmed theories, e.g., when planetary motion is modeled as two point masses orbiting one another. Such cases may be considered "embedded toy models," since they are at the same time models of an underlying well-developed theory *and* extremely simplified and idealized models of phenomena. This contrasts with what has been called "autonomous toy models" (Reutlinger et al. 2017): that is, extremely stylized models that are not models of a theory (and which, in some cases, "seem to bear no relevant relation to a well-confirmed framework theory" at all; ibid.: 11). When the lack of such a relevant relation to an underlying theory is due to the absence of well-developed theoretical resources that one might otherwise draw on, we may properly deem such autonomous toy models "exploratory" in the sense discussed in the previous paragraph. Whereas, in the case of embedded toy models, the underlying theory usually contains some of the resources required for successfully "de-idealizing" and applying a model to a given target situation for predictive purposes, in the case of autonomous toy models—and of exploratory models, more generally—empirical prediction, let alone numerical accuracy, is rarely the explicit goal. Instead, researchers often deploy exploratory models with the aim at generating intelligible explanations of certain types of phenomena. The more we succeed in cultivating our ability to "recognise qualitatively characteristic consequences [...] without performing exact calculations" (De Regt and Dieks 2005, p. 151), the more trust we place in exploratory models, all the while recognizing their intrinsic limitations. While we must guard against mistaking any subjective "aha experiences" or fleeting feelings of familiarity for signs of the truth of our models, we should not downplay the importance of understanding to successful scientific practice. When researchers “refer to the results of their simulations by saying ‘we trust our results’ or ‘we trust our computer simulations’” they not only claim that the results are true (or approximately true), but also that they “understand why they are correct (or approximately correct)” (Durán 2018, p. 98). Recognizing exploratory fruitfulness as contributing to the credibility of a model does not entail that such fruitfulness can somehow compensate for other deficiencies (e.g., representational failure, or lack of realism); rather it amounts to yet another acknowledgment that the credibility of a model is the joint result of features of the model, aspects of the world, and the cognitive goals and interests of its users.

10.7 Summary

Simulation-based methods in science are deeply intertwined with the development of credible conceptual models, where the latter may be variously thought of as (pragmatically adequate) representations of real-world targets, as constructions of “credible worlds” that display internal and external coherence, or as exploratory tools that may aid our understanding even before a well-developed theoretical account of a phenomenon, or class of phenomena, becomes available. While it would be wrong to think that computer simulation methods are a “mere application” of the conceptual model to a particular problem situation, it would likewise be misleading to assume that the credibility of a simulation can somehow be divorced from that of the model. To be sure, verifying that a computer simulation performs as intended and ascertaining furthermore that its results are valid across a range of empirical contexts are important steps in establishing a simulation’s credibility. Yet, whether or not the results of a simulation are credible also hinges on the credibility, or lack thereof, of the underlying model. Yet such credibility, as this chapter has aimed to show, is not up to the model alone, but is jointly constituted by features of the model and the overall goals and cognitive interests of its users.

In much the same way that we demand of credible human interlocutors that they be competent and trustworthy—that is, able and willing to give reliable information across a range of relevant questions, without sudden unexpected failures—credible conceptual models should make reliable and relevant information available to those that depend on them. While this will often be a matter of how faithfully a model represents an actual target system, it would be hasty to think that this exhausts the concerns of scientists who depend on models and simulations. Sometimes, creating a model that constitutes a “credible world” in its own right—even if does not map on to the actual world—can help advance our understanding, and on yet other occasions, the most credible model may simply be the one that shows the most promise in generating fruitful new lines of research. Whether a conceptual model enjoys credibility, then, is as much a matter of its intrinsic structure and its relation to the world at large as it is a reflection of the goals and cognitive interests of its users.

References

Caro, J. J., Eddy, D. M., Kan, H., Kaltz, C., Patel, B., Eldessouki, R., & Briggs, A. H. (2014). Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value in Health*, *17*(2), 174–182.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.

Chaloupka, F. J., Tauras, J. A., & Grossman, M. (2000). The economics of addiction. In P. Jha & F. J. Chaloupka (Eds.), *Tobacco control in developing countries* (pp. 107–129). Oxford: Oxford University Press.

Chao, K.-K. (2014). Models and credibility. *Philosophy of the Social Sciences, 44*(5), 588–605.

De Regt, H., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, *144*(1), 137–170.

Durán, J. M. (2018). *Computer simulations in science and engineering: Concepts, practices, perspectives*. New York: Springer.

Gelfert, A. (2016). *How to do science with models: A philosophical primer*. Cham: Springer.

Goldfarb, R., Leonard, T., & Suranovic, S. (2001). Are rival theories of smoking underdetermined? *Journal of Economic Methodology*, *8*(2), 229–251.

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, *70*(1), 81–99.

Knutti, R. (2008). Should we believe model predictions of future climate change? *Philosophical Transactions of the Royal Society A*, 4647–4664.

Kuhn, T. S. (1977) Objectivity, value judgment, and theory choice. In *The essential tension*. Chicago: The University of Chicago Press.

Law, A. M., & Kelton, W. D. (1991/2000). *Simulation modeling and analysis*. New York: McGraw-Hill.

Lenhard, J. (2007). Computer simulation: The cooperation between experimenting and modeling. *Philosophy of Science*, *74*(2), 176–194.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4), 421–431.

Mäki, U. (2009). MISSing the world: Models are isolations and credible surrogate systems. *Erkenntnis, 70*(1), 29–43.

Massimi, M. (2018). Perspectival modeling. *Philosophy of Science, 85*(3), 335–359.

McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin, 6*(3), 467–472.

Merchant, C. (1980). *The death of nature: Women, ecology, and the scientific revolution*. San Francisco: Harper and Row.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. Morrison & M. S. Morgan (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 10–37). Cambridge: Cambridge University Press.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science, 263*(5147), 641–646.

Parker, W. (2009). Confirmation and adequacy-for-purpose in climate modelling. *Proceedings of the Aristotelian Society, 83*(1), 233–249.

Pitt, J. (2007). Speak to me. *Metascience, 16*(1), 51–59.

Reutlinger, A., Hangleiter, D., & Hartmann, S. (2017). Understanding (with) toy models. *British Journal for the Philosophy of Science,* 1–31. Advance article https://doi.org/10.1093/bjps/axx005.

Robinson, S. (2008). Conceptual modeling for simulation, Part I: Definition and requirements. *Journal of the Operational Research Society, 59*(3), 278–290.

Robinson, S. (2011). Choosing the right model: Conceptual modeling for simulation. In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, & M. Fu (Eds.), *Proceedings of the 2011 Winter Simulation Conference* (pp. 1428–1440). IEEE.

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling, 90*(3), 229–244.

Spinnewyn, F. (1981). Rational habit formation. *European Economic Review, 15*(1), 91–109.

Suárez, M. (1999). Theories, models, and representations. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 75–83). New York: Plenum Publishers.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology, 7*(1), 1–31.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, Mass.: Harvard University Press.

Winsberg, E. (2001). Simulations, models, and theories: Complex physical systems and their representations. *Philosophy of Science*, *68*(Proceedings), S442–S454.

Winsberg, E. (2018). *Philosophy and climate science*. Cambridge: Cambridge University Press.

Yuengert, A. M. (2006). Model selection and multiple research goals: The case of rational addiction. *Journal of Economic Methodology, 13*(1), 77–96.

Zacharias, G. L., MacMillan, J., & Van Hemel, S. B. (2008). *Behavioral modeling and simulation: From individuals to societies*. Washington, DC: The National Academies Press.