2019

Models Dont Decompose That Way: A Holistic View of Idealized Models

Collin Rice
Bryn Mawr College, crice3@brynmawr.edu

Citation
Rice, Collin. 2019. "Models Don’t Decompose That Way: A Holistic View of Idealized Models." British Journal for the Philosophy of Science 70.1: 179-208.

Models Don't Decompose That Way: A Holistic View of Idealized Models

Collin Rice

Abstract

Many (if not most) accounts of scientific modeling assume that models can be decomposed into the contributions made by their accurate and inaccurate parts. These accounts then argue that the inaccurate parts of the model can be justified by distorting only what is irrelevant. In this paper, I argue that this decompositional strategy requires three assumptions that are not typically met by our best scientific models. In response, I propose an alternative view in which idealized models are characterized as holistically distorted representations that are justified by allowing for the application of various (mathematical) modeling techniques.

1. Introduction

Many (if not most) accounts of scientific modeling—across a range of debates—assume that models can be decomposed into the contributions made by their accurate and inaccurate parts.[1] These accounts also assume that real-world systems can be decomposed into the contributions made by the features that are relevant and irrelevant to the target phenomenon. It is then typically claimed that the accurate parts of a successful[2] scientific model map onto what is relevant (or

[1] A notable exception are fictionalist accounts of models (Frigg [2010]; Godfrey-Smith [2009]). However, since these fictionalist accounts focus on the ontology of models rather than on how models explain, how idealizations contribution to explanations, robustness analysis, realism, etc. they will not be my focus here. Moreover, I think the problems of the decompositional strategy can be solved without the additional metaphysical commitments to hypothetical systems that are involved with most fictionalist accounts of what models are.

[2] In most cases "successful" will mean explanatory. However, since some of the accounts that fall under the decompositional strategy are not explicitly focused on explanation it would be misleading to include that as a requirement of being successful. The point here is that the models that best achieve the goals of scientific inquiry will be those that meet these three assumptions.

important) and the inaccurate parts of the model only distort what is irrelevant (or unimportant). This allows one to argue that the accurate parts of the model are what “do the real work” while the inaccurate parts of the model are justified by distorting only what is irrelevant. This *decompositional strategy* is central to most (or at least a wide range) of our current accounts of how to model complex systems, how models explain, how idealizations contribute to model explanations, robustness analysis, and how idealized modeling is compatible with scientific realism.

In this paper, I argue that many of our best scientific models cannot be decomposed in the ways required by the decompositional strategy. To be clear, I will not be arguing that decomposition is a completely wrong-headed strategy—such decomposition is epistemically convenient when it obtains. However, I will contend that the decompositional strategy cannot be the whole story since it requires three assumptions that are not typically met by our best scientific models. As a result, decompositional accounts of modeling, explanation, idealization, robustness, realism, etc. need to be supplemented, or perhaps supplanted, by an alternative approach.

Therefore, after arguing against the decompositional strategy, I will argue for what I call the *holistic distortion view* of how idealizations contribute to scientific models. According to this view, most idealized models in science are holistically (i.e. pervasively) distorted representations of their target systems (Rice [forthcoming]). In other words, idealized models in science are typically greater than the sum of their accurate and inaccurate parts because they result from the complex interaction of various modeling assumptions and idealizations that produce a pervasive misrepresentation of most of the features of their target system(s). The use of these holistically distorted representations is justified, I will argue, because they allow for the application of (mathematical) modeling techniques that provide epistemic access to the kinds of information scientists are interested in. Finally, since this holistic distortion view moves us away from focusing on accurate representation relations, I will argue that a promising alternative is to appeal to universality classes in order to link idealized models with their target system(s) in ways that allow for explanation and understanding (Batterman and Rice [2014]; Rice [forthcoming]). Universality can enable scientific modelers to discover relationships of counterfactual dependence between certain key features and an explanandum even if those relationships hold for drastically different reasons in the idealized model system than they do in the real-world system(s). Therefore, even if the model pervasively distorts its target system(s), it may still be possible to use the model to explain because many of the patterns of counterfactual dependence displayed by the model system will be similar to those of the model's target system(s).

The following section presents several forms of the decompositional strategy that have been proposed across the philosophy of science literature and extracts three key assumptions required for the decompositional strategy to work. Then, Section 3 argues that many of our best scientific models cannot be decomposed in the ways required by those assumptions. In response, Section 4 proposes the holistic distortion view of idealized models. The final section concludes and offers some suggestions for how to develop the holistic distortion view further.

2. The Decompositional Strategy

In this section, I argue that a wide range of accounts, across numerous philosophical debates, can all be grouped under a general decompositional strategy that is committed to some version of the following three assumptions:

(1) *Target Decomposition Assumption:* The real-world system is decomposable such that the contributions of the features that are relevant to (e.g. difference-makers for) the occurrence of the target phenomenon can be isolated from the contributions of features that are irrelevant (or are largely insignificant) to the target phenomenon.[3]

(2) *Model Decomposition Assumption:* The scientific model is decomposable such that the contributions of its accurate parts can be isolated from the contributions of its inaccurate (i.e. idealized or abstracted) parts.

(3) *Mapping Assumption:* When successful, the accurate parts of the model can be mapped onto the relevant parts of the real-world system and the inaccurate parts of the model only distort the irrelevant parts of the real-world system.[4]

While the following accounts are committed to these assumptions to different degrees and for different reasons, I think the discussion shows that they are all equally committed to a general decompositional strategy that requires some form of these three assumptions.[5]

2.1. Mechanistic modeling and decomposition

[3] One reviewer suggested a distinction between a feature's being irrelevant and a feature's being relatively unimportant. While this distinction can certainly be made, I don't think this subtle difference in degree has much of an impact on the arguments that follow. Nonetheless, to avoid any confusion, my use of irrelevance (and relevance) here is meant to capture claims concerning causal irrelevance (e.g. difference making) and relative unimportance (e.g. having a negligible impact). The important point is that the target system is supposed to be decomposed into a set of relevant (or significant) and irrelevant (or insignificant) features—even though different versions of the decompositional strategy will determine what is relevant and irrelevant in slightly different ways.

[4] It is worth noting that Weisberg's [2013] similarity account seems to have some resources for resisting this particular version of the mapping assumption since his account allows the relevant features to be included in the model and yet be distorted in various ways that make them less similar than the target system. However, in what follows I focus on Weisberg's account of minimalist idealization in his [2007] paper and in Chapter 6 of his [2013] book, which is more directly committed to the mapping assumption as I have formulated it here.

[5] Given the range of views, from a number of debates, it is impossible to find the perfect wording that captures all the subtle details of every version of the general decompositional strategy. I think putting these assumptions in terms of relevant/irrelevant and accurate/inaccurate does the best job of capturing what these views have in common. Indeed, however these accounts determine relevance/irrelevance, the claim is that the model will be successful/explanatory just in case it accurately (to some degree) captures the relevant (or most important) features and only uses idealizations to distort irrelevant (or largely insignificant) features. Each account shares this general approach—although they may establish relevance/irrelevance and accuracy/inaccuracy in slightly different ways.

Many forms of decomposition have their roots in the mechanistic approach to modeling complex systems (Bechtel and Richardson [1993]; Craver [2006], [2007]; Craver and Darden [2013]; Machamer, Darden, and Craver [2000]). In general, the mechanistic approach focuses on decomposing systems into their constituent parts and localizing their characteristic activities in order to identify the modular (i.e. dissociable) contributions made by the parts of complex systems (Bechtel and Richardson [1993], p. 7). However, as Bechtel and Richardson repeatedly emphasize:

> Pursuing decomposition and localization is to impose an assumption on the nature of the system whose activities one is trying to explain: it is assuming that it is *decomposable*. A decomposable system is modular in character, with each component operating primarily according to its own intrinsically determined principles. (Bechtel and Richardson [1993], pp. 24-25)

Indeed, the mechanistic approach to modeling assumes that the target system is decomposable into parts that each make a modular (i.e. dissociable) causal contribution to the functioning of the whole (Bechtel and Richardson [1993], p. 24; Levy [2014], p. 5; Woodward [2003], p. 48). It is then assumed that the contributions of the set of relevant parts and interactions of the target mechanism(s) can be segregated (i.e. isolated) from the rest of the surrounding environment and studied independently (Bechtel and Richardson [1993]).

In addition, most accounts of mechanistic modeling assume that successful models (e.g. those that can explain) are those models whose components and interactions *accurately represent* the relevant (i.e. difference-making) causal relationships among the components of the target mechanism(s) and leave out—i.e. abstract away—irrelevant features.[6] In fact, the literature on mechanisms rarely addresses the relationship between mechanistic models and mechanisms themselves since it is routinely assumed that 'the parts and organization of the model typically

[6] It is unclear exactly what role idealizations play in mechanistic modeling accounts. While Kaplan and Craver [2011] note that idealization is compatible with their view, they provide no additional details about the role idealizations play within mechanistic modeling.

map directly to parts and organization of the object being investigated' (Matthewson and Calcott [2011], p. 738). However, some mechanistic views do make it explicit that, 'the goal is to describe correctly enough (to model or mirror more or less accurately) the relevant aspects of the mechanisms under investigation' (Craver and Darden [2013], p. 94).[7] Indeed, according to mechanistic accounts, for a model to explain is just for it to accurately describe the relevant parts of the target mechanism(s). Kaplan and Craver make this idea explicit in their endorsement of what they call a model-to-mechanism-mapping (3M) requirement in which mechanistic models explain in virtue of having their components, activities, properties, and organizational features map onto the relevant causal relations among the components of the target mechanism(s) (Kaplan and Craver [2011], p. 611). In sum, successful mechanistic models will accurately represent the relevant components and interactions of the target mechanism and will leave out irrelevant features.

2.2. Accounts of scientific explanation that require decomposition

The decompositional strategy is also prominent in several causal accounts of how models are able to explain. Indeed, most causal accounts of explanation require a model that explains to accurately represent (or describe) the difference-making causal relationships that produced the explanandum and isolate those factors from other irrelevant causes (Strevens [2009]; Woodward [2003]).

For example, Michael Strevens tells us that, 'no causal account of explanation—certainly not the kairetic account—allows non-veridical models to explain' (Strevens [2009], p. 320). Indeed, on Strevens's kairetic account, 'A standalone explanation of an event *e* is a causal model

[7] See also Glennan [2005].

for *e* containing only difference-makers for *e*' in which, 'the derivation of *e,* mirrors a part of the causal process by which *e* was produced' (Strevens [2009], pp. 73-75). In other words, models that explain do so in virtue of accurately representing an isolable set of difference-making causal factors and leaving out (or idealizing away) other irrelevant causes.

In addition, James Woodward suggests that causal models explain when they 'correctly describe', 'trace or mirror', or are 'true or approximately so' with respect to the relevant causal relations that hold between the explanans and the explanandum (Woodward [2003], pp. 201-203). Furthermore, in a later paper, Woodward suggests that, 'good explanations should both *include* information about all factors which are such that changes in them are associated with some change in the *explanandum-outcome* of interest and *not include* factors such that no changes in them are associated with changes in the *explanandum-outcome*' (Woodward [2010], p. 291). In other words, good explanations should accurately represent difference-making causal factors and abstract away irrelevant causal factors.

More generally, these causal accounts require that the target system can be decomposed into difference-making and non-difference-making causes. They also require that models can be used to isolate difference-making causes from other irrelevant factors. Finally, these accounts claim that models explain in virtue of accurately representing the isolable set of difference-making causes that actually produced the target explanandum (and leaving out irrelevant factors).

2.3. Accounts of idealization that require decomposition

In addition, the decompositional strategy is central to many accounts of how *idealizations* contribute to models that explain (Elgin and Sober [2002]; Strevens [2009]; Weisberg [2007],

[2013]). As a first example, Mehmet Elgin and Elliott Sober [2002] argue that optimality models in evolutionary biology can still explain despite being idealized because the idealizations only distort features that are irrelevant or make little difference to the explanandum. They say:

> A causal model contains an idealization when it correctly describes some of the causal factors at work, but falsely assumes that other factors that affect the outcome are absent. The idealizations in a causal model are *harmless* if correcting them wouldn't make much difference in the predicted value of the effect variable. Harmless idealizations can be explanatory (Elgin and Sober [2002], p. 448).

In other words, according to Elgin and Sober, idealized models can still explain because they only distort features that are irrelevant or largely insignificant. We can see that the distorted features are largely insignificant by noting that removing the idealizations from the model—or replacing them with true assumptions—would not make much difference to the predictions made by the model. Specifically, according to Elgin and Sober, idealized optimality models explain when they correctly describe the (difference-making) role of natural selection in bringing about the explanandum and only idealize other evolutionary factors that are assumed to be irrelevant to the outcome (e.g. drift).

In a similar way, Michael Weisberg describes what he calls *minimalist idealization* as, 'the practice of constructing and studying theoretical models that include only the core causal factors which gave rise to the phenomenon' (Weisberg [2007], p. 642). According to Weisberg, a minimalist model 'accurately captures the core causal factors' since, '[t]he key to explanation is a special set of explanatorily privileged causal factors. Minimalist idealization is what isolates these causes and thus plays a crucial role for explanation' (Weisberg [2007], pp. 643-5).

Among others, Weisberg cites Strevens's [2009] account of idealized models as an example of minimalist idealization.[8] Strevens explains his account of how idealized models can explain in this way:

> The content of an idealized model, then, can be divided into two parts. The first part contains the difference-makers for the explanatory target...The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target. The overlap between an idealized model and reality...is a standalone set of difference-makers for the target. (Strevens [2009], p. 318)

In other words, according to Weisberg and Strevens, idealized models explain by accurately representing, or "overlapping" with, an isolable set of causal difference-makers and using idealizations to indicate (or eliminate) those causal factors that are irrelevant.[9] Indeed, the general goal of these accounts is to show that, 'the causal factors distorted by idealized models are details that do not matter to the explanatory target—they are explanatory irrelevancies. The distortions of the idealized model are thus mitigated' (Strevens [2009], p. 340).

Before moving on, it is worth addressing some possible tweaks to this kind of decompositional approach to idealized models. For example, although Strevens [2009] and Weisberg [2007] suggest a rather strict version of the assumptions of the decompositional strategy, the view of similarity Weisberg defends in his [2013] book seems to allow for some distortion (in terms of dissimilarities) regarding difference-making features. Moreover, rather than claiming that a minimalist model must accurately represent *all* the difference-makers, Weisberg might only require the inclusion of the most important core causal factors. In a similar way, Angela Potochnik [2015] defends a causal view of explanation on which some causally relevant factors can be omitted as long as they do not distort the outcome too much. That is,

[8] Weisberg also includes Robert Batterman's work on asymptotic explanation, Stephan Hartmann's work on physical models, and Nancy Cartwright's account of abstraction as examples of minimalist idealization. I do not think each of these accounts fits Weisberg's minimalist category, but for considerations of space I will not work through those details here.
[9] Moreover, Strevens argues, 'All idealizations, I suggest, work in the same way' (Strevens [2009], p. 341).

Potochnik's account allows for some causally relevant factors to be neglected by models that explain if those factors are relatively unimportant to the causal pattern of interest within the current research program (Potochnik [2015], [forthcoming]). However, although Potochnik's account allows for some difference-making causes to be omitted, she also defends a criteria of explanatory adequacy such that 'a causal explanation must account for all of the significant causal influences on an event' (Potochnik [2015], p. 1179). Indeed, like Elgin and Sober's account, Potochnik's view requires that 'an explanation must take into account all the causal factors with a significant influence on the probability of the event's occurrence' (Potochnik [2015], p. 1178).[10] Moreover, in line with Weisberg [2007], Potochnik suggests that, 'the incorporation of idealizations into explanations is justified by the isolation of a casual pattern—Weisberg's 'core causal factors' (Potochnik [2015], p. 1173). This kind of view allows for some causally relevant factors to be left out of models that explain, but still requires explanations to isolate the causes that have a significant impact on the outcome.

Both these moves attempt to allow for some distortion (or omission) of causally relevant factors by making accurate representation of those factors and their explanatory relevance matters of degree. Although this would weaken the assumptions of the decompositional strategy somewhat, the focus of these accounts is still on accurately capturing (to some sufficiently high degree) or "taking into account" the "most important" causal factors; e.g. those that have a major impact on the probability of the outcome. As a result, although these views use slightly different criteria to determine which causal factors are relevant to the model's explanation, they still

[10] Thanks to an anonymous reviewer for suggesting that I address these potential tweaks. However, although Potochnik's view illustrates one way to potentially weaken the assumptions of the decompositional strategy, it isn't clear to me that her view is an instance of that strategy since she claims there are lots of ways to "account for" causal factors without accurately representing them. Moreover, there are important parallels between certain parts of my holistic distortion view and the view Potochnik defends, but there are also crucial differences; e.g. my account does not require explanations to identify causal patterns.

identify an isolable set of relevant causal factors that idealized models must capture in order to explain. Consequently, these ways of weakening the assumptions of the decompositional strategy still won't be able to account for the kinds of holistic distortion cases presented below where the model pervasively and drastically distorts most of system's difference-making causes—including those that are quite important to the target phenomenon and have a significant impact on the probability of its occurrence. In short, simply narrowing the set of relevant features to a somewhat smaller set than all causal difference makers, or allowing for minor distortions of less significant difference-making features, won't be able to overcome the challenges to the decompositional strategy posed by the kinds of holistic distortion cases presented below.[11]

In sum, all of these accounts require that the real-world system be decomposable into difference-making (or relevant) and non-difference-making (or irrelevant) features—even if they determine which features are relevant (or important) in slightly different ways. They also require that idealized models can be decomposed into their accurate and inaccurate parts. Finally, they claim that idealized models explain when they accurately represent (or capture) the relevant features (e.g. difference-making causes) and the idealized parts of the model only distort irrelevant features.

2.4. Robustness analysis and decomposition

Another way philosophers have employed the decompositional strategy is in applications of robustness analysis (Levins [1966]; Weisberg [2006]; Wimsatt [1981]). Richard Levins [1966] initially argued that, if we investigate a number of distinct models that, 'despite their different assumptions, lead to similar results, we have what we can call a robust theorem that is relatively

[11] Allowing for drastic distortion of difference-making features, or suggesting that there is no set of relevant features that the model ought to capture, would just be to abandon the core tenets of the decompositional strategy.

free of the details of the model' (p. 20). This result, Levins argues, allows us to determine 'whether a result depends on the essentials of the model or on the details of the simplifying assumptions' (p. 20). William Wimsatt [1981] generalizes Levins's ideas by suggesting that:

> All the variants and uses of robustness have a common theme in the distinguishing of the real from the illusory; the reliable from the unreliable; the objective from the subjective; the object of focus from the artifacts of perspective; and, in general, that which is regarded as ontologically and epistemologically trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting. (p. 128)

This is a fairly clear endorsement of the general idea behind the decompositional strategy: robustness analysis enables scientific models to be broken down into the dissociable contributions made by their accurate (i.e. real) and inaccurate (i.e. illusionary) parts where the accurate parts represent the relevant features of the model's target system(s).

In addition, more contemporary accounts of robustness analysis have suggested that when we have a robust result we can sometimes infer additional claims about real-world systems. For example, Weisberg suggests that scientists can use robustness analysis to discover 'robust properties' that are the result of a 'common structure' of a set of idealized models (Weisberg [2006], pp. 736-737). Weisberg then suggests that these mathematical results are typically applied by attempting to 'map [the equations] on to the properties of real or imagined…systems' (Weisberg [2006], p. 738). This translates the common mathematical structure of the models into a common *causal* structure of real (or possible) systems. As a result, the modeler can claim that there is a robust theorem that holds between the common causal structure and the robust property in real-world systems. These further steps require that the model's target systems can be decomposed into a common causal structure and other irrelevant factors in ways that mirror the decomposition of the set of idealized models into a common mathematical structure and other auxiliary assumptions. Moreover, the method requires that we can map the isolable common structure of the models onto the relevant causal structure in the real-world system(s) that is responsible for the robust property. As a result, Weisberg's description of how to fully apply robustness analysis to real-world systems requires a version of all three assumptions of the decompositional strategy.[12]

2.5. Scientific realism and decomposition

Finally, many defenses of scientific realism make use of a version of the decompositional strategy by suggesting that scientific models and theories are *partially* accurate representations. These are sometimes referred to as "selective confirmation" approaches to defending scientific realism (Leplin [1997]; Peters [2013]; Stanford [2003]). For example, Christopher Pincock [2011] has applied a version of the decompositional strategy in precisely this way. The general idea is that models can be decomposed into parts and we can believe that our best scientific models are sometimes accurate with respect to certain relevant parts of real-world systems. Therefore, we can maintain a version of realism because scientific models are sometimes accurate representations of the relevant features of real-world systems even if they distort other (presumably irrelevant) features.

2.6. Three assumptions of the decompositional strategy

What I have tried to show is that for a wide range of accounts in the philosophical literature there is an assumed mapping between the decomposition of scientific models and the decomposition of their target systems.[13] More specifically, I argue that the accounts detailed above each adopt a

[12] While not all applications of robustness analysis require a decompositional approach, often the inferences made require some form of those assumptions. For example, other defenders of robustness analysis have suggested that robustness analysis can allow us to 'be more confident that the result depends not on the falsities we have introduced into the modeling, but rather on the common components' (Kuorikoski et al. [2010], p. 551).
[13] Of course, the decompositional strategy might just claim, "natural systems and models have parts", but that is a rather uncontroversial claim and is too widely applicable to be interesting.

general decompositional strategy that requires some version of the target decomposition assumption, model decomposition assumption, and mapping assumption that were presented at the beginning of this section.

3. Against the Decompositional Strategy

In order for the decompositional strategy to succeed, the three conditions given above must be met. However, in this section, I provide examples of how the model decomposition assumption and the mapping assumption can, and often do, fail to hold. Moreover, I argue that these cases are representative of much larger classes of models that will systematically fail to meet the requirements for applying the decompositional strategy. While I think objections can also be raised regarding the target decomposition assumption, I think philosophical accounts of modeling, explanation, idealization, and robustness ought to be kept independent of metaphysical commitments about the nature of real-world systems whenever possible.[14] Consequently, I will simply note that there is no clear argument for requiring such a strong metaphysical assumption regarding real-world systems and instead focus my critique on the model decomposition assumption and the mapping assumption.

3.1. Many scientific models don't decompose that way

The first problem with the decompositional strategy is that many of our best scientific models cannot be decomposed into the isolable contributions made by their accurate and inaccurate parts. According to the decompositional strategy, we should be able to isolate the contributions of the accurate part(s) of our models from the contributions made by their idealized part(s). This

[14] I also think that the realism debate should avoid having philosophical accounts of realism imposing metaphysical structures on the world a priori. Instead, our metaphysical claims about the world ought to be epistemically justified by first looking at the nature of our models/theories and how they are able to relate to the world. However, filling out the details of this proposal would take us too far away from main focus of this paper.

implies that we should, at least in principle, be able to remove or replace the idealizations within our scientific models while leaving the contributions of the isolated accurate components intact. In other words, if scientific models are truly decompositional in this way, then the idealizations within our best scientific models should be eliminable in the sense that they could in principle be removed (or replaced) without affecting the parts (or contributions) of the model that accurately describe the relevant parts (or features) of the model's target system(s). However, in this section, I argue that for a wide range of scientific models, idealizations cannot be quarantined in this way and their distortions are *pervasive* due to the fact that they are essential to the foundational mathematical frameworks of those models. Although a model's having ineliminable idealizations and being unable to decompose the contributions of the model's accurate and inaccurate parts are conceptually distinct, we can test the model decomposition assumption by considering what occurs to the contributions of the (purportedly) accurate parts of the model when the idealizations are eliminated.[15] In many cases, the idealizations are essential (i.e. ineliminable) for the mathematical frameworks used in the model, which results in the pervasive distortion of most of the system(s) features, which in turn leads to the representation of the relevant (or important) parts of the system being distorted through the lens of the idealized mathematical framework. Consequently, the parts of the model that are supposed to be accurate representations of relevant features can only make their contributions within the context of the idealized mathematical modeling framework that pervasively distorts them (and many other features). When this is the case, the contributions of the idealized parts of the model cannot be isolated from the contributions made by the accurate parts of the model, but are instead *intertwined* within the pervasively distorted representation provided by the mathematical model.

[15] Thanks to an anonymous reviewer who helped me be clearer about these distinctions and the connections between these concepts.

As a first example, we can consider the ideal gas law, which states that $PV = nRT$ where P is pressure, V is volume, T is temperature, n is the number of moles of gas, and R is the constant. This highly idealized equilibrium model is derived from simpler laws of gases (e.g. Boyle's Law and Charles' Law) and can be used to calculate macroscopic changes of measurable variables in real gases. The ideal gas law is at the heart of the kinetic theory of gases, which utilizes the Maxwell-Boltzmann distribution for the velocities of molecules. This distribution is derived by imposing a particular probability distribution on the micostates of the system and then averaging over those microstates to discern macroscale properties of gases. More specifically, the Maxwell-Boltzmann distribution requires that one assume that the molecules are in constant random motion, do not interact, and have velocities that are statistically independent of one another. This enables one to model the speed of the molecules using a Gaussian distribution (i.e. a bell curve) by applying the central limit theorem. Importantly, however, the Maxwell-Boltzmann distribution applies only to ideal gases. In real gases, there are several additional factors (e.g. van der Waals interactions, vertical flows, relativistic speed limits and quantum exchanges) that make the gas particles' speeds often very different from those specified by the Maxwell-Boltzmann distribution.

Indeed, the exact calculations provided by the ideal gas law require a long list of idealizing assumptions including:

1. The gas consists of a large number of identical molecules in constant random motion.
2. The volume occupied by the gas molecules is infinitesimally small compared to the volume of the container; i.e. the molecules do not take up any space.

3. The velocity (components) of each of the particles are statistically independent of one another.

4. The molecules exert no long-range forces on one another and there are no intermolecular forces between the molecules.

5. Collisions between the molecules and the walls of the container are perfectly elastic (or are simply assumed not to occur).

6. The gas obeys the processes of classical Newtonian mechanics.

Of course, each of these assumptions will fail to hold in many real-world gases, and no real-world gas will satisfy all six of them. Despite these distortions, the ideal gas law can be used to explain and understand various behaviors of real gases. Indeed, many real gases behave close to the ways predicted by the ideal gas law and it has been used to explain fundamental features of gas behavior such as diffusion and pressure.

We might, then, ask the following question: What would the ideal gas law look like without these idealizations? Put slightly differently, which parts of the ideal gas law are the dissociable accurate parts that are describing the relevant (difference-making) parts of reality? I contend that we cannot really answer either of these questions because the idealizing assumptions made by the model are introduced to apply foundational mathematical techniques that result in a *pervasively distorted representation* of actual gases. Put differently, the contributions made by the purportedly accurate parts of the model only make their contributions within an idealized mathematical modeling framework that pervasively distorts those features. As a result, the contributions of the idealized and (purportedly) accurate parts of the model are intimately intertwined within a pervasive misrepresentation of the model's target system(s).

The issue here is that—contra Strevens [2009]—the idealizing assumptions within the ideal gas law are not simply noting that certain features of the gas are known to be irrelevant. Instead, they pervasively distort the components and interactions of real gases in order to allow for the application of mathematical tools—e.g. statistical modeling techniques—that enable scientists to extract explanatory information about the system. Without these idealizations the mathematical foundations of the model would not be applicable. It is precisely because these idealizations are essential to the foundational mathematical techniques involved that they are ineliminable from the explanations provided by the model and their distortions are so pervasive. The important point is that there is not an isolable part of the ideal gas model that accurately represents some dissociable set of difference-makers in real gases and is unaffected by the list of idealizing assumptions given above. Instead, an ideal gas is a fundamentally *different kind* of system—an idealized model system—that allows physicists to investigate, explain, and understand real-world gases, but does not accurately represent some dissociable set of difference makers of real gases. The fact that these idealizations are essential to the mathematical frameworks employed by the model demonstrates their pervasive effect on the overall description provided by the model. It also shows that it will be impossible to isolate the contributions made by some accurate part(s) of the model from the contributions made by these inaccurate (i.e. idealized) parts. The distortions introduced are simply too pervasive because they are constitutive of the foundational mathematical techniques required for the model to provide the explanation.

Parallel applications of mathematical modeling can be found in biology (Ariew et al. [2015]; Rice [2012], [2015], [2016]; Walsh et al. [2002]). Indeed, the foundational assumptions made by R. A. Fisher in developing population genetics were inspired by the assumptions underlying the kinetic theory of gases (Morrison [1996], [2004], [2015]).[16] Like Maxwell and Boltzmann, Fisher's general approach was to make various idealizing assumptions about the nature of the individual components of the system and their interactions in order to develop general statistical models of the large-scale behaviors of populations (Morrison [2004]). Assuming that the individual-level events of the population are random and statistically independent allowed Fisher to apply the central limit theorem, which tells us that such a sample will conform to the normal distribution (i.e. a Gaussian bell curve). Then, by assuming the population is infinitely large, one can apply various laws of large numbers to eliminate sampling error (in this case genetic drift).[17] Finally, Fisher averaged over the individual-level events in order to identify statistical features of the overall distribution of genotypic trait types and their fitnesses. In combination, these assumptions allowed Fisher to assume that biological populations approximated a normal distribution whose mean value was the expected outcome (presumably due to natural selection). As a result, Fisher argued that biological populations could be modeled statistically simply by knowing the mean and variance of the overall population. In other words, Fisher's statistical assumptions allowed him to model evolving populations in such a way that no knowledge of the parts or their interactions is required. As Margaret Morrison puts it, Fisher saw that, 'the kinetic theory had shown that knowledge of particular individuals was not required in order to formulate general laws governing the behavior of a population' (Morrison [1996], p. S319).

More importantly, with Fisher's statistical modeling we again see that when certain idealizing assumptions are made about the components of the system, scientists can use various

[16] Fisher himself tells us that: 'The whole investigation may be compared to the analytical treatment of the Theory of Gases' (Fisher [1922], p. 321).
[17] The law of large numbers states that, as the samples size increases, the average of the quantities sampled will be closer to the expected outcome.

mathematical techniques to construct models with which to investigate, explain, and understand real-world systems.[18] In order to do so, however, often requires that they move to a drastically distorted representation of those systems in order to apply those modeling techniques. The success of Fisher's approach was due to his replacing actual populations with highly idealized model populations that relied on the statistical assumptions he adopted from gas theory. As a result of his idealizing assumptions, Fisher *constructed* an infinite model population in which evolution did not involve migration, genetic recombination, genetic interaction, or drift. This resulted in a *pervasively distorted representation of the evolutionary processes in any real-world population.* Indeed, Fisher's work on population genetics was able to provide various explanations 'only by invoking a very *unrealistic and abstract* model of a population' (Morrison [2015], p. 24). These kinds of statistical modeling techniques would later become the foundation of modern population genetics. The most important parallel, for my purposes, is that once again the contributions made by the idealizing assumptions cannot be isolated from the contributions made by the accurate parts of these models because the idealizing assumptions are *necessary for applying the mathematical framework* used by these scientific modelers.

While some idealized models can perhaps be decomposed into the contributions made by their accurate and inaccurate parts, I argue that this decomposition is impossible for many of our best scientific models because the idealizations are essential to the foundational mathematical frameworks used within those models. As a result, without the idealizations, the mathematical techniques will not be applicable and so the explanation and understanding provided by the

[18] As Morrison explains, 'Fisher's mathematisation of selection created a new framework in which its operation was understood as an irreducibly statistical phenomenon, a reconceptualisation that emerges in conjunction with the application of specific mathematical...techniques' (Morrison [2015], p. 41). See Walsh et al. [2002] and Ariew et al. [2015] for a similar interpretation.

model would be inaccessible.[19] In addition, the essential role these idealizations play in the mathematical foundations of many scientific models shows that the contributions of the inaccurate parts cannot be isolated in the way required by the decompositional strategy. Instead, the purportedly accurate parts can only make their contributions within an idealized modeling framework that drastically distorts the model's target system(s). Therefore, contrary to the decompositional strategy, in many cases idealizations are not innocent bystanders that can be quarantined by only distorting irrelevant (or insignificant) features; instead, they are deeply invested collaborators that allow for the application of various mathematical modeling techniques.

3.2. Many idealizations distort difference-making features

The second challenge to the decompositional strategy is that even if we assume the real-world system and the idealized model are decomposable in the ways required, the model's idealizations will often distort difference-making (i.e. relevant) features of the model's target system(s). As a result, we cannot map the accurate parts of the model onto what is relevant and its inaccurate parts onto what is irrelevant.

For example, the Hardy-Weinberg equilibrium model is used to explain and understand various features of heredity and variation. The model establishes a mathematical relation between genotypic frequencies that captures the genetic structure described by Mendelism. The model tells us that if we have a pair of alleles, A_1 and A_2, at a particular locus and in the initial population the ratio of A_1 to A_2 is p to q, the distribution for all succeeding generations will be,

[19] It is, however, possible that removing or replacing the idealizing assumptions would result in a modified version of the model, a different kind of model, or a different explanation. The most important point here is not whether the idealizations can be eliminated from the model, but instead is the pervasive nature of the distortions they introduce due to their role in the foundational mathematical frameworks of the model. This is enough to show the failure of the model decompositional assumption. My thanks to an anonymous reviewer for helping me clarify this point.

$$p^2 A_1A_1 + 2pqA_1A_2 + q^2A_2A_2$$

regardless of the distribution of genotypes in the initial population (or generation). This mathematical model describes the frequencies of different genotypes at a single locus in an infinitely large population, in which mutation, selection, sampling error, and migration do not occur, all members of the population breed, all mating is completely random, all organisms have the same number of offspring, there is no intergenerational overlap, and all these conditions are held constant. However, each of these distorted features makes a difference to the evolution of most (if not all) real-world biological populations. For example, in all real finite populations there will be some non-negligible chance that trait frequencies will diverge from the values predicted by selection; i.e. drift makes a difference to every real-world population (Rice [2015]). As a result, removing sampling error (i.e. drift) from this mathematical model distorts a difference-making feature of every real-world population. In addition, the Hardy-Weinberg model assumes that there is no intergenerational overlap, but this is false of (almost) every real-world population and is a difference-making feature for many evolutionary outcomes (Levy [2011]).

In addition, as was the case in the examples above, many of these idealizing assumptions are necessary for the mathematical foundations of the model (Morrison [2015]). For example, without the assumptions of infinite population size and random mating, the stability of genotypic frequencies across generations will be violated—that is, without these assumptions the very mathematical framework used by the model that allows it to capture key features of heredity and variation will no longer be applicable.

More generally, biological modelers frequently utilize several idealizations that distort the actual processes involved in mating, genetic recombination, inheritance, and selection in order to apply various mathematical modeling techniques. In doing so, they often distort features of real-world systems that make a difference to their evolutionary outcomes (Potochnik [forthcoming]; Rice [2012], [2015], [2016]). Indeed, using idealizations that distort difference-making factors is pervasive in biological modeling—even within mechanistic modeling (Love and Nathan [2015]).[20]

What is more, idealization of difference-makers is widespread in several other sciences; e.g. models in chemistry and physics can be used to explain despite misrepresenting the fundamental kinds of components, interactions, and properties that exist in real systems (Batterman [2002]; Batterman and Rice [2014]; Bokulich [2011], [2012]; Morrison [2015]).[21] As a result of their distortion of difference makers, the distortions introduced by the inaccurate parts of these models cannot be isolated to the distortion of the irrelevant parts (or features) of their target system(s). Therefore, once again, for a wide range of cases across numerous sciences one of the three core assumptions required to apply the decompositional strategy fails to hold.

4. An Alternative Approach: The Holistic Distortion View of Idealized Models

The general goal of the decompositional strategy is to show that the accurate parts of the model are what "do the real work" while the inaccurate parts of the model are justified by distorting only what is known (or assumed) to be irrelevant. However, the arguments above show that for a

[20] For example, even more distortion of difference-making features is involved in game-theoretic modeling, which routinely uses idealizing assumptions such as random pairing of players, symmetric contests, constant payoff structures across iterations of the game, etc.

[21] For example, Alisa Bokulich ([2011], [2012]) discusses various cases in which fictions that drastically distort relevant features of the target system are employed to explain in physics and chemistry. Batterman and Rice [2014] also discuss examples of models in physics and biology that distort difference-making causes and yet are able to provide explanations.

wide range of cases across many scientific disciplines this decompositional strategy will fail. As a result, an alternative approach is needed. In what remains, I will argue for what I call the *holistic distortion view* of idealized models (Rice [forthcoming]). In its most general form, holism is just the thesis that the whole is more than the sum of its parts. I contend that many idealized models in science are holistic distortions that cannot be decomposed into the contributions made by their accurate and inaccurate parts.

It is important to note, however, that my holistic distortion view does not (directly) entail other more traditional forms of holism concerning semantics, metaphysics, or confirmation. Instead, I only advocate a more holistic approach *concerning philosophical attempts to analyze the use of idealized models in science*. Specifically, I contend that many (if not most) idealized models should be characterized as holistically distorted representations of their target system(s) that are greater than the sum of their accurate and inaccurate parts. The decompositional strategy mistakenly ignores the myriad ways in which the explanations and understanding provided by scientific models are typically the result of a rich and complicated mixture of various modeling assumptions whose contributions cannot be studied in isolation. Therefore, my holistic view is a *methodological prescription* for philosophers' attempts to understand how to model complex systems, how models explain, how idealizations contribute to model explanations, robustness analysis, and how idealized modeling is compatible with scientific realism. Most importantly, philosophy of science must move beyond accounts of these aspects of modeling that require the assumptions of the decompositional strategy.

The second part of my view is that idealizing assumptions that result in holistic distortions are often essential to the explanations provided by scientific models because they allow for the application of various (mathematical) modeling techniques (Cartwright [1983], Ch.

7; Rice [forthcoming]). In addition to the cases discussed above, a particularly instructive example is physicists' use of the thermodynamic limit. The thermodynamic limit 'is the limit in which (roughly speaking) the number of particles of the system…approaches infinity' (Batterman [2010], p. 7). What is important to note is that often, 'This limiting idealization is essential for the explanation because for a finite number of particles the statistical mechanical analogs of the thermodynamic functions cannot exhibit the nonanalytic behavior necessary to represent the qualitatively distinct behaviors we observe' (Batterman [2010], p. 7). Indeed, for a wide range of modeling techniques in physics, the thermodynamic limit is an *essential mathematical operation*. For example, as Morrison explains in the case of modeling phase transitions:

> The occurrence of phase transitions requires a mathematical technique known as taking the 'thermodynamic limit,' $N \rightarrow \infty$; in other words we need to assume that a system contains an infinite number of particles in order to understand the behavior of a real, finite system…[since] the assumption that the system is infinite is *necessary* for the symmetry breaking associated with phase transitions to occur. In other words, we have a description of a physically unrealizable situation (an infinite system) that is *required* to explain a physically realizable phenomenon (the occurrence of phase transitions). (Morrison [2009], p. 128).[22]

Like the cases discussed above, the thermodynamic limit is not introduced simply as a way of ignoring what is irrelevant to the target explanandum or as a method for calculational expediency. Instead, these idealized mathematical descriptions function as *a necessary condition* for using the mathematical techniques required to explain and understand the phenomenon of interest (Batterman [2002], Morrison [2009], Kadanoff [2000]).

As I noted above, additional instances of the use of idealizations to allow for the application of mathematical techniques can be found across biological modeling. As another example, optimization models that provide adaptationist explanations make use of numerous

[22] Or as physicist Leo Kadanoff [2000] puts it, 'The existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom' (p. 238).

idealizations that allow for exact calculations of the equilibrium point of the evolving system (Rice [2012], [2015], [2016]). Rather than distorting irrelevant factors so scientists can focus on the accurate representation of difference-making features (e.g. natural selection), these models purposefully move scientists away from even attempting to accurately represent some isolable part of the dynamical processes that led to the explanandum. Instead, in these cases, the idealizing assumptions enable scientists to apply various mathematical techniques that allow them to calculate exactly how changes in the parameters involved in the system's constraints and tradeoffs will result in changes in the expected equilibrium point of the system. The explanation is then provided by showing how the optimal strategy (counterfactually) depends on those constraints and tradeoffs—despite the fact that the optimal strategy is the expected outcome within the idealized model for very different reasons than in the model's target system(s).

Indeed, as Mark Pexton notes, in many scientific models there are 'ineliminable misrepresentations of the true ontology of a system…[They] are necessary in some models because we require them to frame a system in a certain way in order to extract modal information' (Pexton [2014], p. 2344). It is this last part about the way in which idealizations allow scientists to frame the system as a whole to extract modal information about the system that I think has been largely missed by the philosophical literature (although Cartwright [1983], Ch. 7 and parts of Potochnik [2015] are important exceptions).[23] The source of this failure is that most philosophers (and many scientists) have sought to quarantine idealization in order to show that the falsehoods used in science are "harmless" because they only distort irrelevant or

[23] Potochnik's [2015] account in which idealizations can be used to frame the system in different ways that make different causal patterns accessible is similar in that the reframing of the system make positive epistemic contributions. The key difference between our views is that I would resist the claim that this is always done by enabling the model system to display a causal pattern.

unimportant features and do not "get in the way" of the accurate parts of scientific models that do the real work.

As an alternative project, I suggest that philosophers of science ought to provide accounts of the *pervasive* and *unique* contributions that idealizations that distort difference-making (i.e. relevant) features can make to models *conceived of as holistic distortions*; e.g. by enabling scientific modelers to use certain mathematical modeling tools that extract the explanations and understanding they are interested in.[24] Rather than attempting to isolate individual idealizations and show that their distortions are irrelevant (one at a time), we need to analyze how the practice of constructing holistically distorted models is able to consistently make positive contributions to achieving the goals of scientific inquiry—e.g. explanation and understanding—without having to provide an accurate representation of an isolated set of difference-making features.[25]

Here, then, is the core of my holistic distortion view. First, in a wide range of cases, idealized models pervasively distort the fundamental nature of the entities, processes, and features of their target systems—including difference makers. That is, their distortions are holistic, rather than piecemeal. Second, these idealizing assumptions often move scientists to an entirely different representational framework in which the mathematical tools necessary to explain and understand the phenomenon of interest are applicable. These different mathematical modeling frameworks represent different features and patterns of the system (e.g. statistical patterns at the population level) in different ways (e.g. as continuous processes or isolable

[24] This is not intended to suggest that no philosophers have discussed some of the positive contributions of idealizations (e.g. see Strevens [2009] or Potochnik [2015]), only that I advocate focusing on the development of those accounts in line with the holistic distortion view that allows for the distortion of difference-makers instead of accounts that require the assumptions of the decompositional strategy. These contributions also provide access to explanations that would otherwise not be accessible rather than accomplishing what could also be accomplished by the use of abstraction of irrelevant features. Consequently, these contributions go beyond the indication of the irrelevance of some features—this is what makes these idealizations' contributions to holistic distortions unique.
[25] I do not intend to suggest that these are the only goals of scientific inquiry. They are merely representative of the kinds of goals that scientists have been able to achieve with holistically distorted models.

factors) and allow for the use of different techniques for deriving the behavior of the system (e.g. using statistical limit theorems). As a result, these idealizations are often necessary for the models to provide a particular explanation (or understanding) of the target phenomenon—one that is sometimes the only explanation available. A model without these idealizing assumptions would be unable to use the mathematical techniques required for extracting the explanatory information of interest (Rice [forthcoming]). The goal, therefore, should be to justify scientists' use of these holistically distorted representations in terms of the explanations and understanding they provide that would otherwise be unattainable. Analyzing these unique contributions of idealizations that distort difference-making features is in stark contrast to trying to show that the inaccurate parts of scientific models distort only what is irrelevant (or only tell us about what is irrelevant).

The remaining challenge, then, is to show how such holistically distorted representations can still provide explanatory information given that most accounts of explanation (e.g. those discussed above) require accurate representation of difference-making features in order for models to explain. While I advocate a pluralistic approach to understanding the relationships between idealized models and real-world systems, one particularly promising strategy is to appeal to a feature called *universality* (Batterman and Rice [2014]; Rice [forthcoming]). The term universality comes from mathematical physics, but in its most general form it is just an expression of the fact that many systems that are (perhaps extremely) heterogeneous in their physical features will nonetheless display similar patterns of behavior at macroscales. The group of systems that will display similar macrobehaviors despite differences in their physical details are said to be in the same *universality class.* As physicist Leo Kadanoff [2013] puts it, 'Whenever two systems show an unexpected or deeply rooted identity of behavior they are said to be in the same universality class' (p. 178). In other words, universality guarantees that a class of systems, called the universality class, will display the same general patterns of macrobehavior even if the components, interactions, dynamics, and physical details of those systems are very different. For example, universality can connect the behavior of systems as diverse as magnets and fluids (Batterman [2002]). It can also show why a large class of biological systems will all reach the same equilibrium point (e.g. a 1:1 sex ratio) despite differences in the kinds of species, methods of inheritance, size of the population, etc. (Batterman and Rice [2014]). In addition, within these universality classes there are often various *model systems*—that is, universality classes that include a range of real-world systems will often also include some (imaginary or possible) systems described by scientific models. When this is the case, the model system(s) and the real-world system(s) will display similar macroscale behaviors despite having perhaps drastic differences in their fundamental components, interactions, and other features. This connection, I contend, is often what enables scientists to use highly idealized models that holistically distort their target system(s) to explain various behaviors of their target system(s). In other words, universality can show us that there is a class of systems, which includes the idealized model system and its target system(s), that will exhibit the same macroscale patterns of behavior and this link is often what allows scientists to justifiably use holistically distorted models to explain and understand the behavior of real systems (Batterman and Rice [2014]; Rice [forthcoming]).

It is important to distinguish this appeal to the empirical fact of universality to link model systems with their target systems from the use of mathematical techniques to *explain universality itself*. For example, the renormalization group (RG), has been used to explain why a wide range of fluids (and magnets) all display universal behaviors near their critical points (Batterman [2002]). RG is a mathematical technique that effectively transforms a system's original

Hamiltonian into another Hamiltonian that describes a system with fewer degrees of freedom. Repeatedly applying this technique at successively larger scales eventually identifies a fixed point Hamiltonian that preserves the original Hamiltonian's form and phenomenological behavior, but eliminates many of the irrelevant physical details. Furthermore, demonstrating that diverse fluids all flow to the same fixed points in the space of Hamiltonians under repeated applications of RG explains why those systems will display the same macrobehavior despite perhaps drastic differences in their physical details. As a result, RG enables physicists to explicitly delimit the class of systems that will display similar macroscale behaviors; e.g. having the same critical exponent when undergoing a phase transition.

RG effectively differentiates a set of relevant features from a set of microphysical details that are irrelevant and, in doing so, explains why the universal macroscale behaviors are stable across that class of systems—this is what enables RG to provide an explanation *of universality*. In other words, rather than aiming for the construction of models that accurately represent relevant features and idealize irrelevant features, RG is a mathematical modeling technique—whose application also requires the thermodynamic limit—that systematically distinguishes those features that are relevant and irrelevant in order to provide an explanation of the stability of the macroscale behaviors across a diverse range of (real, possible, and model) systems. Moreover, in addition to explaining various instances of universality using RG, physicists have also discovered several so-called “minimal models” (e.g. the lattice gas automaton model) within those universality classes that display the same patterns of behavior and can be used to further investigate and explain the behaviors of real-world systems (Batterman and Rice [2014]).[26]

While providing additional demonstrations that explain universality itself will typically

[26] Furthermore, while RG is an instructive example, universality can also be explained using other methods; e.g. appealing to some kind of multiple realizability.

provide greater degrees of understanding and justification (and may be required to provide a complete explanation for some explananda), providing a complete explanation of why universality occurs (e.g. by using RG) is not a requirement for appealing to the empirical fact of universality to connect the behavior of various real, possible and model systems in ways that allow for idealized models to be justifiably used to discover explanatory information.[27] That is, scientists can justifiably use idealized models within a universality class to explain the behaviors of real-world systems in that class *even when they fail to have a complete explanation of why that universality class occurs.* Indeed, there are many universality classes out there that contain various real, possible, and model systems, but whose universality has yet to be explicitly delimited (i.e. specifying precisely the range of systems in the class) or completely explained (i.e. explaining why those patterns are stable across that class of systems).

In this context, the universality class consists of the set of real, possible and model systems that display the patterns of counterfactual dependence of interest to the scientific modeler. In some cases, the idealized models that are within the same universality class as their target systems will just be those that accurately represent the difference-making features of those systems.[28] However, in many other cases, the same macrobehaviors will be produced by *different* sets of difference-makers and extremely heterogeneous components and interactions across a range of possible systems. In these cases, scientists often construct a pervasively distorted model in order to apply the (mathematical) modeling techniques and theories they have on hand. The resulting idealized model is then used to investigate how various (hopefully measurable) features are counterfactually related to the phenomenon of interest. When the holistically distorted model

[27] Indeed, scientists often explain by appealing to idealized models that are in the same universality class as their target systems before any explicit delimiting or explaining of universality has been attempted.
[28] For example, this will occur when the same set of difference-makers are necessary and sufficient conditions for the occurrence of the target explanandum in every real and possible system.

is (or perhaps can be shown to be) in the same universality class as the system(s) of interest, scientists can justifiably use the holistically distorted model to explain because it can be used to discover information about which features are counterfactually relevant (and irrelevant) to the target explanandum.

For example, the ideal gas law need not accurately represent the ontology of real gases or the actual processes that relate pressure and temperature in order to show us how pressure counterfactually depends on temperature (and vice versa). All that is required is that the model enable us to see how changes in one feature will result in changes in the other and universality can guarantee that those macroscale patterns of counterfactual dependence will be preserved even though the model is a holistically distorted representation of the components, processes, and interactions of real gases. This is because the macroscale patterns of counterfactual dependence between measurable features of real gases will be realized by *any system in the universality class*—including the pervasively distorted model system.

As a second example, a biological optimization model can show us how changes in the constraints and tradeoffs among various features of the system will change the equilibrium point of the evolving population *even if the model describes a selection process that is nothing like the one that produced the explanandum in its real-world target system(s)* (Rice [2012], [2015], [2016], [forthcoming]). The explanation is provided by showing how the equilibrium point of the evolving system is counterfactually related to certain features of the target system, not by accurately representing those features or some modular part of the dynamics that connects those features to the explanandum in the real-world system. In these cases, biological optimization models show how the target explanandum is counterfactually dependent on certain constraints (e.g. the amount of resources available) and tradeoffs (e.g. between average energy intake and average predation risk) even though the model represents these features in a highly distorted way. Moreover, the optimization model describes a selection process—which connects these features to the target explanandum—that drastically distorts the selection process that occurs in real biological populations; e.g. these models typically represent selection without a temporal dimension and assume that offspring perfectly resemble their parents within an infinitely large population. Nonetheless, biologists have discovered that (and in some cases demonstrated why) these kinds of counterfactual dependencies are stable across a range of real, possible, and model systems that are heterogeneous in many of their details including having different species, interactions, inheritance processes, etc. That is, although it describes a system that is drastically different from its target system(s), the optimization model—in virtue of being in the same universality class—still displays certain macroscale patterns of counterfactual dependence between certain constraints and tradeoffs and the target explanandum that can be used to provide an explanation. Revealing these patterns of counterfactual dependence by applying mathematical modeling techniques, however, is importantly different from accurately representing those features and the ways they produce (or result in) the target explanandum in real-world systems.

In sum, highly idealized models can allow scientists to discover counterfactual dependencies that hold between certain features and the target explanandum without having to accurately represent the entities, processes, mechanisms, or difference-makers of their target system(s). Importantly, this kind of information about counterfactual dependencies is widely held to be important to many forms of explanation (Bokulich [2011], [2012]; Rice [2015]; Woodward [2003]). As Woodward puts it, '[an] explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways' (Woodward [2003], p.11). In addition, Bokulich argues that, ‘in order for a model M to explain a given phenomenon P, we require that the counterfactual structure of M be isomorphic in the relevant respects to the counterfactual structure of P’ (Bokulich [2011], p. 39). Indeed, information about counterfactual dependence is central to several accounts of how models are able to explain (Bokulich [2012]; Rice [2015], [forthcoming]; Saatsi and Pexton [2013]; Woodward [2003]).[29]

The key thing to notice is that idealized models can provide information about counterfactual dependencies even when they fail to accurately represent the system’s components and interactions or the actual processes that link those components and interactions to the target explanandum. In many instances this is because universality guarantees that the model system’s patterns of macroscale behavior will be similar to those of the target system(s) even if the actual entities and processes of those systems are extremely different. In other words, universality can enable scientific modelers to discover relationships of counterfactual dependence between certain features and an explanandum even if those relationships hold for drastically different reasons in the idealized model system than they do in the real-world system(s). Therefore, even if the model drastically and pervasively distorts the fundamental nature of the entities and processes of real-world systems in order to use various mathematical modeling techniques, it can still be used to explain because many of the patterns of counterfactual dependence that hold in the model system will be similar to those of real-world systems—those counterfactual relations will just hold for (perhaps very) different reasons in the model system and perhaps only in limiting cases. Generalizing the concept of universality allows us to capture this stability of various patterns of behavior (e.g. counterfactual dependencies) that are largely independent of the physical components, interactions, and features of a heterogeneous

[29] While I believe there are both causal and noncausal explanations in science, the emphasis on counterfactual information is compatible with many of the exclusively casual accounts of explanation on offer.

class of (real, possible and model) systems (see Batterman and Rice [2014] and Rice [forthcoming] for additional details).

In this way, universality enables us to see how idealized models can be used to explain even when they are holistically distorted representations of their target system(s). Indeed, discovering stable universal behaviors allows scientific modelers to move away from attempting to accurately represent some isolable set of difference-making parts of real-world systems and instead focus on discovering highly idealized models within the same universality class that will enable them to apply the mathematical and theoretical techniques necessary to discover the counterfactual dependence information required to explain and understand various phenomena.

One of the key projects of philosophy of science should be to understand how scientists' use of highly idealized models has been so successful in terms of generating explanations and understanding of our world. I have argued that there are at least two links between idealized models and their target systems that can be used to justify scientists' use of idealized models to explain and understand real-world phenomena: (1) accurate representation of difference makers, or (2) being within the same universality class.

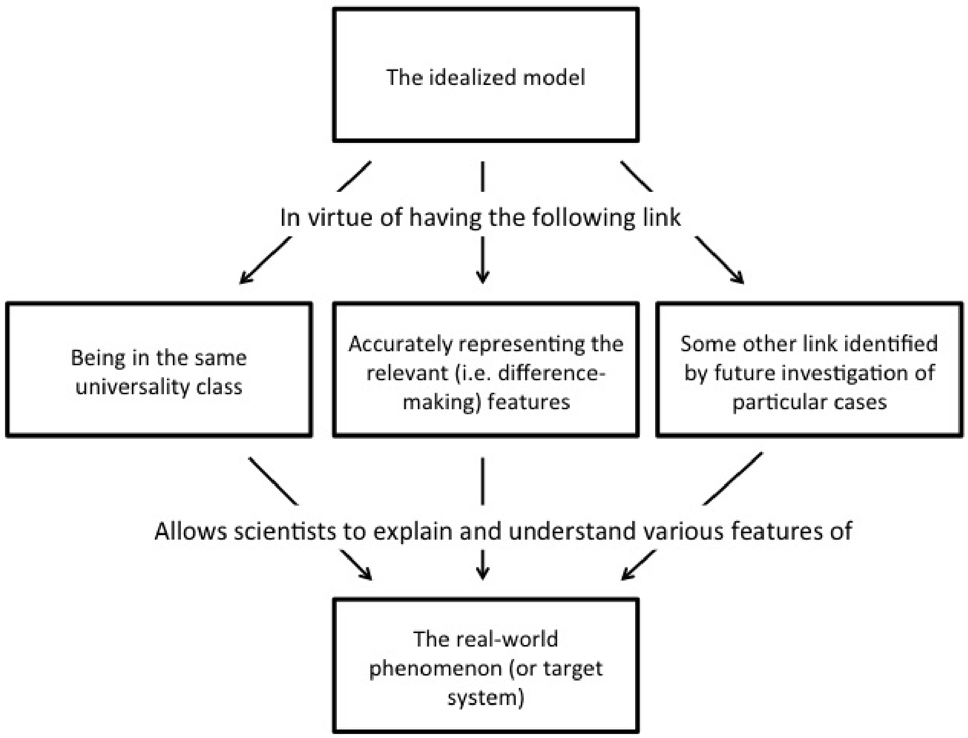

Figure 1: The different links that can hold between idealized models and their target systems that allow scientists to use the idealized model to explain and understand real-world phenomena.

While appealing to universality is a promising approach to justifying the use of holistically distorted models, I want to emphasize that I think one of the key lessons of the above discussion is that philosophers need to be more pluralistic about the kinds of relationships that can hold between idealized models and real-world systems that are sufficient for explanation and understanding. Some idealized models will accurately represent difference-makers of their target system(s), others will be in the same universality class as their target system(s), and others will perhaps exploit some alternative connection to their target system(s) (see Figure 1 above). Indeed, given the diverse array of modeling techniques and goals across various disciplines we should expect there to be an equally diverse set of ways that models can relate to real-world systems and provide the explanations and understanding of interest to scientific modelers.

In sum, the four main claims of my holistic distortion view of idealized models are the following:

(1) Many (if not most) idealized models in science ought to be characterized as holistically distorted representations of their target system(s).

(2) Idealizing assumptions that result in holistically distorted representations typically make essential contributions by allowing for the application of various (mathematical) modeling techniques.

(3) The use of such holistically distorted models ought to be justified by their ability to provide epistemic access to explanations and understanding that would otherwise be inaccessible.[30]

(4) Given the diversity of modeling techniques and goals found in scientific practice, we require a more pluralistic approach to investigating the relationships between models and their target systems that justify their use in developing explanations and understanding.

The task going forward is to provide the justification called for in (3) and (4) by looking at particular examples and enumerating the ways that holistically distorted models provide access to explanations and understanding that would otherwise be inaccessible. I have proposed one additional way this can be done: by appealing to the fact that the holistically distorted model is within the same universality class as its target system(s). I have provided an outline of that approach here, but, like any new account, some of the details of this approach will have to be worked out in other papers. However, these four claims do provide the foundation for a

[30] Holistically distorted models can also allow for predictions that would otherwise be inaccessible, but that kind of contribution is beyond the scope of this paper.

fundamentally different way of thinking about idealized modeling in science that avoids the mistakes of the decompositional strategy.

5. Conclusion

The decompositional strategy is pervasive across a wide range of debates in the philosophy of science. I have argued that the assumptions underlying these decompositional approaches will fail to hold for a wide range of cases of scientific modeling. In response, I have proposed the holistic distortion view, which offers an alternative approach to characterizing and justifying the use of idealized models. Going forward, I suggest that philosophers of science take up the challenge of showing how scientific modelers are justified in using idealized models that holistically distort their target systems by analyzing the ways in which various modeling techniques can produce explanations and understanding that would otherwise be inaccessible. In other words, philosophers of science need to investigate the ubiquitous and unique contributions that various holistically distorted scientific representations make to scientific inquiry rather than continuing their attempts to ignore, remove, isolate, or quarantine the roles idealizations play within scientific theorizing.

Acknowledgements

I would like to thank Robert Batterman, Angela Potochnik, André Ariew, Yasha Rohwer, and two anonymous reviewers for helpful comments and feedback on earlier versions of this paper. I would also like to thank audiences at the University of California, Irvine, the Munich Center for Mathematical Philosophy, Bryn Mawr College, and the 2016 Philosophy of Science Association Meeting for feedback and discussion. This work was supported by funding from the University of California, Irvine's Department of Logic and Philosophy of Science, a Lycoming College faculty development grant, a fellowship from the Munich Center for Mathematical Philosophy, and a research grant from Bryn Mawr College.

Collin Rice

Department of Philosophy

Bryn Mawr College

101 N. Merion Ave

Bryn Mawr, PA 19010

crice3@brynmawr.edu

References

Ariew, A. Rice, C. and Rohwer, Y. [2015]: 'Autonomous Statistical Explanations and Natural Selection'. *The British Journal for the Philosophy of Science*, **66**(3), pp. 635-658.

Batterman, R. W. [2002]: *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford: Oxford University Press.

Batterman, R. W. [2010]: 'On the Explanatory Role of Mathematics in Empirical Science', *British Journal for the Philosophy of Science*, **61**, pp. 1-25.

Batterman, R. W. and Rice, C. [2014]: 'Minimal Model Explanations', *Philosophy of Science*, **81***(3)*, pp. 349-376.

Bechtel, W. and Richardson, R. C. [1993]: *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton: Princeton University Press.

Bokulich, A. [2011]: 'How Scientific Models Can Explain', *Synthese*, **180**, 33-45.

Bokulich, A. [2012]: 'Distinguishing Explanatory from Nonexplanatory Fictions', *Philosophy of Science*, **79**(5), pp. 725-737.

Cartwright, N. [1983]: *How the Laws of Physics Lie,* Oxford: Oxford University Press.

Craver, C. [2006]: 'When Mechanistic Models Explain', *Synthese*, **153**, pp. 355-376.

Craver, C. [2007]: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Oxford University Press.

Craver, C. and Darden, L. [2013]: *In Search of Mechanisms: Discoveries Across the Life Sciences,* Chicago: University of Chicago Press.

Elgin, M., and Sober, E. [2002]: 'Cartwright on Explanation and Idealization', *Erkenntnis*, **57**, pp. 441-450.

Fisher, R. A. [1922]: 'On the Dominance Ratio', *Proceeding of the Royal Society of Edinburgh, V,* **43**, pp. 321-341.

Frigg, R. [2010]: 'Models and Fiction', *Synthese*, **172**, pp. 251-268.

Glennan, S. [2005]: 'Modeling Mechanisms', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, pp. 443-464.

Godfrey-Smith, P. [2009]: 'Models and Fictions in Science', *Philosophical Studies*, **143**, pp. 101-116.

Kaplan, D. M. and Craver, C. F. [2011]: 'The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective', *Philosophy of Science*, **78**, pp. 601-627.

Kadanoff, L. P. [2000]: *Statistical Physics: Statics, Dynamics, and Renormalization,* Singapore: World Scientific.

Kadanoff, L. P. [2013]: 'Theories of matter: Infinities and renormalization', In Robert Batterman (Ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, pp. 141–188.

Kuorikoski, J. Lehtinen, A. and Marchionni, C. [2010]: 'Economic Modeling as Robustness Analysis', *The British Journal for the Philosophy of Science*, **61**, pp. 541-567.

Levins, R. [1966]: 'The Strategy of Model Building in Population Biology', In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press, pp. 18-27.

Levy, A. [2011]: 'Makes a Difference'. Review of Michael Strevens' Depth', *Biology and Philosophy*, **26**, pp. 459-467.

Levy, A. [2014]: 'Machine-likeness and Explanation by Decomposition', *Philosophers' Imprint,* **14**(6), pp. 1-15.

Leplin, J. [1997]: *A Novel Defense of Scientific Realism*, New York: Oxford University Press.
Love, A. and Nathan, M. J. [2015]: 'The Idealization of Causation in Mechanistic Explanation', *Philosophy of Science*, **82**(5), pp. 761-774.
Machamer, P. Darden, L. and Craver, C. [2000]: 'Thinking About Mechanisms', *Philosophy of Science*, **67**, pp.1-25.
Matthewson, J. and Calcott, B. [2011]: 'Mechanistic Models of Population-level Phenomena', *Biology and Philosophy*, **26**(5), pp. 737-756.
Morrison, M. [1996]: 'Physical Models and Biological Contexts', *Philosophy of Science*, **64**, pp. S315-S324.
Morrison, M. [2004]: 'Population Genetics and Population Thinking: Mathematics and the Role of the Individual', *Philosophy of Science*, **71**, pp.1189-1200.
Morrison, M. [2009]: 'Understanding in Physics and Biology', In Henk W. de Regt, Sabina Leonelli, and Kai Eigner (eds.), *Scientific Understanding: Philosophical Perspectives*, Pittsburgh: Pittsburgh University Press.
Morrison, M. [2015]: *Reconstructing Reality: Models, Mathematics, and Simulations*, New York: Oxford University Press.
Peters, D. [2013]: 'What Elements of Successful Scientific Theories are the Correct Targets for "Selective" Scientific Realism?', *Philosophy of Science*, **81**, pp. 377-397.
Pexton, M. [2014]: 'How Dimensional Analysis Can Explain', *Synthese*, **191**, pp. 2333-2351.
Pincock, C. [2011]: 'Modeling Reality', *Synthese*, **180**, pp. 19-32.
Potochnik, A. [2015]: 'Causal Patterns and Adequate Explanations', *Philosophical Studies*, **172**(5), pp. 1163-1182.
Potochnik, A. [forthcoming]: *Idealization and the Aims of Science*, Chicago: University of Chicago Press.
Rice, C. [2012]: 'Optimality Explanations: A Plea for an Alternative Approach', *Biology & Philosophy*, **27**(5), pp. 685-703.
Rice, C. [2015]: 'Moving Beyond Causes: Optimality Models and Scientific Explanation', *Nous*, **49**(3), pp. 589-615.
Rice, C. [2016]: 'Factive Scientific Understanding Without Accurate Representation', *Biology and Philosophy*, **31**(1), pp. 81-102.
Rice, C. [forthcoming]. 'Idealized Models, Holistic Distortions, and Universality', *Synthese*, DOI 10.1007/s11229-017-1357-4
Saatsi, J. and Pexton, M. [2013]: 'Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Non-causal Explanations', *Philosophy of Science*, **80**, pp. 613–624.
Stanford, P. K. [2003]: 'No Refuge for Realism: Selective Confirmation and the History of Science', *Philosophy of Science*, **70**, pp. 913-925.
Strevens, M. [2009]: *Depth: An Account of Scientific Explanation*, Cambridge: Harvard University Press.
Walsh, D., Lewens, T. and Ariew, A. [2002]: 'Trials of Life: Natural Selection and Random Drift', *Philosophy of Science*, **69**(3), pp. 452-473.
Weisberg, M. [2006]: 'Robustness Analysis', *Philosophy of Science*, **73**, pp. 730-742.
Weisberg, M. [2007]: 'Three Kinds of Idealization', *Journal of Philosophy*, **104**(12), pp. 639-659.
Weisberg, M. [2013]: *Simulation and Similarity: Using Models to Understand Our World*, New York: Oxford University Press.

Wimsatt, W. C. [1981]: ‘Robustness, Reliability, and Overdetermination’, In M. Brewer and B. Collins (eds.), *Scientific Inquiry and the Social Sciences*, San Francisco: Jossey-Bass: pp. 124-163.

Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Woodward, J. [2010]: ‘Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation’, *Biology and Philosophy*, **25**, pp. 287-318.