

ARTICLE

A Model-Based Approach to the Problem of Variable Choice

Collin Rice 

Department of Philosophy, Colorado State University, Fort Collins, Colorado
Email: corice@colostate.edu

(Received 15 February 2024; revised 17 September 2024; accepted 02 October 2024; first published online 11 November 2024)

Abstract

This article distinguishes causal modeling, metaphysical, epistemic, and modeling reasons for variable/parameter choice. I argue in favor of justifying variable/parameter choices by appealing to modeling reasons concerning the limitations of the available measurements, experimental data, modeling techniques, and modeling frameworks. I use this “tyranny of availability” to identify normative criteria for variable/parameter choice that apply across most scientific modeling contexts and investigate their metaphysical and epistemological implications.

1. Introduction

A common problem in scientific practice is selecting a set of variables and parameters to include in one’s scientific model.¹ As Jim Woodward (2016) has recently argued, the problem of variable choice is a *scientific* problem that goes beyond philosophers’ attempts to analyze which properties constitute natural kinds or which predicates are “gerrymandered.” To motivate this claim, Woodward quotes physicist Herbert Bernard Callan: “It should perhaps be noted that the choice of variables in terms of which a given problem is formulated, while a seemingly innocuous step, is often the most crucial step in the solution.” (1985, 465).²

The aim of this article is to evaluate various justifications that might be given for why certain variables or parameters are the “right” variables to include in a scientific model. I will argue that—contrary to Woodward’s (2016) and Robert Batterman’s (2021) recent proposals—rather than appealing to casual modeling, metaphysical, or epistemic reasons, scientific modelers often do, and should, appeal to what I call

¹ This also includes cases in which various variables/parameters are *constructed*—e.g., *via de novo* introduction, averaging, coarse-graining.

² It is worth noting that perhaps the problem of variable/parameter choice does not arise for every scientific model. However, even in nonmathematical models scientists need to decide which features to include/represent.

modeling reasons as the primary justification for their choice of a certain set of variables/parameters. These modeling reasons include considerations of which measurements, experimental data, theories, modeling frameworks, and mathematical modeling techniques are currently available with which to construct a viable scientific model. That is, these considerations focus on the constraints involved in trying to simply construct a workable model from the limited set of available modeling resources.

Assuming that we would like our philosophical views to be useful for practicing scientists, I contend that criteria for variable/parameter choice ought to balance the trade-off between (1) being able to offer specific guidelines for practicing scientific modelers, and (2) being as generally applicable as possible across scientific modeling contexts.³ As we will see, some accounts provide very specific criteria for variable choice but are rather limited in their scope of application. Others are more generally applicable, but at the cost of providing little detailed guidance for practicing scientific modelers. In contrast, I will argue that by focusing on the types of modeling constraints that arise from needing to construct a viable model from limited modeling resources we can distinguish generally applicable modeling reasons for variable choice from other more commonly recognized metaphysical, epistemic, or context-specific aims/goals. Indeed, to the extent that across most scientific modeling contexts, modelers need to select the “right” variable/parameters for constructing a viable model from the available resources, these modeling reasons will be applicable across a wide range of modeling contexts. The general applicability of these modeling reasons not only highlights their importance but also shows that more narrow accounts leave out several important considerations involved in variable/parameter choice.⁴

Focusing on these model-based justifications highlights what I call the “tyranny of availability” that greatly constrains variable and parameter choice across scientific practice. Rather than picking out a single best set of variables and parameters for casual modeling purposes (Woodward 2016) or that are of interest to one’s audience (Potochnik 2017), I will argue that—across a wide range of scientific modeling contexts—scientists’ choices of variables and parameters are highly constrained by the availability of the mathematical frameworks, modeling techniques, theories, measurements, data, and so forth with which to construct a scientific model. As a result, the variables or parameters that are favored for other reasons *typically must be drawn from within the set of variables and parameters delimited by these modeling constraints*. This gives modeling reasons a kind of priority over other reasons for variable/parameter choice and makes them the most widely applicable normative guides for scientific model building.⁵ While scientific modeling practices can/should be evaluated within specific modeling contexts, the general applicability of modeling

³ Thanks to Julia Bursten for suggesting this way of clarifying the balance between these theoretical aims.

⁴ Thanks to an anonymous reviewer for pressing me to be clearer about these desiderata.

⁵ I am here following a widely used methodology in philosophy of science that aims to provide suggestions about what scientists *ought* to do by drawing on examples of successful scientific practice. It is beyond the scope of this article to argue for this general methodology on behalf of the field.

reasons means they offer helpful guidelines for scientific modelers across much broader swatches of scientific practice.

A consequence of this view is that we ought to be cautious about drawing strong metaphysical or epistemological conclusions concerning the status of the variables/parameters that end up in our best scientific models. For example, instead of being the most “natural” variables with which to characterize a phenomenon, we can typically only say these are the most natural of the variables/parameters *from among the set of variables/parameters that might be used to construct a viable model given scientists’ limited modeling resources*. Similarly, instead of being the best variables/parameters with which to explain or understand a phenomenon, full stop, we ought to say that, *of the available variables/parameters with which scientists can construct a viable model*, the selected variables/parameters allow for the best explanation or understanding of the phenomenon. As a result, modeling reasons for variable/parameter choice have important metaphysical and epistemological implications.

The article will proceed as follows. In the following section, I distinguish causal modeling, metaphysical, epistemic, and modeling reasons that might be offered for the selection of a set of variables/parameters and argue that modeling reasons ought to be preferred and given priority in scientific practice. Section 3 then contrasts Batterman’s discussion of mesoscale modeling techniques in physics with an example of mesoscale modeling in biology to show that, while there will sometimes be metaphysical reasons for considering the variables within a scientific model to be natural kinds, in many other cases, modeling reasons can justify the choice of (the same types of) variables without warranting further metaphysical claims regarding their naturalness. Using these cases as a guide, in section 4 I lay out a model-based approach to variable/parameter choice and present specific criteria that ought to guide scientists’ choices of variables/parameters. Although other reasons might override them in certain cases, I argue that these modeling reasons yield generally applicable constraints/criteria that ought to guide variable/parameter choice across model-based science. Next, section 5 investigates what kind of priority modeling reasons ought to be afforded in scientific practice. Section 6 responds to some possible objections and the final section concludes.

2. Causal-Modeling, Metaphysical, Epistemic, and Modeling Reasons for Variable Choice

2.1 Woodward’s Causal Modeling Reasons

Woodward’s (2016) analysis of the problem of variable choice is explicitly instrumental in that it is tied to achieving specific aims of scientific inquiry within a particular context: “My view . . . is that the problem of variable choice should be approached within a means/ends framework: cognitive inquiries can have various goals or ends and one can justify or rationalize candidate criteria for variable choice by showing that they are effective means to these ends” (1051). More specifically, Woodward tells us that “The broad goal of inquiry on which I focus in what follows is causal representation/explanation” (ibid.). Thus, what is distinctive about Woodward’s approach is the specification of the specific aim of causal modeling/explanation and the selection of variables based solely on their ability to help scientists accomplish that aim (ibid., 1053).

By tying his approach to the specific aims of causal representation and explanation, Woodward is able to use his interventionist framework to offer fairly detailed criteria for variable choice:

- (1) Choose variables that are well-defined targets for interventions in the sense that they describe quantities or properties for which there is a clear answer about what would happen if they were to be manipulated or intervened on.
- (2) Choose variables that have unambiguous effects on other variables of interest when they are manipulated.
- (3) Choose variables that can (in principle) be manipulated independently of the values taken by other variables.
- (4) Choose variables that satisfy the causal Markov condition (or other standard screening off relations).
- (5) Choose variables that allow for the formulation of (as close as possible to) deterministic causal relationships.
- (6) Look for variables that figure in causal relationships that are *stable* across changes in background conditions.
- (7) Look for variables such that the resulting causal graph *accurately represents causal dependency relations* (ibid., 1054–5).

While Woodward's causal modeling criteria are acceptable as far as they go, I do not think they go far enough. Specifically, the issue is that, because there are numerous other scientific aims besides causal representation and explanation, the preceding criteria will fail to be normatively helpful in other modeling contexts. In short, the problem of variable choice arises in a wide range of cases in which the aim or purpose of the model is something other than to provide a causal explanation/representation. Woodward's criteria will be unhelpful to scientific modelers in those cases. Woodward is aware of this issue but is still hopeful that general guidance might be provided regarding the problem of variable choice (ibid., 1048). In what follows, I argue that model-based criteria for variable choice are more generally applicable beyond cases that focus on the aims of manipulation, control, or providing causal explanations.

Unfortunately, I don't think other authors referring to Woodward's approach as "pragmatic" is particularly helpful here since there can obviously be pragmatic considerations that are disconnected from this sort of causal modeling approach (and the word "pragmatic" is often used in ways that are so general that they will apply to almost any aspect of scientific practice).⁶ As a result, I will instead refer to this Woodwardian type of reason for variable choice as a "causal modeling" reason. It is this focus on pursuing causal modeling ends that limits the applicability of Woodward's approach rather than his appeal to pragmatic considerations.

I will argue that what distinguishes what I call "modeling reasons" from Woodwardian causal modeling reasons is that modeling reasons are motivated by the general scientific task of building a useable model from limited modeling resources (the tyranny of availability)—independently of whether the modeler aims to construct a causal model of the phenomenon or has some other modeling aim. That is, I will argue that modeling reasons for variable choice are applicable across most

⁶ I address the question of whether modeling reasons are pragmatic reasons more in section 6.

(if not all) modeling contexts—regardless of the more context-specific purposes (or ends) scientists have for their models.

2.2 *Metaphysical Reasons*

In contrast with Woodward's causal modeling approach, other philosophers have argued that certain kinds (and the variables that represent them) are "natural kinds" or "natural properties." Thus, one might argue that the variables/parameters that ought to be included in scientific models are those that represent these natural kinds—what Batterman calls "natural kind variables" (2021, 126). One attempt to do this has been to try to identify the properties that will figure in some final or most fundamental physics (Lewis 1983; Sider 2011). However, as both Woodward (2016, 1056) and Batterman (2021, 45) note, this analytic metaphysics approach isn't particularly helpful for practicing scientific modelers given that it depends on metaphysicians' intuitive judgments about what is natural and we simply do not have access to any such final physics. Moreover, such an approach seems to require a rather strong form of reductionism to evaluate the naturalness of the properties used in biology or other "special sciences." At the very least, it would be useful to have normative criteria to offer scientific modelers that do not depend solely on metaphysical intuitions of philosophers, having access to a final physics, or reducing the properties of all models to properties in fundamental physics.

Alternatively, one might attempt to provide scientifically-motivated *a posteriori* reasons for considering certain variables to be more "natural." Indeed, Batterman attempts to do precisely this when he contends that he is providing "nonpragmatic" reasons for considering certain variables to be natural *simpliciter* (ibid., 124). Rather than appealing to considerations of fundamentality, Batterman proposes that, "The ultimate conclusion will be that the middle-out engineering approach to many-body systems is often *ontologically superior* to one based in fundamental theory" (ibid., 121; my emphasis). Moreover, he tells us that he is, "addressing metaphysical concerns about the proper way to carve nature at its joints" (ibid.). Despite his ontological focus, however, Batterman is rather explicit that "we *do not* need to engage in metaphysical analysis" (ibid.) and that "sometimes we have non-metaphysical and non-methodological reasons with which to identify or delimit a privileged class of natural-kind variables" (ibid., 126). I interpret Batterman as rejecting various attempts by metaphysicians (and philosophers of physics) to establish which properties are natural using *a priori* reasoning concerning fundamentality. Yet, he does intend to provide *a posteriori* reasons for drawing conclusions concerning the naturalness of mesoscale variables.

Most of the *a posteriori* reasons offered by Batterman appeal to the ontological implications of the Fluctuation-Dissipation Theorem (FDT), which I discuss in more detail in the next section. After explaining the role of FDT in justifying mesoscale modeling techniques, Batterman tells us that, "*The Fluctuation-Dissipation theorem for many-body systems is really a profound result with ontological consequences*. It tells us that we need to take mesoscale structures and the parameters . . . that code for those structures via correlation functions to be genuine features of the world" (ibid., 110). In other words, Batterman argues that we have scientifically-motivated reasons "for treating the mesoscale parameters as, in a rather strong sense, among those that

should be considered natural kinds” (ibid., 25). As a result of their naturalness, Batterman contends that, “They are the right variables *with which to model and investigate* various aspects of the bulk behavior of many-body systems” (ibid. 127; my emphasis). That is, the FDT has important ontological consequences that, in turn, justify the selection of mesoscale parameters as the best with which to model the target phenomenon.

However, the same problems arise here as with Woodward’s approach. While the FDT does allow for interesting and important inferences concerning which variables ought to be prioritized in various cases within many-body physics, it is unclear how such normative guidance would help with variable choice in other contexts. That is, while Batterman’s *a posteriori* reasons concerning naturalness are certainly unique and interesting, their dependence on the discovery of principles like the FDT limits their applicability in other contexts. For example, in many parts of science there are no physical or theoretical reasons to suppose that the variables included in a scientific model are the most natural, or even that they represent properties that exist. Moreover, the same mesoscale modeling techniques discussed by Batterman—for example, homogenization—have been applied in modeling contexts outside of physics in which there is little hope of establishing the naturalness of the mesoscale variables/parameters included in the model. I discuss such a case in the next section. As a result, while Batterman’s metaphysical approach to variable choice is independent of pursuing causal modeling aims of science, it fails to yield generally applicable normative guidance for scientific modelers.

2.3. Epistemic Reasons

While Woodward’s account is tied to the specific aim of causal modeling, his focus on the aim of providing *explanations* is perhaps applicable across a much wider range of modeling contexts (though it is unclear how well Woodward’s interventionist-based criteria would apply to instances of noncausal modeling/explanation). This suggests that there might be unique reasons/justifications for variable/parameter choice that are involved in aiming to accomplish *more general epistemic aims of science* such as knowledge, explanation, or understanding. Batterman (2021), too, appeals to various kinds of epistemic reasons: “I will argue that mesoscale quantities and parameters are, *for the purpose of understanding* the bulk behavior of many-body systems, much superior to quantities and parameters at fundamental atomic scales” (2; my emphasis). And later on, “This chapter aims to justify these variables/parameters as *better able to figure in explanations, as better able to provide descriptions and understanding of certain behaviors*” (ibid., 121; my emphasis). Similarly, Angela Potochnik (2017) has argued that the variables that ought to be included in a scientific model are those that contribute to the causal pattern of interest to one’s audience because this will best promote understanding. Rather than tying variable choice to causal modeling contexts or appealing to metaphysical naturalness, this epistemic approach aims to justify scientists’ variable choices by showing how they contribute to more general epistemic aims of scientific inquiry.

Turning this epistemic approach into more detailed normative criteria for variable/parameter choice, of course, requires offering an account of these epistemic aims and how scientific models play a role in accomplishing them—or at least some

story about why certain kinds of variables/parameters will best aid in accomplishing these epistemic aims. But this highlights a problem for the epistemological approach. Given the myriad of different accounts of scientific knowledge, explanation, and understanding, epistemic considerations will likely result in a plethora of different criteria for variable choice. This is true both within and across different epistemic aims.

For example, like Woodward, many philosophers have argued that the aim of explanation will be best achieved by models that include variables and parameters that capture the difference-making causes of the system (Strevens 2008). However, other philosophers have argued that the variables that ought to be included in causal explanations should be tailored to the intended audience rather than overly focused on difference-making causes (Potochnik 2017). Moreover, many other philosophers have argued that there are noncausal scientific explanations that would not be captured by these causal accounts and have offered varying accounts of how noncausal explanations work (Batterman and Rice 2014; Bokulich 2011; Rice 2021; Khalifa 2017). Thus, even if we focus exclusively on the aim of explanation, adopting even a modest pluralism will result in multiple conflicting normative criteria for variable/parameter choice.

This problem is only exacerbated by considering additional epistemic aims like understanding or knowledge. Indeed, several philosophers have argued that knowledge, understanding, and explanation are importantly different (Elgin 2017; Kvanvig 2003; Rice 2021). Moreover, the philosophical literature on these other epistemic aims—for example, understanding—has generated a plethora of accounts of how scientific models contribute to them as well (Elgin 2017; Khalifa 2017; Potochnik 2017; Rice 2021).

Consequently, the combination of there being multiple epistemic aims of science and a multitude of conflicting accounts of each of those aims makes it unclear how this epistemic approach would generate widely applicable and consistent rules for variable/parameter choice. That is, given the diversity of epistemic aims and the different ways they are specified *the epistemic approach's increase in generality results in less specificity (or consistency) in terms of the best means by which those epistemic ends are accomplished*. What is more, given that these epistemic aims are not the only aims of scientific inquiry, these criteria are, once again, only normatively useful in so far as these epistemic aims are the goal of scientists' modeling projects. Therefore, while epistemic aims are more general, it is unclear how they can provide generally applicable, specific, and consistent normative criteria for variable choice.

At this point, one might suggest that there are common types of modeling considerations that promote each of these epistemic aims, for example, identifying dependency relations or empirically supported truths (Khalifa 2017). I reply that, even if there are some general commonalities between these different epistemic aims, those commonalities won't be robust enough to yield specific criteria for variable/parameter choice that would best accomplish the various epistemic aims of science. For example, telling scientific modelers to select variables/parameters that are involved in dependency relations is of limited use given the plethora of dependency relations that one might include in a scientific model. Consequently, the normative guidance extracted from what all epistemic aims of science have in common (or what all accounts agree on) is likely to be rather thin.

Another response might simply grant that this diversity of epistemic aims and different means for achieving them is precisely what motivates only evaluating models as being adequate for specific purposes within specific contexts. That is, one might suggest that this diversity of aims is precisely why very few general principles regarding variable/parameter choice are likely to be found.⁷ However, in the rest of the article, I identify numerous more widely applicable criteria for variable/parameter choice that can be applied across (almost) any context in which scientists need to construct a model—regardless of their more specific aims for the model. Specifically, I suggest that—while investigating particular modeling contexts is certainly important—more general criteria (or constraints) for variable/parameter choice can be also identified by investigating the constraints imposed by the tyranny of availability.

2.4 Modeling Reasons

In light of the limitations of the above approaches, I contend that a new approach is needed. Fortunately, Batterman (2021) suggests an alternative by appealing to various kinds of *modeling* considerations to motivate the use of mesoscale variables/parameters to model the bulk behaviors of many-body systems. I think these reasons are importantly different from the more metaphysical reasons Batterman offers by appealing to the FDT. Rather than being ontologically guaranteed to exist, mesoscale parameters are often the “right” parameters because they *allow for successful modeling of the target phenomenon given the available modeling resources*. For example, immediately after telling us that the middle-out approach is ontologically superior, Batterman clarifies that: “By ontological superiority, I mean that these quantities or kinds *allow for much more effective modeling* of the mesoscale regularities exhibited by many-body systems” (2021, 121; my emphasis). In addition, Batterman often characterizes the “naturalness” of these variables in similar terms: “I argue that mesoscale parameters—order parameters and material parameters—are *natural* variables in the sense that they are *the best variables with which to characterize* certain dominant, lawful behaviors of many-body systems” (ibid.; my emphasis).

Following these suggestions from Batterman, I propose that there can be straightforward and generally applicable *modeling reasons* for preferring a set of variables/parameters that are independent of causal modeling, metaphysical, or epistemological considerations. These modeling reasons include using certain (sometimes artificially constructed) parameters because they are necessary to employ the available modeling techniques that have been successfully applied to similar kinds of phenomena in other contexts. For example, modelers in ecology, economics, and the climate science have attempted to model systems in terms of particular kinds of variables and parameters as a means to employing the modeling frameworks used to study phase transitions in physics (Jhun et al. 2017; Rice 2021). Indeed, in many cases, scientific modelers will select certain variables/parameters that are analogous to those used in other similar modeling contexts to enable them to use already developed models or modeling frameworks.

⁷ Thanks to an anonymous reviewer for raising this objection.

Another modeling reason is using certain variables/parameters because they can be compared to the available measurements or experimental data sets from which a model might be constructed and tested. As Batterman notes, one reason physicists focus on mesoscale features is because, “these quantities provide a rather direct connection with measurements we can actually perform on many-body systems” (Batterman 2021, 66). Multiscale modelers in biology agree: “The starting point for a ‘middle-out’ approach to modeling biological systems may be influenced by a number of factors, including the ready availability of relevant experimental data” (Walker and Southgate 2009, 451). Indeed, scientific modelers often select certain variables and parameters because they can be directly compared with the available experimental data and measurements.

A third common type of modeling reason is when scientific modelers use coarse-grained or homogenized variables/parameters because they greatly reduce calculation times or allow for models/simulations to yield more precise solutions. For example, Eric Winsberg notes that in climate modeling, parameterization involves “replacing missing processes—ones that are too small-scale or complex to be physically represented in the discretized model—by a more simple mathematical description” (2018, 48). Moreover, these parameters are not physical since “no such value exists in nature,” but they are instead “artifacts of the computation scheme” (ibid., 49). In this case, scientific modelers make use of certain parameters because they make it possible to build models/simulations that can be solved or run in significantly less time, not because those parameters are thought to capture real, natural, or causal properties of the target system(s).

What distinguishes these kinds of modeling reasons is that they need not be tied to more commonly recognized causal modeling, metaphysical, or epistemic aims of science such as manipulation, prediction, understanding, or explanation. Instead, they are *motivated by the very task of constructing a viable/workable model from limited modeling resources*—regardless of scientists’ further goals for that particular model. As a result, I contend that these modeling reasons are importantly distinct from (Woodwardian-style) causal modeling, metaphysical, or epistemic reasons for variable choice discussed in the preceding text. While models are always built with some further purpose in mind (Giere 2004; Weisberg 2017) and can, and should, be evaluated according to their adequacy for those purposes (Parker 2020), the types of justifications offered for various modeling decisions—for example, variable or parameter choices—should also (when possible) be evaluated *across a wide-range of scientific modeling contexts* independently of more specific aims/purposes for the model. That is, we can *both* evaluate models using means-end reasoning within specific contexts *and* identify more generally applicable criteria/reasons for variable/parameter choice that apply across a wider range of contexts.⁸ Of course, in practice, these generally applicable model-based reasons will interact with more context-specific modeling purposes/aims in various ways. The ways these reasons interact and constrain one another is an important and interesting avenue for future philosophical research. Nonetheless, their being tied to the common challenge of

⁸ The view I defend here is perfectly compatible with also evaluating models based on their adequacy for purpose, but the details of how these views interact is a subject for future research (see Parker 2020, 467).

constructing a viable model from limited modeling resources highlights what is distinctive about modeling reasons for variable/parameter choice, shows that more narrow accounts miss important considerations that influence variable/parameter choice, and gives us strong reasons to believe modeling reasons will be generally applicable across a wide range of scientific modeling contexts.

It is by focusing in on this most generally applicable model-building context that we find scientists consistently appealing to model-based reasons to justify their selection of certain variables/parameters to include within their models. Because they are built into the general scientific task of constructing a scientific model from limited modeling resources, I contend that these model-based reasons are the most generally applicable reasons for variable or parameter choice—though, in practice, their application will almost always be combined with other kinds of reasons for variable/parameter choice. That is, regardless of a scientist’s more specific aims for their model, or the type of model they aim to construct, (almost) all scientific modelers must confront the tyranny of availability due to their being limited modeling techniques, frameworks, measurements, and so forth from which to construct their models.

3. A Tale of Two Sciences: Mesoscale Modeling in Physics and Biology

To draw out what is distinctive about these kinds of modeling reasons and use them to extract more specific normative criteria for variable/parameter choice, it will be useful to compare Batterman’s analysis of mesoscale modeling in physics where ontological reasons drive the selection of particular variables/parameters with a case of mesoscale modeling in biology where scientists appeal to model-based reasons for choosing mesoscale variables/parameters.⁹

3.1 Mesoscale Modeling in Physics

Batterman (2021) persuasively argues that there can be ontological reasons for variable choice in cases where mesoscale variables can be thought of as the “natural variables” with which to model and characterize a many-body system. As I noted in the preceding text, Batterman’s arguments rely heavily on the FDT. Many-body systems are often in flux between states of equilibrium and nonequilibrium. The FDT “states an equivalence between the response of such a system to a small external disturbance (a push of some kind) and internal fluctuations in the absence of such a disturbance” (ibid., 21). In other words, the FDT tells us that the evolution of the system back to its equilibrium will be the same regardless of whether the disturbance is internal or external. More specifically, systems that are perturbed out of their equilibrium state by an internal perturbation will evolve back to equilibrium by having the correlations between different spatial and temporal regions decay over time. The FDT “asserts that the response of the system to an external push will decay in the same way” (ibid., 22). Moreover, because the system’s evolution back to its

⁹ These case studies are intended to highlight what is *distinctive* about modeling reasons. They are not meant to establish a “general applicability” claim. The argument for these reasons’ general applicability is grounded in their being built into the general task of constructing a model from limited resources across multiple modeling contexts.

equilibrium state is characterized in terms of the decay of these correlations between different regions of the systems (that are unobservable from the perspective of the individual particles), *the FDT guarantees that there will be mesoscale correlational structures that are essential to understanding these changes in the system* (ibid., 136). This result, in turn, helps explain why hydrodynamic methods that focus on finding parameters that directly code for those mesoscale correlational structures are so successful.

For these reasons, Batterman argues that there are physical/ontological reasons, grounded in the FDT, for why mesoscale parameters are the “right” or “natural” variables with which to model these many-body systems (ibid., ch. 7). That is, the FDT guarantees that these mesoscale correlational structures will exist and be essential to understanding the bulk behaviors of the system. This, Batterman argues, means we should treat them as “natural variables” and that “[t]hey are the right variables with which to model and investigate various aspects of the bulk behavior of many-body systems” (ibid., 127).

In contrast with Batterman’s proposal that, “there are theoretical, scientific reasons for treating the mesoscale parameters as, in a rather strong sense, among those that *should be considered natural kinds*” (ibid., 25), I argue that, in other cases, there can be compelling modeling reasons for treating mesoscale variables as the right variables with which to model a phenomenon, but that these modeling reasons are often insufficient to warrant this stronger claim concerning natural kinds. That is, I argue that there can be model-based reasons for selecting mesoscale variables/parameters that are independent of the kind of ontological reasons/backing discussed by Batterman. To see how this is possible, we can look to an example of mesoscale modeling in biology.

3.2 Mesoscale Modeling in Biology without Appealing to Natural Variables

The same mesoscale modeling techniques discussed by Batterman—for example, homogenization—have been successfully applied in biological contexts as well. However, here, I will argue, the primary justification for focusing on particular mesoscale parameters/variables appeals to modeling reasons rather than ontological justifications.

As an example, Martha J. Garlick et al. (2011) apply homogenization techniques to model the spread of chronic wasting disease (CWD) in mule deer. Diffusion models, developed in physics to model the spread of gases, have been used to model various biological problems such as the spread of genes/organisms. The issue is that these approaches assume that diffusion takes place across homogeneous landscapes and conditions. However, “most animals do not diffuse like particles” (ibid., 2089) and their spread/movement is different depending on the landscapes in which they exist. This is because, “[Organisms] are greatly influenced by habitat type, moving slowly through landscapes that provide needed resources and more quickly through inhospitable regions and are therefore much more likely to be found some places than others” (ibid.). Simply averaging over the entire system would ignore these important heterogeneities at mesoscales. However, modeling all the smaller scale details of the system would be “daunting to implement in a model, particularly at large spatial scales” (ibid., 2090).

In response, Garlick et al. adopt a homogenization approach that aims to capture the relevant mesoscale differences in terms of types of landscapes by incorporating important small-scale variations into the averages, while keeping the model computationally tractable by ignoring most of the (irrelevant) small-scale variations in the system (ibid., 2090–91). In addition to these computational savings, these modelers justify their use of certain idealizing assumptions that eliminate various variables/parameters from the model because doing so enables them to apply homogenization modeling techniques that have been successfully applied in other similar contexts.¹⁰ As they put it, their model, “will ignore other modeling considerations such as seasonal and sex differences in movement and age structure of the disease for purposes of developing the homogenization approach” (ibid., 2092; my emphasis).¹¹ Moreover, their decision to include certain variables and parameters was justified by claiming that it “allows us to apply a homogenization procedure similar to the one derived for [other similar contexts]” (ibid., 2093). More generally, the desire to employ an existing and previously successful modeling technique is presented as justification for including certain mesoscale variables and parameters and ignoring or idealizing other factors.

The next question is at which mesoscale of the system ought these variables and parameters be described? For these modelers, this modeling decision is dictated by the available measurements and experimental data. In particular, they carved the landscape up into 30×30 kilometer blocks because this is the scale at which the US Geological Survey Landcover Institute had collected their measurements of mule deer movements through different landscapes (ibid.). Carving the landscape up this way and homogenizing over these mesoscale regions resulted in mesoscale ecological diffusion equations for the different parts of the landscape that made use of different motility coefficients designed to capture/incorporate various contributions of the terrain and resources to animal movement (ibid., 2102).

After constructing this homogenization model for the spread of CWD, these modelers argue for the utility of using this modeling approach by pointing out that the nonhomogenized model took 45 hours to run a one-year simulation whereas the homogenized model took just 3.85 seconds! In other words, the decision to include only certain mesoscale variables is further justified by showing that using the homogenization technique provides similar results in $1/42000$ of the time. This is, without a doubt, “a substantial computational savings” (ibid., 2103).

In summary, these biological modelers do not appeal to ontological facts, principles, or theorems that would lead us to expect their mesoscale modeling approach to capture certain “natural” variables of these biological systems. Instead, they offer several different modeling reasons tied to the ability to construct, apply, test, and run their idealized model. Specifically, they offer the following reasons for selecting particular variables/parameters:

¹⁰ For example, the model assumes that death from causes other than CWD are balanced by births and ignores indirect ways that the disease might be spread.

¹¹ Of course, this is a form of means-end reasoning. However, this is motivated by the limited modeling resources available and is generally applicable regardless of one’s further aims/purposes for the model.

- (1) The (existing) homogenization modeling techniques that focus on these mesoscale variables/parameters have been successfully applied to similar kinds of problems in physics and other areas of biology.
- (2) Other factors are ignored for the purposes of implementing the homogenization approach.
- (3) The available data/measurements concern variables/parameters at a particular mesoscale.
- (4) Computational savings.

None of the preceding justifications require us to show that there is some kind of physical theorem or metaphysical reason why this modeling approach is applicable, universal, or likely to be successful. That is, we need not argue that these parameters/variables are more natural to argue that they are, nonetheless, the “right” parameters/variables with which to characterize/model the phenomenon. Moreover, these variables/parameters need not be targets for intervention or manipulation of the kind focused on in causal modeling. Finally, these modelers do not appeal to epistemic reasons concerning improved understanding or explanation, but rather appeal to the available modeling approaches, measurements, and computational resources as justification for their variable/parameter choices. Therefore, in this case, I argue that modeling reasons alone provide sufficient reasons for these modelers’ variable/parameter choices without appealing to (Woodwardian-style) causal modeling, metaphysical, or epistemic reasons.

Of course, just because scientists appeal to these considerations does not mean they *ought* to. However, the fact that this is a *successful* instances of scientific modeling in which the modelers are able to construct a viable/usable scientific model from the existing resources that is able to accomplish their aims does provide reasons to think they are justified. By choosing variables to use the available modeling techniques (that were previously applied to similar types of phenomena), drawing on the available data/measurements, and limiting the necessary computational resources, these modelers were able to build a successful scientific model for accomplishing their purposes from the limited modeling resources available. That is, using these modeling reasons for variable choice were crucial to the success of their modeling project.

4. Adopting a Model-Based Approach to Variable/Parameter Choice

I now draw on the preceding example to identify several model-based justifications that might be given for selecting particular variables/parameters. I then use these model-based reasons to construct a list of criteria that ought to normatively guide variable/parameter choice across scientific practice.

4.1 Certain Variables/Parameters Are Necessary for Using the Available Modeling Techniques

A primary consideration in adopting a model-based approach to variable/parameter choice is that, in many cases, the inclusion of certain types of parameters, conserved quantities, and limiting cases is *necessary for employing the currently available mathematical modeling techniques*. In the preceding case, we saw that the inclusion of certain (idealized) features and the exclusion of various complicating factors was

done “for purposes of developing the homogenization approach” (ibid., 2092). That is, these modelers wanted to use an existing modeling technique and included/excluded various variables/parameters as a means to being able to employ that modeling approach. Other examples include the inclusion of certain variables/parameters that are “self-similar” to apply certain multiscale modeling techniques (Batterman 2021; Rice 2021) and the inclusion of variables for population averages and variation that are necessary to apply certain statistical modeling frameworks (Ariew et al. 2015).

The flipside of these reasons for including certain variables/parameters is that wanting to adopt a particular modeling framework is also often used to justify *ignoring* certain variables/parameters. For example, R. A. Fisher (1930) motivated his inclusion of random mating and a sufficiently large number of genes in his biological models so that he could apply statistical modeling techniques. This modeling choice, in turn, motivated Fisher to build biological models that ignored smaller scale variations:

Finally, and perhaps most importantly, he assumed that the factors were sufficiently numerous so that some small quantities could be neglected; in other words, large numbers of genes were treated in a way similar to large numbers of molecules and atoms in statistical mechanics. As a result, Fisher was able to calculate statistical averages that applied to populations of genes in a way analogous to calculating the behaviour of molecules that constitute a gas. (Morrison 2004, 1197)

These cases illustrate that, often, scientific modelers motivate/justify their choices regarding which variables/parameters to include (and whether to randomize them or introduce various limits) by their desire to use the available modeling frameworks and the necessity of those variables/parameters for employing those modeling techniques. More generally, scientific modelers are often constrained by the limited range of representational/modeling frameworks that have been developed for modeling a certain type of phenomenon. While scientific modelers certainly can, and sometimes do, construct new types of models, in most cases scientific modelers prefer to choose from among the already existing types of models and make variable/parameter choices motivated by the desire to use existing modeling techniques.

4.2 If It Ain't Broke, Don't Fix It

A related, but importantly different, justification for using certain variables/parameters within a model is the desire to use modeling techniques *that have been successfully applied to other similar problems in the past*. In the preceding case, the modelers justify their use of particular variable/parameters to use homogenization techniques because such a modeling approach had been successfully applied to similar kinds of problems by other ecologists and physicists. Moreover, they incorporated the variables/parameters, such as coefficients for “ecological diffusion” that were developed by those other modelers. Christopher Pincock (2012) describes a similar case in which the use of a variable that represents an organism’s “domain of danger” is central to Hamilton’s selfish-herd modeling research program. Interestingly, Pincock argues that this proxy variable is an idealization because there is evidence that predation risk is not accurately captured by an organisms’

domain of danger. But biologists continue to use the variable because it enables them to develop models of predation risk and apply them in similar contexts. As Pincock puts it, “[T]he central anchor for Hamilton’s program is the claim that predation risk is accurately reflected in the relative size of a prey organism’s domain of danger . . . this anchor is an idealization. *It is used to formulate a model despite the fact that the scientists using it believe it to be false*” (ibid., 492; my emphasis). In these cases, the general motivation is to include variables/parameters that enable for the application of a previously successful modeling technique and to use the parameters that have been specially developed for constructing models of a particular type of phenomenon. As a result, it isn’t just the employment of an existing modeling strategy but also the use of specific ways of deploying those strategies to solve particular types of problems that motivates the inclusion of certain types of variables/parameters within a scientific model.

4.3 Confronting the “Tyranny of Scales”

Another modeling consideration that motivates/justifies the inclusion of certain variables/parameters within a model are the challenges raised by the so-called tyranny of scales (Oden 2006; Batterman 2013; Bokulich 2021; Rice 2021). Essentially the problem is that most of the available scientific models have been developed to represent process at particular characteristic scales (or a narrow range of scales). However, when modeling multiscale phenomena, often features and processes at a wide range of spatial and temporal scales need to be modeled and put in communication with one another. As J. T. Oden explains:

Virtually all simulation methods known at the beginning of the twenty-first century were valid only for limited ranges of spatial and temporal scales. Those conventional methods, however, cannot cope with physical phenomena operating across large ranges of scale At those ranges, the power of the tyranny of scales renders useless virtually all conventional methods. (2006, 29–30)

This situation creates the modeling challenge of needing to select models and variables that *can be more easily integrated with other types of models used to describe features and processes at other scales*. As a result, in most multiscale modeling situations, “Multiscale modeling, besides modeling the system, needs to address the issue of how to bridge the gaps between different methodologies and between models at different scales” (Castiglione et al. 2014, 7). As a specific example, several biological modelers’ motivation for building middle-out mesoscale cellular automata is that these models “can be integrated with other modeling modalities (e.g. partial or ordinary differential equations) to model multi-scale phenomena” (Walker and Southgate 2009, 450). Batterman provides a similar example in which modeling multiscale phenomena in biological contexts, “requires linking together different types of modeling at various levels.” (Bassingthwaight et al. 2009, 597). In short, often scientific modelers choose certain variables/parameters because they enable them to construct models at particular scales that can be more easily integrated with different types of models used to represent features at other scales.

4.4 Comparison with the Available Measurements or Experimental Data

Another kind of justification seen in the preceding case is the inclusion of certain variables/parameters because they allow for direct comparisons with the available experimental data/measurements. Indeed, as Batterman argues at numerous places in *A Middle Way*, “One of the most important aspects of the hydrodynamic description in terms of correlation functions is its rather direct connection with experiment.” (2021, 15). Numerous biological modelers agree: “The starting point for a ‘middle-out’ approach to modeling biological systems may be influenced by a number of factors, including the ready availability of relevant experimental data” (Walker and Southgate 2009, 451; my emphasis). I take these justifications to be appeals to generally applicable modeling reasons because they are motivated by the need to have measurements and data with which to build, test, and verify scientific models. And often, despite certain variables/parameters being desirable for various causal modeling or epistemic purposes, if they are not measurable, then they are often not the “right” variables. For example, in many cases, although directly representing fitness would often be the most explanatory/predictive, various proxies for fitness (e.g., amount of food consumed or number of eggs fertilized) are used in biological models instead because they are directly measurable.

4.5 Computational Savings

Finally, as we saw in the spatial ecology case mentioned previously, often the choice of a particular set of variables/parameters is justified by showing that the approach yields accurate (or similar) results while using fewer computational resources. Rather than suggesting that this makes the explanations provided by the model better or better tracks the true ontology of the system, scientific modelers often argue that certain variables/parameters (or modeling techniques) ought to be employed because they can provide similar calculations/results in less time or with fewer computational resources. While certainly pragmatic (given that time is a resource), computational savings are not tied exclusively to causal modeling contexts as are Woodwardian casual modeling reasons. Instead, computational savings can motivate variable/parameter choice regardless of the modeler’s more specific aims.

4.6 Model-Based Criteria for Variable/Parameter Choice

Using these model-based considerations, let me now lay out a (clearly nonexhaustive) list of some of the criteria for variable/parameter choice that might be motivated/justified by what I have been calling modeling reasons:

- (1) Choose variables/parameters that are necessary to use the available mathematical models or modeling techniques.
- (2) Choose variables/parameters that enable the use of modeling techniques that have been successfully applied to similar types of problems/phenomena in other contexts.
- (3) Choose variables/parameters that will best allow the model to “communicate” or “pass information” to different types of models used to represent other scales (or features) of the system.

- (4) Choose variables/parameters such that at least some of the features of the model can be constructed from, or compared against, the available measurements or data.
- (5) Choose variables/parameters that utilize the least computational resources/time.

Like Woodward's criteria, these model-based criteria are intended to be "general normative guides" or "rules of thumb." That is, they provide general guidelines for how scientists ought to make variable/parameter choices rather than being exceptionless rules that must be adhered to in every case. Indeed, there will be modeling contexts in which each of these normative criteria ought to be violated—for example, when the best option is for scientists to construct a new kind of model for a novel type of phenomenon. However, this does not mean that these modeling reasons fail to be broadly applicable across scientific modeling contexts. Something can be a good reason in most or all instances even if other reasons override it in particular contexts/cases. Relatedly, it is crucial to remember that these model-based criteria ought to be weighed against each other and applied collectively rather than in isolation. When considered collectively, these criteria can provide clear guidance for practicing scientific modelers across a wide range of modeling contexts.

5. The Priority of Modeling Reasons

A primary reason for emphasizing these modeling reasons for variable/parameter choice is that they are widely applicable because *they are built into the very question of which variables/parameters to include in a model*. Indeed, the tyranny of availability gives rise to constraints on variable/parameter choice whenever a scientist aims to construct a model of the phenomenon from limited modeling resources. I contend that this gives modeling reasons a kind of *priority* as general constraints on scientific model building. In addition to their widespread applicability, my claim that modeling reasons have a kind of priority in scientific practice is also motivated by noting that model-based reasons typically *constrain* other justifications/motivations for variable/parameter choice. Rather than a claim about temporal sequence (in practice, these different types of reasons will often interact and intertwine in complex ways), my claim is that scientific modelers often must identify a set of variables/parameters that will enable them to construct a workable model of the phenomenon and will then choose *from among that constrained set* the variables/parameters that will best serve their more specific modeling purposes.

This priority of model-based reasons entails that scientific modelers' frequent appeals to model-based justifications for variable/parameter choice have important epistemological and metaphysical consequences. Specifically, the selection of a set of variables/parameters to include in a scientific model ought not be interpreted as a purely metaphysical or epistemological claim about these variables/parameters being the closest carving of nature at its joints, or the best for explaining or understanding the phenomenon, full stop. Instead, these claims about the metaphysically or epistemically (or otherwise) preferred variables/parameters *need to be relativized to the available variables/parameters with which scientists can construct a viable model*. This means we should instead say things like "these are the variables/parameters that

allow for the best explanation of the phenomenon from among those variables/parameters with which scientists can construct a workable model.” Similarly, we can typically only say that “these are the most natural variables/parameters from among the set of variables/parameters that might be used to construct a viable scientific model given the existing modeling resources.”

More generally, my model-based approach to variable/parameter choice focused on the availability of modeling resources shows that, although investigating how models can be adequate for specific purposes has revealed numerous important insights, we can also identify more general normative guides for scientists by looking at features of scientific practice that arise across most (or perhaps all) modeling contexts. This can help provide more generally applicable philosophical accounts of scientific modeling that then ought to be combined with more context-specific considerations. In sum, philosophers of science analyzing scientific modeling practices need not be completely particularist (or case-specific) in their conclusions; nor do they need to make pronouncements about exceptionless universal principles/rules for all of scientific practice. Instead, philosophers of science can, and should, investigate specific modeling contexts, while also looking for more general lessons that can be applied across larger swatches of scientific practice. I contend that the modeling reasons identified above are generally applicable normative guides for scientific modelers when deciding which variables/parameters to include in a scientific model. That is, of the kinds of reasons considered here, modeling reasons best navigate the trade-off of being widely applicable, while being able offer specific guidelines for variable/parameter choice.

6. Objections and Replies

There are several possible objections to the preceding arguments for a model-based approach to variable choice. In this final section, I try to address some of them.

First, one might object to my suggestion that these reasons for variable/parameter choice are really distinct. Indeed, in scientific practice, casual modeling, metaphysical, epistemic, and modeling reasons often seem intertwined rather than being distinct options. While I think this *in practice* claim is certainly right, I offer three brief replies. One response is to simply acknowledge that while distinguishing these kinds of reasons is an idealized representation of what happens in actual model construction, it is nonetheless a useful philosophical exercise to distinguish them so they can be considered and analyzed independently. In particular, distinguishing these reasons for variable/parameter choice helps us better analyze the justification/warrant they provide for certain modeling decisions—even if they are always applied in combination within scientific practice. Another response is to note that although more than one type of reason will often be applied across the same model construction process, those reasons need not be given equal weight. Instead, as I suggested, I think modeling reasons ought to be given priority because they apply across all model-building contexts and are typically constraints on other types of considerations given for variable choice. A final response appeals to the lessons drawn from the cases discussed in the preceding text. Specifically, I think it is crucial to differentiate modeling reasons from metaphysical reasons to avoid conflating the presence of certain variables/parameters in our scientific models with their being

natural kinds (or being plausible targets for intervention). While I think Batterman has convincingly shown that we can sometimes have metaphysical reasons that justify why certain variables/parameters ought to appear in scientific models, we ought to be careful about overapplying any kind of “reading our metaphysics off our best scientific models.”

A second objection might suggest that a generalized version of Woodward’s account that focused on “pragmatic” reasons generally would capture these modeling reasons. First, my argument is that these modeling reasons are importantly different from Woodward’s causal modeling reasons; I am not arguing that modeling reasons are never pragmatically motivated or are completely distinct from more context-specific reasons. My claim is that modeling reasons for variable choice are not unique to specific causal modeling contexts or aims and that this makes them importantly different from the kinds of reasons appealed to by Woodward.

Of course, one might suggest that Woodward’s account is too narrow and that what we might call “pragmatic reasons” are just anything that pays attention to utility, context, or resource constraints. If we adopt this extremely generic sense of pragmatic, then I agree that modeling reasons would (almost always) be pragmatic reasons. But this vague use of the term pragmatic would include both generally applicable modeling reasons motivated by limited modeling resources and highly context-specific reasons tied to specific modeling aims/purposes. While I commend this focus on the constraints/contexts in which models are built, in addition to highlighting the various purposes to which models can be put, it is also important to highlight the kinds of reasons that can be given for variable/parameter choice *across all (or at least most) modeling contexts*. A key part of what is interesting and useful about identifying these modeling reasons is that they focus on the general challenge of attempting to construct useable/workable scientific models from limited modeling resources. Although, in practice, these generally applicable modeling reasons will typically interact with more context-specific pragmatic purposes for models—and those interactions are interesting to investigate in their own right—it is also important to highlight the general considerations that limit/constrain scientific model building across large swatches of scientific practice.

A final objection might suggest that even though modeling reasons can be sufficient for variable/parameter choices, if we have ontological reasons for choosing certain variables they should always override the modeling reasons. A somewhat different way that Batterman (2021) poses this idea is to suggest that the reason why certain kinds of mathematical modeling techniques are able to be successful/effective across multiple contexts is because of the way the world is. Indeed, as Batterman suggests, even without the FDT, the success of mesoscale modeling strategies gives us reason to think that one ought to include those mesoscale variables and eschew smaller scale details. I think this generalization of the lessons of Batterman’s FDT cases is on the right track, but that more specific versions of it involving the reality and naturalness of particular variables/parameters are less feasible. There certainly are ways that real systems are—for example, having dependence and independence relations, various degrees of autonomy, separable scales—that are essential for our ability to successfully use (mesoscale) mathematical models to explain and understand their behaviors. This warrants some kind of realist, or naturalist, claims about how our mathematical models track features of reality. But this rough “tracking

of reality” claim is much weaker than suggesting that the effectiveness of mathematical models that include particular variables/parameters enables us to infer that those variables/parameters reflect, mirror, or accurately represent natural kinds more generally. Mathematical modeling can be justified by tracking features of reality without the additional claim that the best variables/parameters to use within a mathematical model will (always) be ones that accurately describe real or natural properties.

7. Conclusion

This article has distinguished causal modeling, metaphysical, epistemic, and modeling reasons for choosing certain variables and parameters with which to model and characterize a phenomenon. I have argued that scientific modelers typically do, and should, justify their choice of a particular set of variables and parameters by appealing to modeling reasons concerning the available measurements/data, computational resources, modeling techniques, theories, and modeling frameworks. Adopting this approach has enabled the identification of specific and generally applicable criteria that ought to guide variable/parameter choice across scientific modeling contexts. Going forward, philosophers of science should continue to investigate the ways in which the available modeling resources constrain scientists’ choices about which variables/parameters to include in their models. Doing so will help illuminate the context(s) in which scientific model construction takes place, help guide scientists’ selection of variables/parameters, and clarify the inferences that ought to be drawn concerning the variables/parameters within our best scientific models.

Acknowledgments. Thanks to Robert Batterman, Julia Bursten, Kareem Khalifa, Domenica Romagni, and the audience at the Philosophy of Experiment Workshop at the University of Stockholm for comments and feedback on earlier versions of this paper. This work was supported in part by the endowment fund of the Department of Philosophy, Colorado State University.

References

- Ariew, André, Collin Rice, and Yasha Rohwer. 2015. “Autonomous Statistical Explanations and Natural Selection.” *British Journal for the Philosophy of Science* 66 (3):635–58. <https://doi.org/10.1093/bjps/axt054>
- Bassingthwaighe, James, Peter Hunter, and Denis Noble. 2009. “The Cardiac Physiome: Perspectives for the Future.” *Experimental Physiology* 94 (5):597–605. <https://doi.org/10.1113/expphysiol.2008.044099>
- Batterman, Robert W. 2013. “The Tyranny of Scales.” In *The Oxford Handbook of Philosophy of Physics*, edited by Robert Batterman, 255–86. Oxford: Oxford University Press.
- Batterman, Robert W. 2021. *A Middle Way: A Non-Fundamental Approach to Many-Body Physics*. Oxford: Oxford University Press.
- Batterman, Robert W., and Collin Rice. 2014. “Minimal Model Explanations.” *Philosophy of Science* 81 (3):349–76. <https://doi.org/10.1086/676677>
- Bokulich, Alisa. 2011. “How Scientific Models Can Explain.” *Synthese* 180:33–45. <https://doi.org/10.1007/s11229-009-9565-1>
- Bokulich, Alisa. 2021. “Taming the Tyranny of Scales: Models and Scales in the Geosciences.” *Synthese* 199:14167–99. <https://doi.org/10.1007/s11229-021-03416-w>
- Callan, Herbert. 1985. *Thermodynamics and an Introduction to Thermostatistics*. New York: Wiley.
- Castiglione, Filippo, Francesco Pappalardo, Carlo Bianca, Giulia Russo, and Santo Motta. 2014. “Modeling Biology Spanning Different Scales: An Open Challenge.” *Biomed Research International* 2014:1–9. <https://doi.org/10.1155/2014/902545>

- Elgin, Catherine Z. 2017. *True Enough*. Cambridge, MA: MIT Press.
- Fisher, Ronald A. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- Garlick, Martha J., James A. Powell, Mevin B. Hooten, and Leslie R. McFarlane. 2011. "Homogenization of Large-Scale Movement Models in Ecology." *Bulletin of Mathematical Biology* 73:2088–2108. <https://doi.org/10.1007/s11538-010-9612-6>
- Giere, Ronald. 2004. "How Models Are Used to Represent Reality." *Philosophy of Science* 71 (5):742–52. <https://doi.org/10.1086/425063>
- Jhun, Jennifer, Patricia Palacios, and James O. Weatherall. 2017. "Market Crashes as Critical Phenomena: Explanation, Idealization and Universality in Econophysics." *Synthese* 195 (10):4477–4505. <https://doi.org/10.1007/s11229-017-1415-y>
- Khalifa, Kareem. 2017. *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press.
- Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.
- Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (4):343–77.
- Morrison, Margaret. 2004. "Population Genetics and Population Thinking: Mathematics and the Role of the Individual." *Philosophy of Science* 71 (5):1189–1200. <https://doi.org/10.1086/425241>
- Oden, J. Tinsley. 2006. *Simulation-Based Engineering Science: Revolutionizing Engineering Science through Simulation*. Washington, DC: NSF Publications. www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- Parker, Wendy. 2020. "Model Evaluation: An Adequacy-for-Purpose View." *Philosophy of Science* 87 (3):457–77. <https://doi.org/10.1086/708691>
- Pincock, Christopher. 2012. "Mathematical Models of Biological Patterns: Lessons from Hamilton's Selfish Herd." *Biology and Philosophy* 27: 481–496. <https://doi.org/10.1007/s10539-012-9320-8>
- Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago: University of Chicago Press.
- Rice, Collin. 2021. *Leveraging Distortions: Explanation, Idealization, and Universality in Science*. Cambridge, MA: MIT Press.
- Sider, Ted. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Strevens, Michael. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.
- Walker, Dawn C., and Jennifer Southgate. 2009. "The Virtual Cell—A Candidate Co-Ordinator for 'Middle-Out' Modelling of Biological Systems." *Briefings of Bioinformatics* 10 (4):450–61. <https://doi.org/10.1093/bib/bbp010>
- Weisberg, Michael. 2017. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
- Winsberg, Eric. 2018. *Philosophy and Climate Science*. Cambridge: Cambridge University Press.
- Woodward, James. 2016. "The Problem of Variable Choice." *Synthese* 193:1047–72. <https://doi.org/10.1007/s11229-015-0810-5>

Cite this article: Rice, Collin. 2025. "A Model-Based Approach to the Problem of Variable Choice." *Philosophy of Science* 92 (2):432–452. <https://doi.org/10.1017/psa.2024.52>