**The Experiment-Applicability Tradeoff in Scientific Model Building**

**Citation:** Rice, C. The experiment-applicability tradeoff in scientific model building. *Synthese* **206**, 124 (2025). <https://doi.org/10.1007/s11229-025-05215-z>

**Please Cite Published Version!**

**Collin Rice**

**Colorado State University**

**Department of Philosophy**

**corice@colostate.edu**

**Abstract.** Ever since Richard Levins’s (1966) influential article “The Strategy of Model Building in Population Biology” there has been a growing discussion surrounding the types of tradeoffs that confront scientific model builders (Orzack and Sober 1993; Odenbaugh 2003; Weisberg 2006, 2012). However, almost all the discussion of Levins’s modeling tradeoffs has focused on what Michael Weisberg calls the ‘representational ideals’ of scientific models. In this paper I argue that this emphasis on representational aims misses many of the *pragmatic* tradeoffs that scientific modelers confront due to limited experimental data, measurement tools, modeling frameworks, and other modeling resources. In response, this paper aims to investigate the pragmatic modeling tradeoff between (1) having a model be constructable from, and testable against, the available experimental data and (2) building models that are able to generalize across a wide range of contexts of application. I argue that this experiment-applicability tradeoff is a relationship of attenuation rather than a strict or necessary tradeoff between model properties. I then use three case studies to show that, rather than a strict theoretical limit, how this tradeoff is best navigated is highly context sensitive. I then explore the philosophical implications of this tradeoff for how we ought to think about theories as collections of models, how models connect with experiments, and how modelers balance various modeling aims.

**1. Introduction**

Ever since Richard Levins’s (1966) influential article “The Strategy of Model Building in Population Biology” there has been a growing discussion surrounding the types of tradeoffs that scientific model builders face (Matthewson and Weisberg 2009; Odenbaugh 2003; Orzack and Sober 1993; Walmsley 2021; Weisberg 2006, 2013). Levins’s tradeoffs between modeling aims such as realism, generality, and precision connect with several philosophical debates concerning scientific realism, unification, predictive power, explanation, and idealization. However, almost all the discussions of these modeling tradeoffs have focused on what Michael Weisberg (2006, 2007, 2013) calls the ‘representational ideals’ of scientific models. That is, the focus has been on determining how best to navigate the tradeoffs between different *representational* goals that scientific modelers have concerning which features of the target system(s) to include and how accurately or precisely to represent those features. While this investigation into the representational aims of models has been insightful[[1]](#footnote-1), in this paper I argue that it misses many of the *pragmatic* modeling tradeoffs that scientific modelers must confront due to limited experimental data, measurement tools, modeling frameworks, and other modeling resources. In other words, I argue that focusing on the representational relationships between models and their target system(s) has caused the philosophical literature on tradeoffs in model building to miss other kinds of tradeoffs that arise out of the pragmatic constraints that influence the choice of various modeling strategies.

In response, this paper aims to investigate an important, but previously unanalyzed, pragmatic modeling tradeoff that grows out of the constraints and limitations of the contexts in which scientific models are constructed and used. Specifically, beyond just representing target systems in different ways, scientific modelers would like their models to make contact with the available experiments and measurements that can be performed within particular real-world systems. That is, scientific modelers often aim to have their models be informed by, or comparable against, experimental data. Indeed, several philosophers of science have noted that, scientific modelers often choose “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.” (Rice 2024, 9).[[2]](#footnote-2) In addition, scientific modelers typically aim to have their models be generally applicable across a wide range of contexts. That is, scientific modelers aim to construct models that can be used in the investigation of many different types of systems. However, these two modeling aims often pull scientific modelers in opposing directions. As a result, I will argue that scientific modelers often face a tradeoff between (1) having a model be constructable from, or testable against, the available experiments/measurements and (2) building models that are able to generalize across a wide range of contexts of application. More specifically, I will argue that this experiment-applicability tradeoff means that accomplishing one of these modeling aims *tends* to make accomplishing the other aim more difficult—i.e., it is a relationship of *attenuation* rather than a strict/necessary tradeoff.

My argument depends on a sensible distinction between the pragmatic (or practical) features involved in constructing models for particular purposes/contexts and the semantic (or representational) properties of models concerning how accurately they represent their target systems (Odenbaugh 2003; Parker 2020).[[3]](#footnote-3) While several philosophers have investigated various kinds of pragmatic constraints on scientific modeling (Odenbaugh 2003; Parker 2020; Rice 2024), these discussions have not analyzed the specific types/kinds of *tradeoffs* that exist *between* these different pragmatic constraints or aims in any detail.[[4]](#footnote-4) For example, although Rice’s recent article on variable choice briefly mentions that various pragmatic modeling criteria “ought to be weighed against each other and applied collectively” (2024, 448), it does not offer any detailed investigation into the specific tradeoffs between those aims. In addition, although Wendy Parker (2020) discusses some of the pragmatic considerations involved in building models for specific purposes, she does not offer any analysis of the various tradeoffs that hold between those pragmatic aims/goals. Thus, while the present paper certainly relates to the projects being pursued in those other articles, it covers new ground by directly investigating the ways that these pragmatic modeling aims tradeoff with (or constrain) one another.

After identifying the pragmatic experiment-applicability tradeoff, I move to argue that how the tradeoff is best navigated in scientific practice is highly context dependent. I argue for this context-dependence by looking at three case studies in which differences in the modeling context result in different answers about how to best navigate the experiment-applicability tradeoff. This shows that, contrary to how they have been framed by other philosophers, many of the tradeoffs of model building are not absolute limits on the maximization of the features of models. Instead, modeling tradeoffs are often instantiated in context-dependent ways that differentially impact the decisions of practicing scientific modelers.

In addition to its interest to the scientific modeling and experimentation literatures, identifying this experiment-applicability tradeoff also has several interesting implications for other debates within philosophy of science. First, it helps make sense of the claim that scientific theories are often collections of models that navigate this tradeoff (and others) in different ways (Levins 1966; 1960; Sober 2000). Second, it allows us to better elaborate how such (collections of) models mediate between (generally applicable) theories and (localized) experimental results (Morgan and Morrison 1999; Suppes 1962). Third, it shows how important the availability of experimental data is for navigating the pragmatic constraints of model building—i.e. scientific experimentation and scientific modeling are often deeply intertwined (Morrison 2009; Parke 2014; Winsberg 2003). Fourth, investigating the context sensitivity of the experiment-applicability tradeoff helps clarify the ways that scientific modelers must balance a multitude of intertwined modeling aims that often tradeoff with one another in a variety of ways.

 The following section briefly reviews the literature on tradeoffs in scientific model building that has grown out of Levins’s original paper. Then, in Section 3, I present the experiment-applicability tradeoff and argue that it is unique in being largely independent of (philosophers’ characterizations of) Levins’s representational goals of realism, generality, and precision. Next, in Section 4, I use three case studies to show the context-sensitive ways scientific modelers navigate this tradeoff differently in different modeling contexts. Finally, Section 5 explores the implications of the experiment-applicability tradeoff for other debates in philosophy of science and the various relationships between models and experiments.

**2. Tradeoffs in Scientific Model Building**

*2.1 Levins’s tradeoffs*

Levins’s original discussion focused on arguing against what he calls a “naïve, brute force approach” (1966, 421) that would attempt to build models that would faithfully reflect each of the plethora of factors involved in biological populations. According to Levins, this approach fails for three reasons:

1. There are too many parameters to measure.
2. The equations are insoluable analytically—even for computers.
3. Even if they could be solved, their solutions would be so complicated as to have no meaning for us. (Levins 1966, 421)

Interestingly, Levins’s *critiques* of the brute force approach acknowledge several of the pragmatic constraints on scientific modeling that I wish to highlight here. Specifically, regarding (1), he notes that one of the challenges facing population biologists is that “there are too many parameters to measure; some are still only vaguely defined; many would require a lifetime each for their measurement” (Levins 1966, 421). Yet, this challenge of building models from the available measurements plays little to no clear role in Levins’s positive view of the tradeoffs that biological modelers must address and the modeling strategies they adopt as a result. Instead, Levins suggests that rather than including *all* the factors of the target system, “clearly we have to simplify the models in a way that preserves the essential features of the problem” (1966, 421). He then proposes that “It is of course desirable to work with manageable models which maximize generality, realism and precision toward the overlapping but not identical goals of understanding, predicting and modifying nature. But this cannot be done.” (Levins 1966, 422). In this paragraph Levins shifts from various modeling challenges for the brute force approach concerning measurement, solvability, and complexity (i.e. building manageable models) towards three *representational* aims of models. Notably, Levins initially suggests that there are really *four* desiderata—manageability, realism, generality, and precision—but then moves on to only discuss tradeoffs between the latter three representational aims. Specifically, Levins argues that achieving all three of the aims of generality, realism, and precision in the same model is impossible.[[5]](#footnote-5) At most, model builders can maximize two of them. He goes on to present various modeling strategies for coping with this limitation. The first strategy involves sacrificing generality in order to increase realism and precision. The second sacrifices realism in order to improve generality and precision. And the third sacrifices precision to realism and generality.[[6]](#footnote-6)

*2.2. Orzack and Sober’s formalization of generality, precision, and realism*

In their analysis of Levins’s tradeoffs, Steven Orzack and Elliott Sober (1993) focus exclusively on Levins’s discussion of the aims of generality, precision, and realism. Specifically, they critique Levins’s proposal of a three-way tradeoff by offering formal definitions of these representational aims—which Levins notably does not do. According to Orzack and Sober, generality, realism and precision of a mathematical model ought to be formalized as follows:

(G) If one model applies to more real-world systems than another, it is

more general,

(R) If one model takes account of more independent variables known

to have an effect than another model, it is more realistic,

(P) If a model generates point predictions for output parameters, it is

precise (1993, 534).

Using these formal definitions of model properties, Orzack and Sober go on to argue that Levins’s proposal that we maximize two of these properties while sacrificing a third fails. Specifically, they argue that when one model is a limiting case of another, generality, realism, and precision can all be maximized simultaneously; contrary to Levins’s view. What I want to note here is that Orzack and Sober analyze Levins’s view solely in terms of their formalized definitions of the semantic/representational properties of models. As a result of this characterization of Levins’s view, there is no discussion of the role of employing measurable parameters in one’s model. Nor is there any discussion of whether the models involved can be solved analytically or would generate understanding for the modelers. In short, there is no discussion of the need to build *manageable* models from existing modeling resources.

The problem is that these formal definitions of generality, realism, and precision takes Orzack and Sober’s discussion fairly far away from the pragmatic constraints on the practice of model building that are the focus of much of Levins’s discussion. This issue is clearly illustrated by Jay Odenbaugh (2003) who argues that “The fundamental problem with Orzack and Sober’s analysis is that their notion of a ‘necessary trade-off’ concerns the logical or semantical relations between model properties. However, Levins’s notion…concerns our cognitive limitations and use of mathematical representations and not the logical or semantic properties of models alone” (1506). Following Odenbaugh, while I think Orzack and Sober’s formalization of the semantic properties of models is fine as far as it goes, it leaves out many other pragmatic aspects of model building in favor of focusing on the logical relationships between the semantic properties of models in limiting cases.

*2.3 Weisberg’s representational ideals*

Later on, John Matthewson and Michael Weisberg characterized Levins’s tradeoffs in terms of attempting to provide several distinct ‘representational aims’. Specifically, in several places, Weisberg (2006, 2007, 2013) uses his framework of understanding idealization strategies via various *representational ideals* to characterize the modeling tradeoffs discussed by Levins. For example, Weisberg suggests that the brute force approach aims at the representational goal of ‘completeness’ and that Levins’s other idealization strategies for modeling aim at alternative representational goals such as ‘generality’. Indeed, on Weisberg’s framing, Levins’s three idealization strategies “try to include only the most important or the most relevant aspects, which depends on the interests of the theorists” (2006, 633). This focus on which relevant features to represent in the model leads to Weisberg’s central question for the idealization approach: “How do we know when to be satisfied with the *accuracy* of our models, given that we have committed ourselves to their being inaccurate from the start?” (2006, 633). Again, this frames the aims and modeling decisions involved in selecting a modeling strategy largely in terms of which features to include in the representation, how accurately to represent those features, and which features to idealize. In later work, Weisberg explicitly argues that Levin’s tradeoffs help explain the construction of multiple models that aim to achieve different representational ideals (2007, 2013) or are the result of the logic of representation (Weisberg 2004; Matthewson and Weisberg 2009). These representational aims are, again, focused exclusively on the semantic relationships between models and their target system(s). For example, Matthewson and Weisberg explicitly interpret Levins’s modeling aim of generality as “concerning the model-target relationship”, “representational fidelity criteria”, and “how well a model describes the causal structure of the target system” (Matthewson and Weisberg 2009, 181). In general, Weisberg’s framework of representational ideals focuses attention on what the modeler aims to *represent* about the target system(s). Of course, the rest of Weisberg’s important work on modeling explicitly addresses many of the pragmatic aspects of modeling that are my focus here. However, when it comes to Levins’s modeling tradeoffs, Matthewson and Weisberg repeatedly frame them solely in terms of balancing the various representational ideals (or aims) that scientific modelers have.

**3. A New Kind of Tradeoff: Pragmatics, Experiments, and Applicability**

While this literature concerning the tradeoffs among representational aims of modelers has certainly produced important insights, I contend that it misses other important tradeoffs that practicing scientific modelers face. For one thing, accomplishing different representational aims with respect to some target system(s) is only *one of many* aims of scientific modelers. That is, the aims scientific modelers seek to optimize are broader than just representational aims. Specifically, I argue that the practice of model building gives rise to numerous tradeoffs and modeling decisions that result from *pragmatic* tradeoffs that are largely independent of the representational aims of generality, precision, and realism. As a result, to fully characterize the tradeoffs involved in the practice of model building, we need to expand beyond strict/necessary tradeoffs between representational aims/desiderata and look at the pragmatic constraints highlighted by Levins’s original critiques of the brute force approach. In short, the tradeoffs of scientific modeling are not just tradeoffs about the *representational ends/ideals* concerning relationships between the model and its target system(s), but also includes numerous tradeoffs concerning *how to build useful models from the available* *experimental and modeling resources*.[[7]](#footnote-7) Expanding the discussion of tradeoffs in model building in these ways has the potential to breathe new life into these debates and bring the literature on modeling tradeoffs into conversation with other philosophical questions concerning the nature and role of experimentation (Knuuttila and Loettgers 2018; Parke 2014; Suppes 1962) and the pragmatic application of models for specific purposes (Parker 2020; Rice 2024), without framing these questions solely in terms of representational aims or a model’s semantical properties. I suspect that further tradeoffs between various modeling aims will be identified by future investigations—e.g. tradeoffs between representational, pragmatic, or epistemic aims of scientific modelers.

As Odenbaugh (2003, 2006) argues, Levins’s discussion is largely concerned with the pragmatics of the practice of model building rather than tradeoffs that necessarily hold between well-defined representational desiderata. Although Levins goes on to focus on (rather vague conceptions of) a model’s generality, precision, and realism, the original problems he raises for the brute force approach “are all problems that arise as products of scientists *and* their models, and *not* from the models alone” (Odenbaugh 2006, 612). It is this pragmatic focus on the construction and use of models by modelers and the resulting constraints on model building that suggests other ways that Levins’s framework of tradeoffs might be usefully developed.[[8]](#footnote-8)

Specifically, in the rest of this paper, I investigate an important, but unanalyzed, pragmatically motivated tradeoff faced by model builders and argue that neither side of the tradeoff is best interpreted as a representational/semantic relationship between a model and its target system(s). As a result, this tradeoff is importantly different from the modeling tradeoffs identified by other philosophers.

 While there are numerous pragmatic constraints and limitations that scientific modelers face, I here focus on two prominent modeling aims that I argue exhibit a tradeoff within scientific practice:

1. Constructing models whose parameters and predictions are derivable from, or comparable against, the available experiments/measurements.[[9]](#footnote-9)

(2) Building models that are applicable across a wide range of modeling contexts.

The issue, however, is that these two modeling aims often tradeoff against one another in scientific practice such that accomplishing one of these aims often makes accomplishing the other aim more difficult—i.e., there is a relationship of attenuation between them.

On the one hand, building models whose parameters are more directly comparable with experiments performed in particular systems tends to make the model less applicable in other contexts. One reason for this tradeoff is that the available experiments/measurements are typically performed *within particular systems* and applicability is achieved by using the model to investigate a wide range of *different systems.* As a result, filling in a model’s parameter values by using experiments from particular systems tends to make the model less applicable to other situations where those parameter values are very different, unmeasurable, or perhaps nonexistent.

Another reason this tradeoff occurs is that the spatial and temporal scales of the general patterns that scientists would like to explain and understand are often quite different from the scales at which experimental studies can be performed. As a result, there is often a tension between employing (typically coarse-grained) parameters at the scale of the phenomenon (or pattern) of interest and using parameters at different scales that are comparable with the available experimental data. In ecology, for example, the phenomena ecologists want to understand with their models typically involve many species “across decades to centuries, over large geographical areas” (Pimm 1991, 1). However, the experimental data available is typically at much smaller scales. As Stuart Pimm puts it:

Look at any current ecology journal and you will see that our studies are often very brief: ten years is a long-term study. In the United States, such a study would require a minimum of four consecutive National Science Foundation grants… Studies, especially experimental studies, are generally on a scale of square meters and rarely cover more than a few hectares. The typical community ecologist…tends to study no more than a dozen species” (1991, 1)

The issue is that the scales at which experiments can be performed are often quite different from the scales at which the phenomena of interest occur. Levins echoes this issue concerning the timescales of measurements within particular systems by noting that models such as those built by Watt (1956), “reduce the parameters to those relevant to the short-term behavior of their organism, make fairly accurate measurements, … and end with precise testable predictions applicable to these particular situations” (Levins 1966, 22). What I want to note about this example is not that Watt’s parameters are *precisely defined*, but rather that they target *short-term features* of the system that are *applicable only* *within particular situations*. This enables the parameters of Watt’s models to be directly comparable with experiments performed in specific real-world systems, but inevitably limits their application to other systems. Of course, this challenge of wanting to build models of more general phenomena from experiments within particular systems does not only confront ecologists (Batterman 2021; Walmsley 2021). As other biological multiscale modelers put it: “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). In short, which parameters are ultimately included in a model is not determined just by what makes a difference or is relevant to the phenomenon of interest. Instead, scientific modelers are also heavily influenced by which parameters can be *measured* within particular systems and at what *scales* those experiments are possible (Batterman 2021; Rice 2024). More generally, what I hope to have shown is that being able to compare a model’s parameters with the available experimental data acquired within particular systems, and at particular scales, *tends* to make the model less generally applicable to other situations.

 On the other hand, making models generally applicable by coarse-graining parameters or tapping into general mathematical frameworks (e.g. statistical modeling) tends to result in parameters that are more difficult to compare with the available experimental data. This is because the mathematical parameters that are generally applicable across many systems are often unmeasurable within particular systems or experiments. For example, although various parameterizations of fitness are applicable to a wide range of evolving systems, most operationalizations of fitness are not directly measurable within those systems. This is evidenced by biologists routinely using various “proxies” for fitness such as average energy intake, average number of eggs fertilized, or number of food items consumed.

In many cases, this tension between general applicability and comparability with measurement occurs because broader application is achieved by abstracting away from the details that are heterogeneous across different systems, but it is precisely those heterogeneous details that are possible targets for experimental manipulation. In other cases, the tradeoff occurs because the generally applicable parameters are ‘artifacts’ of the widely applicable computational or modeling frameworks, but as a result have no corresponding variables or features within real-world systems that could be the target of experimental studies. An interesting example here is ecosystem integrity. While widely applicable in a variety of conservation contexts, this parameter is difficult, if not impossible, to directly measure within real-world populations (Rohwer and Marris 2021). As a result, beyond just being *less specific* about parameter values, scientists often construct general models by using *completely* *different* parameters that are widely applicable across multiple systems, but whose values are unmeasurable in actual systems. Consequently, this increase in applicability of a model across different contexts often comes with the cost of having the model’s parameters be difficult or impossible to measure in any real-world system.

Here one might be reminded of Patrick Suppes suggestion that scientific practice often involves the construction of a hierarchy of models that connect theories with experimental data in different ways (Suppes 1962).[[10]](#footnote-10) One important difference here is that Suppes’ view focuses almost exclusively on experimental data that can be presented numerically and subject to statistical analysis (Leonelli 2019, 7). Yet, as Sabina Leonelli notes, “especially within the biological, social, environmental, historical, and health sciences, Suppes’ hierarchy of models proves difficult to apply” (2019, 2). In contrast, my investigation of the experiment-applicability tradeoff does not restrict what is considered data or a model of data in these ways. This enables my analysis to be broadly applicable to areas where the relationships between models and data take forms other than statistical/formal analysis (see Leonelli 2019 for further discussion).

Despite this, Suppes’ use of set theory and statistical analyses of data models might be separated from his suggestion that scientific modeling typically involves the construction of a hierarchy of models that connect theoretical models with data by filling in the model in various ways.[[11]](#footnote-11) However, another key difference, that is highlighted by the case studies discussed below, is that on my view modelers can respond to the experiment-applicability tradeoff by building *singular* models that navigate the experiment-applicability tradeoff in different ways. That is, while scientific modelers certainly *can* build hierarchies of models that connect theory with experiment in different ways, such a collection/hierarchy of models is not a necessary response to the experiment-applicability tradeoff. Instead, individual models (or modelers) might navigate this tradeoff in different ways even if those models are not more specific or more general instances of each other.[[12]](#footnote-12)

This allows us to distinguish two different responses to a modeling tradeoff that are typically run together within the literature. In much of the existing literature, philosophers have proposed that the response to a tradeoff that makes it impossible to maximize various modeling desiderata is to construct *multiple models* that satisfy different desiderata (or aims). We see this clearly in Weisberg’s suggestion that the tradeoffs identified by Levins give us reason to expect the construction of multiple models that idealize the system in different ways—what Weisberg (2007) refers to as multiple models idealization. The question then becomes how to “get a handle on this proliferation” (Weisberg 2013, 174). Here, there is a modeling tradeoff between aims that is responded to by building different models (that are perhaps part of a hierarchy) that accomplish different modeling aims. However, in Levins original discussion, and in the cases discussed below, the initial suggestion is that scientific modelers respond to these modeling tradeoffs by building individual models that accomplish one (or a few) modeling aims while sacrificing others. For Levins, given that generality, precision, and reality cannot all be maximized within a single model, scientific modelers tend to either “sacrifice generality to realism and precision” or “sacrifice precision to realism and generality” (1966, 422). Levins does go on to suggest that these different strategies mean that modelers eventually end up having to “treat the same problem with several alternative models” (423), but this is not strictly speaking necessary; nor do those models necessarily constitute a hierarchy that links theoretical models with experimental data. Therefore, scientific modelers can respond to modeling tradeoffs by either (1) constructing individual models that maximize some modeling aims and sacrifice others or (2) constructing multiple models that accomplish different modeling aims.

One concern here is that if scientific modelers are able to accomplish different modeling aims by building collections or hierarchy of models, then in what sense is there still a modeling tradeoff operating in these cases? After all, there seems to be no need to sacrifice one modeling aim to accomplish another. For example, if Suppes is right that a hierarchy of models is involved in connecting theoretical models with data models, then scientists seem to be able to maximize widespread applicability with some models within the hierarchy and comparability with experimental data via different models. I suggest that this is still an instance of the experiment-applicability tradeoff operating within scientific practice—though that tradeoff is being responded to differently. Generally, the existence of a modeling tradeoff makes it such that multiple modeling aims/desiderata cannot be maximized *within the same model*.[[13]](#footnote-13) Once this modeling tradeoff between two (or three, or more) modeling aims is recognized, scientific modelers face a choice. They can either build a single model that optimizes some modeling aims while sacrificing others or build a collection (or hierarchy) of different models that each focus on different modeling aims. Both, I suggest, are responses to the existence of a modeling tradeoff that makes it impossible (or difficult) to maximize multiple modeling aims within the same model.

Now, of course, if Suppes is correct that “a whole hierarchy of models stands between the model of the basic theory and the complete experimental experience” (Suppes 1962, 260), then every comparison of a model with the available experimental data will involve multiple models including models of data and models of experiment. But this just serves to highlight that the relationship between models and experimental data is a complicated one. This is certainly the case. However, this merely complicates (or analyzes) *one side* of the experiment-applicability tradeoff rather than directly addressing the tradeoff itself or how it might be responded to in different scientific contexts (which is my focus here). I compare my view with Suppes’ semantic view of theories and the building of multiple models a bit more in Section 5.1.

To highlight the uniqueness of this tradeoff, I now argue that neither of these modeling aims (comparability with experiment and general applicability across contexts) are best interpreted as representational ideals or semantic properties of the model. On one side of the tradeoff, being comparable with experiments and measurements is not a semantical feature of the model nor is this comparability with data something that is explicitly represented about the model’s target system(s). It is, instead, a feature *of the modeling context* that involves relationships between the model’s parameters, the measurements/data acquired from various experiments, and the measurability of those parameters in the model’s target system(s). In short, comparability with experiment is not a semantic property of a model determined by a representational relationship between the model and its target system(s), but is determined by how measurable certain features are and whether performing such measurements is feasible given the available resources. Therefore, whether a model’s parameters are comparable with experiment is primarily a pragmatic question about how the model is built, used, and applied in a particular context rather than a semantic question about how accurately or precisely the model represents features of its target system(s).

At this point one might object that choosing to include a particular parameter that is directly comparable to the available experimental data still results in a representational difference in the model’s content.[[14]](#footnote-14) That is, one might argue that even if the parameters are introduced for pragmatic/practical reasons of comparability with experimental data, they are still parameters within the model and so change its representational content. So, why is this newly introduced parameter not a semantic property of the model? My reply is that, although including these parameters does result in a change in the representational content of the model, their introduction is not *motivated* by those representational aims/goals. The claim is *not* that these parameters are somehow separate from the representational content of the model, but that they are not introduced because they accurately represent relevant features of the model’s target system(s). That is, the introduction of these parameters is not driven by the modeler’s aims of more accurately or precisely representing certain features of the model’s target system(s). Instead, many modeling parameters are introduced explicitly because they are comparable with the available experimental data—whether or not they accurately represent features of the model’s target system(s) or are present in those systems at all. Consequently, although they result in the model having different representational content, it is their comparability with the available measurements/data (a pragmatic aim of the modeler) rather than their representing the model’s target system in a particular way (a representational aim of the modeler) that is doing the real work here. Several philosophers have noted that assessing models in terms of their ability to satisfy pragmatic/practical purposes “can be contrasted with one on which model quality is (just) a matter of how accurately and completely a model represents a target” (Parker 2020, 458). To see the point a different way, suppose that including Q or Q’ would make the model equally representationally accurate, but Q’ is much easier to directly compare to experimental data. I suggest that scientific modelers will often choose Q’ for that reason (alone) independent of whether Q or Q’ enables the model to more accurately (or precisely) represent its target system(s).

On the other side of the tradeoff, I argue that there is a crucial distinction between *applying* a modelacross multiple systems and *accurately* *representing* those systems (see Elliott-Graves 2022 for a nice discussion of different senses of generality and applicability). Unfortunately, much of the modeling literature has conflated these concepts by suggesting that models are applicable to a real or possible system just in case the model’s assumptions are satisfied by the system—i.e. if the model accurately represents the relevant features of the system. For example, Weisberg suggests that Levins’s models “apply to many systems because few assumptions were made about the exact nature of the target system being modeled” (2006, 634). That is, they apply to many systems because they make only a few assumptions that need to be satisfied for the model to apply to them. Indeed, as we saw above, Matthewson and Weisberg interpret Levins’s notion of generality in terms of representational fidelity criteria that asks, “how well a model describes the causal structure of the target system” (2009, 181). Similarly, Orzack and Sober argue that Levins’s conception of generality is largely ambiguous because it depends on the interpretation of the model. They illustrate this point primarily by characterizing models as conditional if-then statements and then arguing that it is unclear whether the antecedent of the model is satisfied by real-world systems (1993, 535). However, many scientific models are applied/used to investigate systems that do not satisfy their assumptions; e.g., the ideal gas law (Elgin 2007). Moreover, many models are applied to systems whose relevant causal structures are drastically misrepresented by the model (Potochnik 2017; Rice 2021). Consequently, a model being generally applicable to many systems is importantly different from the question of whether the model accurately represents the relevant causal structures of those systems or how many systems satisfy the assumptions of the model.[[15]](#footnote-15)

As a result of these differences, the experiment-applicability tradeoff is also importantly different from the tradeoff between building general or precise models that has been widely discussed by other philosophers. As Weisberg (2006, 2013) and Matthewson and Weisberg (2009) note, including more features of the system, or describing them more precisely, tends to make the model less able to be satisfied by a wide range of real or possible systems—i.e. this increase in precision of the parameters sacrifices the ability for the model to accurately describe the features of multiple real or possible systems in one (formalized) sense. Above I argued that this sense of generality in terms of having multiple systems satisfy the representational features of the model (what Weisberg calls representational fidelity) is crucially different from the kind of general application I have in mind. In short, models can be applied to systems they do not accurately represent and may not be applied to systems they do accurately represent. Thus, (this sense of) generality and applicability come apart.

More importantly, though, the desire of scientific modelers to be able to compare the parameters of their models with the available experiments and measurements is quite different from having a preference for more precise model parameters. Matthewson and Weisberg (2009) define precision in terms of the “fineness of specification of the model description’s parameters”, which they suggest is “neither an attribute of collected data, nor the output of a calculation based on the model” (178). However, if this is how we define model precision, then comparing the model’s parameters with the available experimental data will not always make the model more precise in this sense (and sometimes just the opposite). And making the model’s parameters more (or less) precise will not always make the model more comparable with the available experiments. Generally, the desire to compare the parameters of a model with the available experiments/measurements is importantly different than having a preference for models with more (or less) precisely defined parameters. As a result, both sides of the experiment-applicability tradeoff are importantly different than the tradeoff between generality and precision discussed by other philosophers.

Before moving on to some case studies, it is crucial to emphasize that this experiment-applicability tradeoff is not a *strict* tradeoff in which increasing one aspect of a model *necessarily* reduces the other aspect. Instead, this is a weak tradeoff in which—all else being equal—increasing one aspect of a model *tends* to make increasing the other aspect more challenging. Matthewson and Weisberg refer to these as relationships of ‘simple attenuation’ which occur “when an increase in the magnitude of one attribute makes the achievement of another more difficult” (2009, 170). I take the above discussion of the general relationships between these two modeling aims to provide several arguments for why we ought to expect this relationship of attenuation to hold across much of scientific modeling. Those arguments, in combination with the three case studies analyzed in the next section, provide reasons for thinking that the experiment-applicability tradeoff influences much (though perhaps not all) of the practice of scientific modeling. In addition, given its dependence on pragmatic features concerning model construction and application, the nature of this tradeoff and how it is best navigated is also highly dependent on various features of the modeling context. In fact, in some cases *both* sides of the tradeoff can be simultaneously increased. Consequently, this tradeoff is not a universal limit on what modelers can accomplish across all contexts. Instead, how the tradeoff functions in scientific practice and how it is best navigated is highly context dependent.

**4. Three Cases of Navigating the Experiment-Applicability Tradeoff**

To illustrate the context-dependent ways this tradeoff is navigated by practicing scientific modelers, in this section I analyze three case studies that show alternative ways of navigating the tradeoff between building models that are comparable with experimental data and constructing models that are applicable across a wide range of modeling contexts. In the first case, the focus is on maximizing comparability with the available experiments/measurements but at the cost of sacrificing the model’s ability to be applied in other cases. In the second case, general applicability is emphasized, but the parameters of the model are difficult to compare with experiments/measurements in particular real-world systems. In the third case, convenient features of the model’s target systems make it such that the construction of models that are both comparable with experimental results within particular systems and broadly applicable across multiple contexts is possible.

*4.1. Maximize comparison with experiment by sacrificing general applicability*

A striking example of the tradeoff between the ability to compare a model’s parameters with the available measurements/experiments and aiming to construct models that are generally applicable can be found in the literature on biological optimality models. For example, Geoffrey A. Parker and John Maynard Smith begin their review of optimality modeling in evolutionary biology by drawing the following distinction:

We distinguish between general models and specific models—though in reality they form part of a continuum. General models have a heuristic function; they give qualitative insights into the range and forms of solution for some common biological problem. The parameters tested may be difficult to measure biologically, because the main aim is to make the analysis and conclusions as simple and direct as possible. Specific models are designed to be applied quantitatively to particular species, and include parameters that can readily be measured. (Parker and Maynard Smith 1990, 27)

Although they frame the distinction between general and specific models—in ways that might be thought to be similar to Levins’s tradeoff between generality and precision—they spell this out in terms of general models providing qualitative insights, but having parameters that are *difficult to measure* and specific models being designed to have quantitative parameters that can be *easily measured* within particular systems. This, I contend, clearly reflects the tradeoff that model builders often face between constructing generally applicable models (that investigate common problems) and constructing models that can be compared to experimental results within particular systems.

 As an example of navigating the tradeoff by sacrificing general applicability in order to be able to compare the model’s parameters directly with experiments, we can look at Kenneth Wilson’s (1994) investigation of different optimality models for the evolution of clutch size in insects. After arguing that experimental data is required to differentiate between different optimality models of insect clutch size, Wilson’s aim in this study is to provide “a series of laboratory experiments, using the bruchid beetle…, that not only provide the fundamental data required to make such predictions but also enable quantitative comparisons to be made between observed behavior patterns and those predicted by the different models” (1994, 366). In other words, the goal is to make the model’s parameters and predictions directly comparable with the available experimental data.

 Specifically, these beetles lay their eggs on seeds, which burrow into the seed after three or four days, feeding on the seed for approximately twenty-six days, before emerging as adult beetles. However, since females often lay multiple eggs on a single seed “as the number of larvae per seed increases so too does the amount of competition between larvae for food.” (Wilson 1994, 367). A consequence of this is a fitness tradeoff between the benefit of additional larvae and competition between those larvae such that “the rate at which an ovipositing female accrues fitness declines as clutch size increases. This trajectory of fitness gain is often referred to as the ‘larval competition curve’” (Wilson 1994, 367). Without going into all the details, the issue is that there are lots of ways to parameterize this larval competition curve and the resulting fitnesses of different clutch sizes. To determine which of these parameterizations was best, Wilson conducted the following experiments on a specific beetle population:

Experiment 1: determined the effect of larval competition on emergence weight.

Experiment 2: determined the effect of female emergence weight on realized fecundity.

Experiment 3: measured the cost of reproduction.

Experiments 4 and 5: measured the effect of travel time and clutch size on oviposition time.

Experiment 6: measured the effect of conspecific females on clutch size.

Experiment 7: measured the effect of emergence weight on lifespan of virgin females.

Experiment 8: looked at the effect of female age and previous oviposition experience on clutch size.

Each of these experiments generated quantitative data drawn from experimental manipulations on a specific real-world beetle population. This experimental data was then compared with a wide range of quantitative parameters and predictions of various clutch size models (Wilson 1994, 375).

 What is most important about this example for my purposes here is that Wilson clearly makes the decision to compare the available models with a wide range of experimental data. However, in doing so, the resulting optimality models are highly restricted in terms of their range of application. Not only does clutch size and the larval competition curve only apply to certain types of organisms, but the detailed experiments were all performed within a single population of a single species. Moreover, the parameters chosen for experimental manipulation are all short-term features of the system that are relatively stable at those shorter time scales. Focusing on these easily measured short-term parameters allows the models to be directly comparable to the available experimental data drawn from specific beetle populations, but this modeling choice makes it difficult to generalize the quantitatively parameterized models so they might be applied to other biological organisms or populations.

*4.2. Maximize applicability by sacrificing the ability to be compared with experiments*

While scientific modelers certainly would (all else being equal) like their models to be directly comparable with experimental data, often the nature of the model’s target systems is such that experimentation to confirm parameters or test the model’s predictions is impossible or exceedingly difficult. For example, Nancy Nersessian describes how “bioengineering scientists face a major challenge in that they usually cannot experiment on biological systems directly. The complexity of the real-world (in vivo) phenomena makes experimentation too difficult, or even impossible, to control.” (Nersessian 2022, 25). More generally, Nersessian notes that biological modelers are routinely faced with severe ‘data constraints’: “The kind of experimental data (time series) needed for building dynamic models and parameter fitting is often not available or difficult to obtain, and the available data are usually noisy.” (Nersessian 2022, 180). In other cases, the relevant experimental data is long in the past, would require making measurements over multiple (human) generations, or involves parameters that change too quickly for the available experimental methods to detect. In short, in many cases of modeling complex systems, the relevant experimental data is often lacking or severely limited.

When the available experimental data is so limited, modelers often opt to have their mathematical models be generally applicable to a wide range of contexts.[[16]](#footnote-16) To build these generally applicable models, modelers routinely tap into mathematical frameworks whose applications can be widespread, but whose parameters are often difficult, if not impossible, to compare with the available experimental data from particular systems.This is often for two reasons: (1) mathematical frameworks often involve parameters that are constructed *de novo* within the representational framework (i.e. they are artifacts of the modeling approach) and (2) models are often made more general by abstracting away from the idiosyncrasies (or differences) of particular systems. The first reason entails that often general mathematical models involve parameters that simply have no measurable correlate in real-world systems. For example, Eric Winsberg describes parameterization in climate modeling as a process that involves using modeling parameters that are not physical since “no such value exists in nature” but are instead “artifacts of the computation scheme.” (Winsberg 2018, 49). The second reason—making models more general via abstraction—means that there are simply far fewer parameters within the model that can be compared with experimental data from particular systems. Despite these limitations on being comparable to experimental data, many modelers aim to construct generally applicable models by using highly general mathematical frameworks and abstracting away from the features of particular systems. In other words, when confronted with the experiment-applicability tradeoff, other modelers sacrifice the ability for their parameters to be directly comparable with experimental data so as to maximize the general applicability of their models.

One of the clearest examples of this response to the experiment-applicability tradeoff is R.A. Fisher’s statistical modeling of biological populations. Instead of measuring fitness within real populations to fill in specific parameters of his statistical models, Fisher abstracts away from any details about the fitness of particular traits in any specific population. For Fisher, fitness is instead represented by his *Malthusian parameter,* which is just the per capita rate of increase and decrease for a particular trait in the population. Fisher’s use of a Malthusian parameter for the fitness of a trait type enabled him to construct a model that is widely applicable to evolving biological populations. Indeed, Fisher uses this parameter to derive his Fundamental Theorem of Natural Selection. However, Fisher’s particular mathematical operationalization of fitness in terms of the Malthusian parameter is a rather *contrived* conception of fitness that is quite different from how philosophers and biologists conceptualize the term. However, using this contrived definition of fitness is essential to Fisher’s model since, “The rigor of the demonstration requires that the terms employed should be used strictly as defined” (Fisher 1930, 35).

Rather than being *representationally accurate* of lots of real (or possible) target systems, Fisher is well aware that “the theorem is exact only for idealized populations, in which fortuitous fluctuations in genetic composition have been excluded” (Fisher 1930, 35). That is, Fisher’s aim is not to build a model that is generally *true* or *accurate* with respect to the features of real biological populations. Nonetheless, the fundamental theorem of natural selection and its Malthusian parameter is intended to be widely *applicable.* As Fisher notes, “The definitions given above may be applied to any characteristic whatever; it is of special interest to apply them to the special characteristic *m* which measures the relative rate of increase and decrease.” (Fisher 1930, 34). Moreover, this fundamental theorem, “bears some remarkable resemblances to the second law of thermodynamics. Both are properties of populations, or aggregates, true *irrespective of the nature of the units which compose them.*” (Fisher 1930, 36, my emphasis). One key difference, however, is that “fitness, although measured by a uniform method, is qualitatively different for every different organism, whereas entropy, like temperature, is taken to have the same meaning for all physical systems” (Fisher 1930, 37). That is, despite Fisher’s operationalizing fitness in terms of the Malthusian parameter in a way that is generally applicable, he acknowledges that the fitness of every organism will be different such that this universal applicability is not mirrored by being generally true of lots of different systems or organisms.

Fisher closes his chapter by noting that “the actuarial information necessary for the calculation of the genetic changes actually in progress in a population of organism, will *always be lacking*; if only because the number of different genotypes for each of which the Malthusian parameter is required will often, perhaps always, exceed the number of organism in the population in addition to the fact that this parameter is very imperfectly known” (Fisher 1930, 44). In other words, actually measuring fitnesses (or at least the Malthusian parameters) for any real population will always be impossible. Moreover, Fisher acknowledges that although we only need to determine the values for the parameters *W, D, C,* and *M* in the equation for population growth, “Our ignorance as to these is, of course, profound” (Fisher 1930, 44) thoughhe hopes that “direct observational methods may yet determine the numerical values which condition the survival and progress of particular species” (Fisher 1930, 47). What we see here is Fisher defending his modeling methodology by showing that it can employ generally applicable statistical techniques that can be *applied* to anybiological population—i.e. the model is widely applicable in many different modeling contexts. But we also see Fisher acknowledging that the parameters of his model will, perhaps always, be impossible to measure via experiments in particular systems. This, I contend, clearly illustrates that in some contexts scientific modelers sacrifice the ability to compare their models with experimental data so as to employ modeling techniques that are widely applicable to many different systems.

*4.3. Mesoscale modeling, stable parameters, and autonomy*

What the above cases show is that sometimes the measurable parameters are difficult to generalize for application to other systems and sometimes the generally applicable parameters are not measurable. Despite this, sometimes the modeling context is poised just right so that both of these pragmatic modeling aims can be increased simultaneously. In general, this occurs when the parameters that are applicable to a wide range of systems (typically due to some kind of multiple realizability or autonomy) are also the features of the system that the available experimental methods are able to measure/track.

 An example of such a modeling context can be found in several kinds of mesoscale modeling in both physics and biology (Batterman 2021; Nersessian 2022). For example, drawing on numerous examples, Robert Batterman presents what he calls an ‘engineering approach’ to modeling many-body systems that is explicitly a ‘middle-out’ modeling strategy. In very rough outline, the strategy is to identify important mesoscale material parameters of the system that track correlations among groups of particles and enable bridges between features at both smaller and larger scales. One important advantage of this approach compared with modeling the system in terms of its smallest scale features is that the mesoscale theory/model “can be connected to fundamental theory at one end, and at the other it is applied directly to the experimental data” (Schwinger 1969, 19). The availability of this experimental data concerning material parameters is due to the stability of these mesoscale correlations within real-world systems. In particular, although many of the smaller scale quantities in the system decay too quickly to be measured (Batterman 2021, 65), these mesoscale material parameters are stable enough (over time) to be measurable. Therefore, by focusing on these mesoscale features of the system, physicists (and engineers) are able to build models whose key parameters “often are experimentally observable through various light, X-ray, and neutron scattering experiments. Thus these quantities provide a rather direct connection with the measurements we can actually perform on many-body systems (Batterman 2021, 66). As a result, by focusing on these mesoscale parameters, scientists can construct models whose parameters are directly comparable with the available experimental data. This enables one side of the experiment-applicability tradeoff to be increased fairly easily in this context.

The mesoscale parameters used in these models are also generally applicable to multiple systems because they are *autonomous* of most of the system’s features at smaller scales. As Batterman puts it “order parameters allow one to say a considerable amount about the macroscopic nature of many-body systems without having to have detailed knowledge about the nature of the system at the most fundamental scale” (2021, 109). This autonomy allows for the same models to be applied across systems whose features at those fundamental scales are quite different. That is, this autonomy allows mesoscale models that focus on material parameters to be widely applicable across many different modeling contexts. Consequently, in some modeling contexts the widely applicable parameters are *also* the parameters for which detailed experimental data is possible. When this occurs, both sides of the experiment-applicability tradeoff can be simultaneously increased.

It is worth noting that this ability to optimize both sides of the experiment-applicability tradeoff is due to features of the target system and the available measurement tools and modeling techniques. The stability of the properties of the system that one would like to measure is essential to their being accessible for experimentation. Similarly, the autonomy of certain features from other features of the system is essential for generalizing the model across multiple contexts. In addition to being stable in these two senses, there needs to exist experimental methods for making precise measurements of the properties represented by the model’s parameters. Finally, there needs to be a modeling technique/framework that focuses on precisely those parameters that are both comparable with experiment and autonomous of most of the other features of particular systems. These are, I contend, all features of the modeling context that go beyond the semantical/representational properties of models themselves. This is one reason why Schwinger describes this as “a pragmatic approach” (1969, 22) and Batterman says this engineering approach “appears manifestly pragmatic” (2021, 104).[[17]](#footnote-17) Consequently, we see how, across all three cases, the experiment-applicability tradeoff is a pragmatic tradeoff that is highly context sensitive and depends on the available experimental data and modeling techniques, the features of the real-world target systems, and the purposes and skills of the scientific modeler. This also serves to further highlight the ways that the experiment-applicability tradeoff depends on the context rather than being a necessary relationship of attenuation across all contexts.

**5. Implications for Other Debates in the Philosophy of Science**

In this final section, I look at some of the implications of the experiment-applicability tradeoff for other debates in the philosophy of science and discuss future directions for investigating the tradeoffs involved in scientific model building.

*5.1 Theories as collections of models that navigate tradeoffs in different ways*

Several philosophers have suggested that theories ought to be construed as collections of models (Suppe 1977; Suppes 1960; Giere 1988). Moreover, Levins notes, “a satisfactory theory is usually a cluster of models” (1966, 431). However, Levins focuses on various ways these models are related via robustness analysis, hierarchical nesting, or aggregated parameters (Levins 1966, 431). More generally, Levins tells us that “the multiplicity of models is imposed by the contradictory demands of a complex heterogeneous nature and a mind that can only cope with a few variables at a time; by the contradictory desiderata of generality, realism and precision.” (1966, 431). That is, clusters of models are constructed to best manage contradictory modeling aims in the face of overwhelming complexity. Relatedly, Weisberg argues that the proliferation of multiple conflicting models is due to Levins’s tradeoffs and that “if a theorist wants to achieve high degrees of generality, accuracy, precision and simplicity, she will need to construct multiple models” (2007, 646-47).

I suggest that what is missing from this picture are alternative reasons for why theories often are constituted by a collection of models. Specifically, in addition to being motivated by tradeoffs between representational ideals, scientists often build multiple models of the same phenomenon because some of those models are theoretically generalizable to a wide range of situations, others are directly comparable with the available experiments and measurements, and still others are constructed because they focus on parameters that are both comparable with the available experimental data and widely applicable. That is, rather than being driven solely by Levins’s tradeoffs between generality, realism and precision, scientists are also motivated to construct collections of models because different models will navigate the experiment-applicability tradeoff in different ways.

More generally, once we recognize that there are numerous pragmatic modeling tradeoffs in addition to the tradeoffs that hold between representational ideals, we see that that are additional reasons why scientists might build multiple models for the same phenomenon. Indeed, since different modelers will navigate this network of modeling tradeoffs in different ways, we should expect communities of scientists to produce clusters of models that are general, precise, realistic, comparable with experiments, widely applicable, understandable, etc. to different degrees. Since no single model can provide all these features (or virtues), different modeling choices about how to navigate modeling tradeoffs will result in an even wider array of models tailored to different purposes or aims. The literature surrounding Levins’s paper has identified some of the reasons that scientists build multiple models. Suppes (1960) and the literature surrounding the hierarchy of models that mediate between theories and data identifies additional reasons. This paper adds yet another reason: building collections of models enables scientists to respond to the experiment-applicability tradeoff in a variety of ways. As a result, scientists theoretical understanding of a (type of) phenomenon will often be derived from the construction, application, and use of a cluster of models that navigate various modeling aims and tradeoffs in different ways.

*5.2 Models as mediators between theory (or widely applicable models) and reality (or data)*

Another prominent view in the modeling literature suggests that models mediate between theory and reality. For example, on Mary Morgan and Margaret Morrison’s (1999) influential view, models are partially autonomous from both theory and data and so can mediate between them. We can see this idea directly reflected in Schwinger’s observation that his engineering approach “can be connected to fundamental theory at one end, and at the other it is applied directly to the experimental data” (1969, 19).

I contend that the experiment-applicability tradeoff—and the variety of ways of navigating it—helps us make sense of different ways that models mediate between generally applicable theories and experimental data. For one thing, the tradeoff clarifies precisely *why* individual models often mediate between general theoretical parameters and experiments performed within particular contexts. Specifically, because scientific modelers are often trying to balance between competing modeling aims of generally applicability and connection with experimental results acquired from particular systems, many models often end up playing mediating roles in order to balance these two modeling goals. That is, being partially autonomous of theory and data so as to mediate between the two is one way of trying to balance the competing modeling aims involved in the experiment-applicability tradeoff. Indeed, rather than constructing individual models that maximize one side of the tradeoff or the other, sometimes modelers will construct mediating models that balance those two aims (without maximizing either) by having the model mediate between theories and experimental data.

While individual models can sometimes strike this balance, often it is collections of models that navigate the tradeoff differently that enable for the connection of generally applicable models and other models that are more directly connected with experimental data. That is, rather than individual models mediating between (distinct) theories and experimental data, I suggest that models often mediate relationships between other models that navigate the experiment-applicability tradeoff differently. Specifically, some models will be theoretically generalizable, some will be directly comparable with experiments performed in particular systems, and some will be able to partially accomplish both. For example, Tarja Knuuttila and Andrea Loettgers (2018) provide an interesting discussion of how what they call ‘synthetic models’ in biology are often used to mediate between generally applicable mathematical models and models that more directly connect with the experimental data. This highlights how many models end up mediating relationships between generally applicable theoretical models and other models that are more directly comparable with the available experimental data (Suppes 1962).[[18]](#footnote-18) In short, it is often because it is difficult to construct models that are both widely applicable and empirically grounded that scientists often use clusters of models to mediate between widely applicable models (that maximize one aim) and models of data (that maximize another).

 Another important claim of the models as mediators view is that scientists often fill in more abstract models (or theories) in various ways to apply them to real physical systems. Suppes (1962), Cartwright (1983), and Morgan and Morrison (1999) all suggest that models need to be filled in, or made more concrete, for them to be applicable to real systems. While this is certainly an important part of applying a model to real systems, the experiment-applicability tradeoff suggests that it is only part of the story. In some cases, application involves bringing a model into contact with experimental data drawn from real systems. Yet, as we saw in the R.A. Fisher case, some modelers maximize applicability by using mathematical parameters that are difficult or impossible to measure in real systems. In short, application of a model is not always achieved by filling in a more theoretical model to make it more realistic, concrete, or connected to empirical data.

*5.3 The available experiments and measurements play essential roles in navigating modeling tradeoffs*

Another interesting implication of the above discussion is that the literatures on scientific experimentation and scientific modeling have much to offer each other (Morgan 2005; Parker 2009; Parke 2014; Winsberg 2009). After all, how to best navigate these modeling tradeoffs depends on which experiments are available, what they can tell us, and how they relate to theoretical parameters within scientific models. Conversely, which parameters scientists aim to compare with experiments and how they perform those experiments depends on the models that scientists have available, their aims/purposes for those models, and the parameters/variables they choose to include in those models (Morgan and Morrison 2009; Rice 2024; Woodward 2016). Therefore, which models are built often depends on which experiments can be performed, and which experiments are performed often depends on what models scientists choose to construct. More generally, as Jack Justus correctly notes:

To study the biological world, ecologists must do two things. One, they must conduct the fieldwork – experimental manipulation, monitoring, sampling, surveys, and so on – that furnishes the raw observational information at the base of all scientific knowledge… But the insights that can be gleaned from this data alone are limited. The second thing ecologists must do, to harness data’s full epistemic power, is to create a model, a representation of the portion of the biological world of interest. Only with such a representation can observational information be integrated and systematized into a more comprehensive and detailed account of how the biological world works. Perhaps it is not as epistemically foundational, but modeling is just as indispensable as observation to achieving significant ecological insights (2021, 74).

In short, biological modeling—as well as modeling in other sciences—must balance the needs of model construction with the constraints involved in collecting observational and experimental data. As a result, the literatures on scientific modeling and scientific experimentation have much to offer each other (Peschard and van Fraassen 2018; Winsberg 2003). While there is a large literature on whether/how to distinguish the epistemic roles of simulations from those of experiments (Morgan 2005; Morrison 2009; Parker 2009; Parke 2014; Winsberg 2003, 2009), this literature has not focused on building, testing, and applying models and simulations *from* experiments and how the simulations and models being used influence *which* experiments are performed to test those models and simulations. Moreover, the focus of that literature has been primarily on whether simulations *are* experiments, but there is little discussion of *the tradeoffs between* the various pragmatic aims of model construction and experimentation.Thus, while investigating the similarities and differences between the epistemic roles of experiments and simulations is certainly a worthwhile philosophical project, philosophers should also continue to investigate the ways in which scientific model-building and experimentation *constrain one another* in different contexts via tradeoffs between multiple aims of scientific practice.

Similarly, while some philosophers of science have focused on evaluating models with respect to their specific purposes in different cases (Parker 2020), these discussions have rarely addressed tradeoffs between various pragmatic aims within those modeling contexts. For example, while Parker correctly mentions that certain practical/pragmatic constraints play a role when we consider how long it takes to run a computer simulation (2020, 465) or how well the model’s parameters compare with available data (2020, 468), she offers no further discussion of how these different pragmatic aims of scientific modelers might tradeoff with one another across different contexts. I suggest that the experiment-applicability tradeoff is one fruitful place to continue weaving some of these philosophical literatures together to better capture the ways that the practice of model building and experimental data collection constrain one another in scientific practice. Indeed, these investigations are likely to reveal several additional tradeoffs between various pragmatic, epistemic, ethical, and other aims of science. Doing so will further contribute to developing several philosophers’ suggestion that representational aims/features of the “model-target relationship [are] only part of the story when it comes to achieving scientific aims” (Parker 2020, 465).

*5.4 Balancing a multitude of modeling aims*

A final implication of the above view is that how to best navigate modeling tradeoffs depends on more than just the representational goals of scientific modelers. Indeed, as we have seen from investigating the experiment-applicability tradeoff, modeling choices are often *highly constrained* by the pragmatic and ontological features of the modeling context. As a result, scientific modelers must balance a number of things that go beyond their representational ideals/aims. These include, but are not exhausted by, the following (rough) types of questions[[19]](#footnote-19):

*Pragmatic questions:*

What experiments are feasible?

What experimental data already exists concerning this type of system?

What modeling (or computational) frameworks and resources are available?

*Ontological questions:*

Which parameters are stable enough to be targets for experimental measurement?

Which parameters display enough autonomy to be applicable across different target systems?

*Epistemological questions:*

Does the model aim to produce knowledge or to be explanatory?

Does the model aim to provide an explanation or to produce understanding?

*Representational questions:*

Which features of the target system is the model intended to accurately represent (and to what degree)?

Which features of the target system does the model distort via idealization or abstraction?

These various aims of modeling practice will often tradeoff with one another in myriad ways that constrain the practice of scientific modeling. However, as I argued above, the literature on tradeoffs in scientific modeling has almost exclusively focused on formally defined representational aims/goals without investigating other kinds of tradeoffs that constrain those modeling decisions. Moreover, while numerous philosophers have investigated these other aims of scientific modeling, there has been no sustained investigation of the tradeoffs that hold between (or across) those various aims and how scientific modelers ought to navigate them in different modeling contexts. Rather than trying to clearly demarcate categories of modeling aims/questions, the point is that there are far more tradeoffs and constraints involved in scientific modeling beyond those that hold between various representational/semantic aims. For example, there are likely tradeoffs between making precise predictions with a model and the time and computational resources required to make those predictions. What is more, there will likely be complex tradeoffs that cut across these categories of modeling aims. For example, how well a model satisfies certain pragmatic aims of modelers might tradeoff with certain epistemic or semantic aims. We can see this clearly in cases where, say, an extremely complex and highly idealized computational model is highly successful at making certain kinds of predictions for some practical purpose; e.g. in climate modeling. The point is not that pragmatic, ontological, and epistemological aims of models will be easy to disentangle (or clearly demarcate), but rather that the structure of tradeoffs in scientific modeling is a much richer and interwoven landscape than is suggested by narrowly focusing only on models’ semantic/representational properties. Going forward, philosophers of science ought to further investigate this much wider and complex mosaic of modeling tradeoffs and the ways they require scientists to balance a variety of considerations beyond just which features to accurately or inaccurately represent in their models.

**6. Conclusion**

This paper has argued that philosophical debates concerning the tradeoffs of scientific model building ought to be expanded to include various pragmatic features of modeling concerning the available experimental data and generally applicable modeling frameworks. While the literature on scientific modeling and idealization has fruitfully drawn attention to the representational choices involved in scientific modeling, philosophers’ discussions of representational ideals miss many of the pragmatic features of model construction and use. Specifically, I have argued that scientific modelers must often confront a tradeoff between having their model be comparable with the available experimental data and having their model be generally applicable across lots of modeling contexts. Rather than a universal or necessary limit on model properties, how this tradeoff is realized and navigated in scientific practice is highly context dependent. Going forward, I suggest that philosophers of science investigate other types of modeling tradeoffs and the different ways they are navigated in different contexts of scientific practice. Doing so will improve our understanding of how scientific modelers balance the relationship between the available modeling frameworks and experimental data, how limitations on modeling resources impact the construction of scientific models, and how scientific modelers balance a multitude of different modeling aims.

**References**

Batterman, R. (2021). *A Middle Way; A Non-Fundamental Approach to Many-Body Physics.* Oxford: Oxford University Press.

Elgin, C. (2007). Understanding and the facts. *Philosophical Studies 132*(1): 33-42.

Elliot-Graves, A. (2022). What are general models about? *European Journal for Philosophy of Science 12*: 74.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* London: Oxford University Press.

Giere, R. (1988). *Explaining Science: A Cognitive Approach.* Chicago: University of Chicago Press.

Justus, J. (2021). *Philosophy of Ecology.* Cambridge: Cambridge University Press.

Knuuttila, T. and Loettgers, A. (2018). Modeling/Experimentation: The Synthetic Strategy in the Study of Genetic Curcuits. In *The Experimental Side of Modeling* (eds. Peschard, I. F. and van Fraassen, B. C.). Minneapolis: University of Minnesota Press, pp. 118-147.

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science 9*: 22.

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

Matthewson, J. (2011). Tradeoffs in model-building: A more target-oriented approach. *Studies in History and Philosophy of Science: Part A 42*(2): 324-333.

Matthewson, J. and Weisberg, M. (2009). The structure of tradeoffs in model building. *Synthese 170*: 169-190.

Morgan, M. S., Morrison, M. (1999). *Models as Mediators: Perspectives on Natural and Social Sciences*. Cambridge: Cambridge University Press.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology 12*(2): 317-29.

Morrison, M. (2009). Models, measurements and computer simulation: The changing face of experimentation. *Philosophical Studies 143*(1): 33-57.

Nersessian, N. (2022). *Interdisciplinarity in the Making: Models and Methods in Frontier Science*. Cambridge: MIT Press.

Odenbaugh, J. (2003). Complex systems, trade-offs, and theoretical population biology: Richard Levins’s ‘strategy of model building in population biology’ revisited. *Philosophy of Science 70*: 1496-1507.

Odenbaugh, J. (2006). The strategy of ‘the strategy of model building in population biology’. *Biology and Philosophy 21*: 607-621.

Orzack, S. and Sober, E. (1993). A critical assessment of Levins’s ‘The strategy of model building (1966). *Quarterly Review of Biology 68*: 534-546.

Parke, E. C. (2014). Experiments, simulations, and epistemic privilege. *Philosophy of Science 81*(4): 516-536.

Parker, G. A. and Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature 348*: 27-33.

Parke, W. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese 169*(3): 483-96.

Parker, W. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science 87*, 457-477.

Peschard, I. R. and van Fraassesn, B.C. (2018). *The Experimental Side of Modeling*. Minneapolis: University of Minnesota Press.

Potochnik, A. (2017). *Idealization and the Aims of Science.* Chicago: Chicago University Press.

Pimm, S. (1991). *The Balance of Nature.* Chicago: University of Chicago Press.

Rice, C. (2021). *Leveraging Distortions: Explanation, Idealization and Universality in Science*. Cambridge: MIT Press.

Rice, C. (2024). A model-based approach to the problem of variable choice. *Philosophy of Science.* doi: 10.1017/psa.2024.52

Rohwer, Y. and Marris, E. (2021). Ecosystem integrity is neither real nor valuable. *Conservation Science and Practice 3*: e411.

Schwinger, J. (1969). Julian Schwinger’s engineering approach to particle theory. *Scientific Research 4*(17): 19-24.

Sober, E. (2000). *Philosophy of Biology*. Boulder, CO: Westview Press.

Suppe, F. (1977). *The Structure of Scientific Theories.* Chicago: University of Illinois Press.

Suppes, P. (1960). A comparison of the meaning and use of models in mathematics and the empirical sciences. *Synthese 12*: 287-300.

Suppes, P. (1962). Models of data. In *Logic, Methodology, and Philosphy of Science: Proceedings of the 1960 International Congress.* Eds. E. Nagel, P. Suppes, and A. Tarski. Pp. 252-261. Stanford, CA: Stanford University Press.

Walker, D. C. and Southgate, J. (2009). The virtual sell—a candidate co-ordinator for ‘middle-out’ modelling of biological systems. *Briefings of Bioinformatics*, *10*(4): 450-461.

Walmsley, L. D. (2021). The strategy of model building in climate science. *Synthese 199*: 745-765.

Watt, K. E. F. (1956). The choice and solution of mathematical models for predicting and maximizing the yield of a fishery. *Journal of Fisheries Board of Canada 13*(5): 613-645.

Weisberg, M. (2006). Forty years of the strategy: Levins on model building and idealization. *Biology and Philosophy 21*: 623-645.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy 104*(12): 639-659.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World.* New York: Oxford University Press.

Winsberg, E. (2003). Simulated experiments: Methodology for a virtual world. *Philosophy of Science*, *70*: 105-125.

Winsberg, E. (2009). A tale of two methods. *Synthese 169*(3): 575-92.

Winsberg, E. (2018). *Philosophy and Climate Science*. Cambridge: Cambridge University Press.

Wilson, K. (1994). Evolution of clutch size in insects. II. A test of static optimality models using the beetle *Callosobruchus maculatus. Journal of Evolutionary Biology 7*: 365-386.

Woodward, J. (2016). The problem of variable choice. *Synthese 193*: 1047-72.

1. This paper is not intended to replace the existing literature surrounding Levins’s paper, but rather to expand/supplement that literature in fruitful ways by identifying tradeoffs and features of modeling that are missed by that discussion’s focus on representational aims or properties of models. [↑](#footnote-ref-1)
2. While Rice (2024) discusses this aim of scientific modeling and some other pragmatic constraints on scientific modeling, that paper does not address the *tradeoff* *between* the pragmatic aims of comparability with experimental data and general application that are my focus here. More generally, while several philosophers have discussed these aims of scientific modeling separately (Parker 2020; Suppes 1962), the tradeoffs between comparing a model with experimental data and the ability for that model to be generally applicable has not been explicitly analyzed and contrasted with the “great deal of philosophical discussion of scientific modeling [that] has focused on when models represent, or successfully represent, their targets” (Parker 2020, 465). [↑](#footnote-ref-2)
3. Of course, one could draw additional distinctions between the various aims of modelers and note that many modeling aims are dependent on each other in complex ways. All that is necessary for my argument here is that there is a distinction between the pragmatic aims of building models for specific purposes and the semantic aims of building models that are accurate representations of certain features of their target systems. [↑](#footnote-ref-3)
4. Rice (2024, 2) does discuss a tradeoff between having criteria for variable choice balance between “(1) being able to offer specific guidelines for practicing scientific modelers, and (2) being as generally applicable as possible across scientific modeling contexts.” However, this is a tradeoff that confronts *philosophers’* attempts to offer useful but widely applicable accounts *of* scientific modeling. It is not a tradeoff between the pragmatic aims that arise within the practice of scientific modeling itself. As such, this paper addresses a very different tradeoff between pragmatic constraints that confront practicing scientific modelers rather than a tradeoff concerning the aims of philosophical theorizing about science. [↑](#footnote-ref-4)
5. After proposing these three options, Levins’s discussion moves on to present his ideas concerning robustness analysis, which will not be my focus here. [↑](#footnote-ref-5)
6. Strikingly, none of these strategies regarding representational aims of models directly address Levins’s original concerns about being able to measure the parameters of the model or compare it with the available measurements. Despite this, some of Levins’s *examples* of these strategies are discussed in terms of comparability with experiment or general applicability. I’ll return to these points later on after arguing that comparability with experiment is importantly different from precision or realism. [↑](#footnote-ref-6)
7. Rice (2024) refers to this as the ‘tyranny of availability’ that scientific modelers must face concerning limited modeling resources. [↑](#footnote-ref-7)
8. Again, this is meant to offer a useful *expansion* of the discussion of tradeoffs in model-building rather than a *replacement* of the existing discussion of Levins’s tradeoffs. [↑](#footnote-ref-8)
9. This aim of modeling practice mirrors one of Levins’s complaints about the brute force approach: that there are too many parameters to measure. [↑](#footnote-ref-9)
10. Thanks to an anonymous reviewer for suggesting this potential connection with Suppes’ view and the suggestion that these two aspects of his view might be separated. [↑](#footnote-ref-10)
11. As I note later, this has some parallels with the ‘models as mediators’ view’s suggestion that models connect theories with physical systems by specifying the theory more concretely. [↑](#footnote-ref-11)
12. Another question here is how to individuate models in order to determine when one model (M’) that is comparable with experiments is just a more/less specific version of another model (M) and when they are two distinct models. However, as I noted above, I think the experiment-applicability tradeoff applies in either case though there are different possible responses to the tradeoff employed via the construction of individual or collections of models. [↑](#footnote-ref-12)
13. This, I take it, is Levins’s original point when he claims that we cannot build individual models that simultaneously maximize generality, precision and realism. [↑](#footnote-ref-13)
14. Thanks to an anonymous reviewer for raising this objection. [↑](#footnote-ref-14)
15. While semantic aspects might be involved in other ways (e.g. concerning what truths can be extracted from using the model), this shows that applicability is largely independent of whether the model is a true or accurate representation of a wide range of real (or possible) systems. [↑](#footnote-ref-15)
16. For example, the logistic growth equation was always intended to be a highly general model of biological growth (Elliot-Graves 2022). [↑](#footnote-ref-16)
17. Batterman goes on to argue that there are additional non-pragmatic reasons for adopting this approach. However, my point here is that these pragmatic aspects of the approach are different from appealing to the representational ideals or semantic properties of models themselves. [↑](#footnote-ref-17)
18. However, unlike Suppes, I do not think these models need to be more or less specific versions of each other in a set-theoretical sense. [↑](#footnote-ref-18)
19. I do not think there is much at stake here about exactly how to differentiate these broad categories of questions; i.e. the specific categories of questions are not essential to my argument/claim here. The important point is that there are multiple aims of modeling that exhibit tradeoffs beyond just those that hold between representational aims. [↑](#footnote-ref-19)