2015

Moving Beyond Causes: Optimality Models and Scientific Explanation

Collin Rice
Bryn Mawr College, crice3@brynmawr.edu

Custom Citation
Rice, Collin. 2015. "Moving Beyond Causes: Optimality Models and Scientific Explanation." Noûs 49.3: 589-615.

Moving Beyond Causes: Optimality Models and Scientific Explanation

Collin Rice

Lycoming College

Abstract. A prominent approach to scientific explanation and modeling claims that for a model to provide an explanation it must accurately represent at least some of the actual causes in the event's causal history. In this paper, I argue that many optimality explanations present a serious challenge to this causal approach. I contend that many optimality models provide highly idealized equilibrium explanations that do not accurately represent the causes of their target system(s). Furthermore, in many contexts, it is in virtue of their independence of causes that optimality models are able to provide a better explanation than competing causal models. Consequently, our account of explanation and modeling must expand beyond the causal approach.

1. Introduction.

Recently philosophers of science have begun to pay more attention to the building of idealized mathematical models (Batterman 2002a, 2002b, 2009; Bueno and Colyvan 2011; Odenbaugh, 2005; Pincock 2007a, 2011, 2012; Rice 2012; Weisberg 2007). An important example of idealized mathematical modeling is the widespread use of optimality models in biology and economics.[1] A prominent approach to scientific explanation claims that for a model to provide an explanation it must accurately represent causes of the explanandum (Craver 2006; Hitchcock and Woodward 2003; Kaplan and Craver 2011; Lewis 1986; Railton 1981; Salmon 1984; Strevens 2004, 2009; Woodward 2003).[2] Indeed, most philosophical accounts of how explanation works are explicitly causal. In this paper, I argue that the explanations provided by many optimality models present a serious challenge to this causal approach.

Since Hempel (1965) it has been widely accepted that a necessary condition for something to count as a scientific explanation is that it be (at least in some sense) *true*. A version of this requirement is present in contemporary causal theories of explanation (Lewis 1986; Salmon 1984; Strevens 2004, 2009; Woodward 2003). For these theories, to explain an event is

to *accurately represent some of the actual causes* (or causal mechanisms) in the event's causal history.[3] This striving for accurate representation of salient causes is also present in a prominent approach to using models as explanations (Craver 2006; Kaplan and Craver 2011; Potochnik 2007; Strevens 2009). For causal accounts, in order for a model to provide a satisfactory explanation it must provide a veridical representation of the salient causes of the target system(s). As Michael Strevens asserts, "no causal account of explanation—certainly not the kairetic account—allows non-veridical models to explain" (Strevens 2009, 320).[4] In addition, Kaplan and Craver argue that, "to explain the phenomenon, the model must…reveal the causal structure of the mechanism" (Kaplan and Craver 2011, 605). Indeed, this veridical representation requirement is made explicit in their model-mechanism-mapping (or 3M) requirement for explanatory models (Kaplan 2011; Kaplan and Craver 2011). According to this approach, the explanatory power of a model comes from its accurate representation of these salient causes.[5] Therefore, idealizations that distort or eliminate difference-making (or otherwise contextually relevant) causes may be justified pragmatically, but should ultimately be removed from the best explanation of a phenomenon.[6]

In contrast, other philosophers have recently argued that idealization and abstraction sometimes play an important role in scientific explanations precisely because they remove—or abstract away from—the various causal details of a system in order to capture the dominant (and sometimes noncausal) features that are responsible for a phenomenon (Batterman 2002b, 2009, 2010; Huneman 2010; Rice 2012; Weslake 2010). The goal of this approach is not to provide an accurate representation of any actual causes or causal mechanisms, but rather to utilize idealized mathematical models and complex derivation techniques in order to capture the dominant, and sometimes noncausal, features of the system(s) responsible for the explanandum. According to this view, sometimes a highly idealized model that does not accurately represent a system's causal details, "can *better* explain and characterize the *dominant* features of the physical phenomenon of interest. That is to say, these idealized models better explain than more detailed, less idealized models" (Batterman 2009, 429).[7] According to this approach, idealizations that distort or omit a system's causally relevant details can make important *contributions* to the model's ability to provide a particular kind of explanation. In other words, sometimes a model is a better explanation in virtue of, rather than in spite of, its being highly idealized and providing little (if any) accurate information about causes.

In this paper, I argue that many optimality explanations present a serious challenge to the causal approach. I contend that many optimality models provide explanations that do not accurately represent the difference-making (or otherwise relevant) causes of their target system(s). Indeed, the key features of optimality explanations move the model in the opposite direction suggested by the causal approach by eliminating (or drastically distorting) many of the target system's causally relevant details. To be clear, I will not be arguing against the claim that many (if not most) explanations in science are causal explanations. Nor will I argue that optimality models cannot be used to provide causal explanations. Rather, I will argue that often the way in which scientists use optimality models to provide explanations is in conflict with the requirements of causal approaches to explanation. Unfortunately, I believe philosophers have often inappropriately attempted to apply their primarily a priori approaches to explanation to these unique instances in biology rather than first considering how these models are actually used by scientists to provide explanations. Therefore, my approach will be to first analyze scientists' use of optimality models independent of any particular theory of explanation and then investigate how these cases fit with our current philosophical theories of explanation. For purposes of space, I will restrict my discussion to examples of optimality explanations from biology. Although I believe many of the claims made in this paper can be applied to optimality explanations within other disciplines (e.g. economics), because there are some important differences, a detailed analysis of those cases will have to be provided elsewhere.

The next section outlines some basic features of the explanations provided by optimality models. Then Section 3 presents examples of how scientists use optimality models to explain biological phenomena. From these examples, I identify three key features of an optimality explanation: equilibrium, idealization, and synchronic representation of structural features. Then, in Sections 4, 5 and 6, I argue that these three components contribute to the explanation in virtue of eliminating or distorting many of the causally relevant details of the target system(s). In addition, I will argue that optimality models provide highly idealized equilibrium explanations, even though they do not accurately represent the salient causes of the explanandum. Furthermore, in many instances, a model that accurately represented causes would actually provide a *worse* (or perhaps no) explanation. As a result, our account of explanation and modeling must expand beyond the causal approach. The final section provides some suggestions for how this expanded account can be formulated.

2. The Core Components of an Optimality Explanation.

Optimality models are distinguished by their use of a mathematical technique called Optimization Theory, whose goal is to identify which values of some *control variable(s)* will optimize the value of some *design variable(s)* in light of some design constraints (Beatty 1980; Maynard Smith 1978; Seger and Stubblefield 1996). An optimality model specifies a constrained set of possible strategies known as the *strategy set*. The design variables to be optimized constitute the model's *currency*.[8] An optimality model also specifies what it means to *optimize* these design variables (e.g. should a design variable be maximized or minimized). This is referred to as the model's *optimization criterion*.

Once the strategy set and optimization criterion have been identified, an optimality model describes an *objective function*, which connects each possible strategy to values of the design variable(s) to be optimized.[9] These equations will build in various context-specific design constraints and tradeoffs among the quantities represented within the model. For example, when building a bridge there may be several design features to be optimized—e.g. weight, width, cost etc. Yet, not all of these can be optimized simultaneously; certain tradeoffs (e.g. more width will require more weight and cost) and context-specific limitations (e.g. limited funds) will constrain the optimal design. These constraints and tradeoffs are built into the optimality model's objective function (or strategy set).

Once these components are specified, one can deduce which of the available strategies will yield the optimal value(s) of the design variable(s). The strategy that optimizes the model's criterion, in light of various constraints and tradeoffs, is deemed the *optimal strategy*. By mathematically representing the important constraints and tradeoffs, an optimality model can demonstrate why a particular strategy is the *best available solution*.

Now that we understand what an optimality model is, we can turn to the question of how an optimality model might be used to provide an explanation. The main thing to note is that merely showing that a strategy is the best available is insufficient to provide a satisfactory explanation. Therefore, if an optimality model is going to provide a satisfactory explanation of biological explananda—e.g. the current trait distribution of a biological population—the model

must also include assumptions about why the optimal strategy *is to be expected*. I refer to these additional assumptions as *optimization assumptions*.

In biology, optimality models usually provide a species of *equilibrium explanation* (Sober 1983).[10] For instance, in biology, an optimality model is often taken to explain the current state of the population by showing that the optimal strategy is (or is related to) the evolving system's equilibrium point. In the simplest cases, we can identify which of the available strategies will *maximize* some currency; e.g. fitness. In these "frequency-independent" cases, it is assumed that the system will tend to increase the model's currency; thereby making the strategy that maximizes the model's currency an equilibrium point. However, sometimes the optimal strategy will depend on what strategies other individuals in the population are playing (or the population as a whole).[11] In these *frequency-dependent* cases, game-theoretic techniques must be used (Lewontin 1961; Maynard Smith 1982). Here, the model's optimization assumptions will aim to show that the optimal strategy (or distribution of strategies) is an *evolutionarily stable equilibrium*, but this stable state may not maximize the model's currency.[12]

In either case, in order to provide a satisfactory explanation, an optimality model must: (1) accurately represent the salient constraints and tradeoffs of the target system(s) and, (2) make accurate optimization assumptions about the target system.[13] Precisely which aspects of these features need to be accurately represented (and to what degree) to provide an explanation will depend on the explanatory interests of the modeler (Matthewson and Weisberg 2009). Given this context specificity, I will simply say that an optimality model can provide a sufficient explanation when it adequately captures these aspects of the physical system *to the degree dictated by the explanatory context*.

Several accounts connect explanation with some notion of *expectability* (Hempel 1965; Salmon 1984; Strevens 2009; Batterman 2002). That is, an explanans explains the explanandum in part because it allows us to see why the explanandum was expected to occur. This minimal—though by no means sufficient—requirement for being an explanation provides some insight into why optimality models are able to provide satisfactory explanations. An optimality model uses mathematical representations to demonstrate that a particular strategy (or set of strategies) is the best available given certain constraints and tradeoffs—i.e., the mathematical model demonstrates why a particular strategy *is locally optimal*. The optimization assumptions then show *why the*

target system is expected to arrive at (and perhaps maintain) the optimal strategy. However, I will argue below that many optimality models are able to satisfy these requirements *without accurately representing causes of the target explanandum*.

3. Examples of Optimality Explanations.

In this section I present two examples in order to illustrate how biologists use optimality models to provide explanations.[14] Each example is representative of a larger class of optimality explanations commonly found in biology. In light of these examples, I identify three key features of an optimality explanation: equilibrium, idealization, and synchronic mathematical representations of structural features. In the following sections I analyze the contributions these features make to the explanation provided by the optimality model.

3.1. A Frequency-Independent Example: Parker's Dung Flies.

Biological optimality models typically assume that natural selection will maximize the model's criterion; e.g. fitness or its proxy. Furthermore, these models commonly assume that natural selection will be able to overcome any other evolutionary factors; e.g. drift. These optimization assumptions entail that the strategy that maximizes the model's currency is the equilibrium point of the evolving population. Another important feature of biological optimality models is that they are usually intended to be *adaptive explanations of population-level explananda*—e.g. why this population has evolved the adaptive trait it has.[15]

As an example, G. A. Parker utilized an optimality model in his attempts to explain why dung flies (*Scatophaga stercoraria*) copulate for 36 minutes on average (Parker 1978). Parker's model is an instance of a general class of biological optimality models that stems from the prey and patch choice models used in foraging theory (Charnov 1976; Krebs 1984; Stephens and Krebs 1986).[16] These foraging models analyze the tradeoff between energy intake from the current patch or food item and the lost opportunity to perform other foraging tasks. Parker's model uses this kind of analysis to calculate the optimal time for dung flies to spend copulating given that time spent copulating is time that cannot be spent on other mating tasks.

First, Parker observed that female dung flies mate with multiple males. He then discovered, by experimentation, that when this occurs the second male fertilizes far more eggs

(80%) than the first (20%). Consequently, after copulating, a male spends some time guarding the female before searching for other mates. The total behavioral cycle time is given by summing search time, copulation time and guard time. Parker then observed that the average time spent searching plus guarding was 156 minutes. Therefore, the total cycle will last 156 + *c* minutes, where *c* is the amount of time spent copulating. Different values of *c* constitute the model's strategy set.

By experiment, Parker found that increasing *c* increases the average number of eggs fertilized. However, there is an important tradeoff: time spent copulating is time that cannot be spent on other parts of the mating cycle. In addition, Parker observed diminishing returns on time spent copulating.[17] These constraints and tradeoffs are captured by the mathematical curve that represents average fertilization as a function of copulation time (Figure 1).

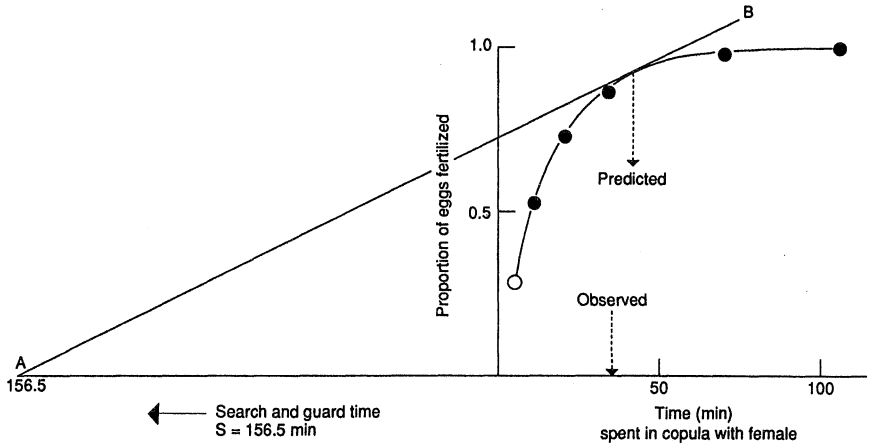

Figure 1: Parker's optimality model used to investigate the copulation time of dung flies (Sober 2000).

According to Parker's optimization criterion, the optimal value for *c* is the value that *maximizes the rate of eggs fertilized* across several iterations of the behavioral cycle. Consequently, the optimal strategy occurs at the point where a line that passes through the origin and intersects the asymptotic curve with the steepest slope (line A-B above) intersects the curve. This optimal strategy occurs when *c* is equal to 41 minutes, which is fairly close to the observed average value of 36 minutes. Given this predictive accuracy, and the fact that Parker's model was based on detailed empirical observations, the model is often thought to have captured the salient constraints and tradeoffs involved in the selection of the trait (Sober 2000).[18]

However, in order for Parker's model to provide an explanation of the observed trait value requires additional assumptions. First we must assume that a strategy's fitness is strictly increasing with increased (average) rate of egg fertilization—i.e. we must assume that Parker's optimization criterion is accurate. Furthermore, we require assumptions that eliminate the influence of other evolutionary factors; e.g. drift or genetic constraints. These assumptions are captured in various idealizations within the optimality model. For example, in order to eliminate drift the model assumes that the population is infinite. Other idealizations—e.g. assuming that phenotypes are inherited perfectly by offspring—result in the elimination of other evolutionary factors. These optimization assumptions entail that the optimal strategy identified by the model is the equilibrium point of the system. In this way, Parker's model can be used to provide an equilibrium explanation for why dung flies copulate for approximately 36 minutes on average.[19]

3.2. Explaining General Patterns: Equilibrium Sex Ratios.

In addition to explaining system-specific phenomena, optimality models are frequently used to explain highly general patterns. For example, life-history models investigate the optimal lifetime reproductive effort across widely heterogeneous populations (Williams 1966; Charnov et. al. 2007). Another commonly referenced group of biological optimality models comes from the theory of sex allocation (Charnov 1982; Fisher 1930; Hamilton 1967; Maynard Smith 1982; Maynard Smith and Price 1976). Although there are numerous examples to choose from, for my purposes, it is sufficient to consider Fisher's original (1930) model of equilibrium sex ratios. Although Fisher's model has been expanded (and criticized) in numerous ways, the general structure of the optimality explanation of a general pattern remains the same. The important point is that Fisher asked a very general biological why question: *why is the sex ratio often 1:1?*

In Fisher's model the strategy set includes all the possible points on a continuum from producing only male offspring to producing only female offspring. He used a simple frequency-dependent optimality model to demonstrate that if the sex ratio were controlled by phenotypic strategies played by parents (which are assumed to be directly determined by their genes), the stable equilibrium sex ratio would be 1:1.

In rough outline, Fisher's model claims the following. Fisher reasoned that if one sex is more common in the population, there will be a fitness payoff to parents who produce the

minority sex since their children will have more mating opportunities. To see this, suppose that male births are less common than female births. A newborn male thus has better mating prospects than a newborn female and is, therefore, expected to have more offspring on average. Therefore, parents who produce males tend to have more grandchildren on average. Consequently, male births will become more common in the population. However, as the population approaches a 1:1 ratio, this fitness advantage fades away. The exact same reasoning applies if we assume that female births are less common. The only state in which selection will not favor the production of the minority sex is when the sex ratio is 1:1. Therefore, a 1:1 sex ratio is the stable equilibrium state of the evolving population.

Fisher's model relies on a key tradeoff between the ability to produce sons and daughters. This tradeoff is described by looking at what economists refer to as the *substitution cost*. In this case, the substitution cost tells us how many sons can be produced if one less daughter is produced. In Fisher's model this tradeoff is perfectly linear—i.e. males and females cost the same amount of resources to produce and so one fewer son means one more daughter and vice versa. This is why his model predicts a 1:1 sex ratio. Charnov (1982, 28-9) shows that generalizing Fisher's model leads to the conclusion that the equilibrium ratio of a population, *r*, can be calculated using the following formula:

$$r = C_2/(C_1 + C_2)$$

Where C_1 is the average resource cost of one son and C_2 is the average resource cost of one daughter. In other words, the equilibrium sex ratio for the population is determined by the substitution cost, C_2/C_1; i.e., a parent can have one daughter, or C_2/C_1 sons. Therefore, the tradeoff between utilizing resources to produce sons and daughters is the key to explaining the equilibrium sex ratio.

In order for these models to provide an explanation, we must assume that natural selection will optimize the model's criterion—e.g., mating opportunities. Additionally, the models assume that other evolutionary factors (e.g. drift) will not deter the population from reaching the equilibrium favored by natural selection. In order to derive this result, these models usually make the idealizing assumptions that organisms reproduce asexually, mate randomly, each have equal access to resources for investing in offspring, the population is infinite in size,

etc.[20] As I will argue below, abstracting away from these heterogeneous details is key to providing an explanation of the kind of general biological pattern Fisher was interested in.

3.3. The Key Features of Optimality Explanations.

These examples are certainly insufficient to provide an account of all optimality modeling in biology. Still, in light of these examples, we can identify three key features common to most optimality explanations. First, these models usually provide a type of *equilibrium explanation*. Second, these models are typically *highly idealized*. Third, most of the explanatory work in these models is done by *synchronic mathematical representations of structural features of the system*.

According to the causal approach, in order to provide an explanation a model must accurately represent the (contextually or metaphysically) salient causes of, or causal mechanisms that gave rise to, the explanandum (Craver 2007; Lewis 1986; Railton 1981; Salmon 1984; Strevens 2004, 2009; Woodward 2003). However, in what follows I will argue that the key features of optimality explanations (those listed above) move the model in the opposite direction suggested by the causal approach. In addition, I will argue that these models are able to provide explanations despite the fact that they do not accurately represent causes of the explanandum.

4. Equilibrium Explanation.

Most of the optimality models in biology are used to provide equilibrium explanations. Part of what is explanatory about this kind of equilibrium model is that it allows us to understand that no matter what the particular causal trajectory of the population had been, within a certain disjunction of possible causal trajectories, the ultimate result would have been the equilibrium state. That is, the model tells us that a wide range of potential causal trajectories would all yield the same result, *but it does not tell us which one is the actual causal trajectory or initial state of the target system* (Sober 1983).

One way to understand the explanatory power of this kind of equilibrium explanation is that we know that the actual causal trajectory of the system is within the disjunction of possibilities that the model shows will lead to the same equilibrium. Michael Strevens argues that an equilibrium explanation of a ball's arriving at the bottom of a basin is causal in this way. He says, "But it is not true that the explanation says nothing at all about the ball's starting point

and trajectory. It identifies the starting point as one of a large class, namely, all starting points at the basin's lip, and likewise identifies the trajectory as one of an equally large class" (Strevens 2009, 288). It is true that the equilibrium model gives us some minimal information about the actual causal history of the system by telling us it is within the disjunction of causal trajectories that would lead to the explanandum. However, I maintain that grounding the model's explanatory power in this minimal information about the actual causes is a mistake. For one thing, it suggests that a model that provided additional information about the actual causes would provide a superior explanation. Indeed, Strevens claims that models in ecology and economics that "black-box" the dynamics of underlying causal mechanisms are at best partial explanations. He says, "Because a model that secretes some mechanisms in its explanatory framework does not confer…'deep' or 'full' understanding of the phenomenon...the black-boxing model is limited in its explanatory power. A deep explanation…must flesh out the model's black boxes rather than leaving the causal details in the framework" (Strevens 2009, 169).[21] Other causal accounts of explanation also favor explanations that fill in the details of micro-level causal mechanisms (e.g. see Kaplan and Craver 2011 and Kaplan 2011).

In contrast, I maintain that there are several reasons for preferring the kind of equilibrium explanations provided by optimality models in virtue of the fact that they *do not cite the actual causal trajectories or mechanisms of particular systems* (Batterman 2002b; Garfinkel 1981; Pincock 2007b, 2011; Weslake 2010). One reason is that citing the actual causal trajectory (or mechanisms) of a particular system means that the optimality model is unable to apply to other systems subject to similar constraints and tradeoffs, but which are heterogeneous in these causes. Many philosophers have recognized the explanatory value of this kind of modal information (Garfinkel 1981; Jackson and Pettit 1992; Woodward 2003). By moving away from the actual causes of any particular system (or set of systems), the optimality model is able to capture a wider range of possible systems that are extremely heterogeneous in their causes.

Here, I do not intend to suggest that an explanation's applying to more possible systems is always objectively better.[22] Rather, I agree with Elliott Sober that the preference for generality or detail is largely a matter of our explanatory interests (Sober 1999). However, in many cases the generality provided by optimality models' exclusion of causal details is what we want. Indeed, sometimes the generality of the explanandum *requires* an explanation that is independent

of the initial conditions and causal trajectories of particular systems. For example, in order to explain evolutionary patterns across causally heterogeneous populations—e.g. the frequency of the 1:1 sex ratio—often the causal details of particular systems need to be eliminated.

In addition, the explanations provided by optimality models are extremely enlightening precisely because they are independent of information about the actual causal trajectories of particular systems, not in spite of this feature. For one thing, it is an important explanatory piece of information that the particular initial conditions and causal trajectory are not required for the target explanandum to occur. As Jackson and Pettit (1992, 177) put it, the explanation tells us that, "if the actual history described by the microcausal explanation had not obtained, the explanandum would still have occurred."[23] Similarly, when discussing Fisher's model, Elliott Sober claims, "Where causal explanation shows how the event to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired" (Sober 1983, 202). Although the actual causal trajectory of the system includes many difference-making (i.e. causally relevant) causal factors, the optimality model explains without referencing those causes. Instead, the optimality model provides us with an important piece of explanatory information: *the initial conditions and causal trajectory of the target system are not important for understanding why the target explanandum occurred because several different causal histories would have led to the same outcome.* In this way, the optimality model provides additional explanatory information that would not be captured by a model that accurately represented the causal trajectories or initial conditions of particular systems.

Eliminating these causal details also allows our explanation to focus on the features of the target system(s) that *are* essential for understanding the target phenomenon. Once we understand that the particular initial conditions and causal trajectory of the system are not important, we can better appreciate the things that really matter: the structural constraints and tradeoffs that the optimality model mathematically represents and the optimization assumptions used to deduce the target explanandum. Therefore, in some instances, optimality models' elimination of causally relevant details allows them to provide a better explanation than models that would "fill in the details" of the actual causes.

5. Idealizations in Optimality Explanations.

Yet even though optimality explanations do not reference the actual initial conditions or causal trajectory of their target system(s), they may still provide a different type of causal explanation by accurately representing some causal mechanisms (or processes) that gave rise to the explanandum. That is, the model may not tell us precisely how the causal history actually proceeded, but may still accurately represent some of the causal relationships (e.g. mechanisms) responsible for the explanandum.[24] However, even if optimality models can be shown to be a kind of causal model, they will almost always fail to meet the veridical standards required by causal approaches to explanation. This is due to the fact that optimality explanations are almost always so highly idealized that *they fail to accurately represent the (salient) causes of their target system(s)*.

To begin, biological optimality model's mathematical representations of constraints and tradeoffs are often inaccurate when compared to the causal mechanisms acting within the population. For example, in Parker's model the assumption is that average fertilization rate increases with increased time spent copulating according to a perfectly asymptotic curve. In addition, many game-theoretic models assume that a conflict is symmetric or that payoffs are constant across iterations of the game in order to use certain mathematical operations required to deduce the equilibrium state (see Hammerstein and Selten 1994 for some examples). But these smooth curves and constant parameters will almost always be inaccurate when compared with the causal mechanisms acting within the model's target system(s).[25]

Next, the strategy sets of most optimality models do not accurately represent the set of strategies actually causally interacting within the system.[26] The model's strategy set, however, is not intended to accurately represent the strategies that were causally interacting in the population, but merely aims to capture the *relevant set of alternative strategies* for the optimization problem. For instance, although a sex ratio model may assume that the strategy set includes all the probabilities that a birth will be a male, this is not intended to claim that in the target population there were individuals playing this range of strategies. Rather, the model assumes that the optimal strategy will evolve *regardless* of which particular distribution of strategies was actually causally interacting within the population's history (as long as this set includes the optimal strategy). The actual strategies are, however, *causally* relevant to the

evolutionary process that occurred. This causal information, however, is explicitly *not* included in the explanation provided by the optimality model.

In addition, a biological model's optimization assumptions are almost always inaccurate when compared with a population's causal mechanisms—even if we consider causal mechanisms at the level of the population. For instance, in foraging models it is often assumed that natural selection will maximize average energy intake. However, this is only one thing that might influence the survival and reproduction of organisms in the population. In fact, the various causal mechanisms (or more generally natural selection) acting in the population often will not optimize the model's criterion. Therefore, in most cases, the optimization process described by the optimality model will not accurately represent any actual causal process (or mechanism) of the target system. The goal of a biological optimality explanation, however, is not to accurately describe the causal dynamics that led to the evolution of a trait. Rather, biological optimality models identify optimal strategies (or states) that are only outcomes or end states of an evolving system that approximates their optimization assumptions in the long run. To borrow a weighted phrase from economics, a model's optimization assumptions are usually taken to be adequate as long as the system behaves "as if" it where optimizing the model's criterion. Therefore, optimality models are frequently used to explain why a system has evolved the optimal strategy *regardless of whether their optimization assumptions are true of the causal processes acting in the system at any point along the way to that strategy*.

Next, many biological optimality models use idealizations concerning the way in which phenotypic strategies are inherited; e.g. that strategies are inherited perfectly by offspring. These idealizations are introduced because it is assumed that changing these inheritance assumptions will have no effect on the occurrence of the phenomenon. In other words, the actual causal processes that underlie these kinds of inheritance assumptions are not important for understanding why the phenomenon occurred. Therefore, we can ignore these causes, or at least specify them inaccurately, without our model losing any explanatory power. However, the causes underlying these inheritance assumptions are important difference-makers in the explanandum's causal history in at least one sense—without some causal process of inheritance nothing would have evolved. Nevertheless, an accurate representation of these causes is not required for the optimality model to explain the target phenomenon.

Another idealization is the assumption that environmental pressures are constant. This idealization is required in order for the optimal strategy to be an obtainable equilibrium state—i.e. the population needs time to arrive at the predicted optima before changes in environmental conditions alter the equilibrium point of the system. However, the causal selection pressures in a population are never constant in this way. Therefore, this idealization grossly distorts the causal mechanisms actually influencing the selection of the trait. However, accurately representing these causal mechanisms is not important to the explanation provided by the optimality model.

Finally, most biological optimality models assume that the population being modeled is infinite (or effectively infinite). However, population size *does* make a difference to every evolutionary process; i.e., drift is a statistical fact of every real-world biological population. Assuming infinite population size has the effect of eliminating drift from the model by utilizing various laws of large numbers. By incorporating the idealization of infinite population size, the optimal trait according to natural selection is what we expect to evolve. Therefore, this idealization, like many of the others, is vital to the optimality model's explanation—without it the features of the model do not entail the occurrence of the explanandum. However, once again, this idealization eliminates (or drastically distorts) the causal mechanisms (or processes) operating within the model's target population(s).[27]

In sum, I have identified six kinds of idealizations frequently utilized in biological optimality explanations:

(1) The model's mathematical curves, equations, or payoff structures are often idealized when compared to the causal processes within the target system(s)).

(2) Idealized strategy sets are intended to capture the relevant alternatives rather than strategies actually causally interacting within a population.

(3) The models' optimization assumptions do not accurately represent a causal mechanism in the system, but only captures the general optimizing tendency of the system in the long run.

(4) There are idealizations regarding causal mechanisms of inheritance.

(5) It is assumed that selection pressures do not change over time.

(6) Infinite population size is assumed to allow for the use of various laws of large numbers in deducing the target explanandum.

When considered together, these idealizations entail that optimality models usually provide little, if any, accurate information about the actual causes, or causal mechanisms, within the model's target system(s). In the end, the highly idealized optimality model represents mathematical relationships between constraints, tradeoffs, and the system's equilibrium point that *do not mirror any causal relationships (or processes) in the target system.* Put different, optimality models fail the kind of "model-to-mechanism-mapping requirement" of causal theories that requires that the "dependencies…among variables in the model correspond to the…causal relations among the components of the target mechanism" (Kaplan and Craver 2011, 611).

The key point here is to recognize that rather than omitting irrelevant causal factors so that we can focus on those causes that make a difference—as say Strevens's (2009) account of idealization would have it—I argue that these idealizations *considered collectively* move us away from attempting to accurately represent causes *at all* towards a representation of relationships that do not mirror any of the causal relationships within the model's target system(s).

In a similar way, James Woodward describes how the ideal gas law, "Abstracts radically from the details of the causal processes involving particular individual molecules and instead focuses on identifying high-level variables that aggregate over many individual causal processes that figure in the general patterns that govern the behavior of the gas" (Woodward 2003, 354). In addition to abstracting radically from individual-level causal processes, I argue that optimality models are typically so idealized that they provide little (if any) accurate information about any of the causes within the model's target system(s)—even if we consider causes at the "macro" level. Instead, the key relationships (what Woodward calls patterns involving high-level variables) that are the focus of many highly idealized optimality models do not correspond to causal relationships within the model's target system(s).

Moreover, many optimality models utilize idealizations that appear to play *essential* roles within their explanations. Robert Batterman (2002, 2009, 2010) has recently described similar cases in physics. According to Batterman, idealizations that introduce limits are sometimes

essential to an explanation because they allow for certain mathematical operations to be performed that would not otherwise apply. In many cases, however, the explanation can only be provided by the idealized model—lose the idealizations and you lose the explanation. Similarly, the explanations provided by at least many optimality models require that certain idealizations be introduced in order to employ certain mathematical techniques used to derive the target explanandum. For instance, the biological examples described above used the idealizing assumptions that the population is infinite, organisms mate randomly, phenotypic strategies are inherited perfectly, etc. Without these idealizations, the features represented in the mathematical model are insufficient for deriving the target explanandum. Even more idealizations are required in many game-theoretic explanations of highly general biological patterns. Therefore, for many optimality explanations it is unclear how the various idealizations could be removed from the model without consequently eliminating the explanation being offered.

By moving away from (even attempting to provide) an accurate representation of causes we are able to provide an explanation that applies to systems in which these causes are different. That is, by introducing various idealizations that distort or eliminate the causal details of particular systems, we are able to provide an explanation that captures a wider range of possible systems.[28] In addition, this kind of explanation is precisely what is *required* in order to explain highly general patterns that range over systems that are heterogeneous in most of their causal details. In order to explain biological patterns often requires assumptions that move us away from the heterogeneous causes of particular systems.

In addition, by showing that the causes of particular systems are irrelevant, the explanation of the highly idealized optimality model contains explanatory information about what is *not* required for the occurrence of the explanandum. Indeed, these idealizations are often introduced because the various causes of particular target systems are not important to the explanation provided by the optimality model. In other words, *these idealizations make essential contributions to the explanation provided by an optimality model because they show why most (if not all) of a system's causal factor(s), mechanism(s), or variable(s) are not important for understanding why the explanandum occurred.* Although many of these details are relevant to the veridical causal explanation, they are not required for the optimality model to explain the phenomenon—this is because extremely different physical causes would have been sufficient so long as the constraints, tradeoffs, and optimization assumptions of the optimality model are

satisfied. This modal (i.e. counterfactual) information is key to explaining many of the explananda we observe (e.g. repeatable patterns) since it shows us why extremely heterogeneous causal systems will nevertheless display similar behavior.

What is more, these idealizations again aid our grasping of the key structural relationships that are important for understanding the phenomenon. By moving away from the causes of the model's target system(s) we are able to see how the dominant constraints and tradeoffs of the target system(s) are responsible for the explanandum. The relevance of these features is demonstrated by providing counterfactual information about how the system would behave if these structural relationships had been different in various ways. In this way, if an optimality model adequately captures the dominant structural features (to the degree dictated by the context), the model can provide a highly idealized equilibrium explanation despite the fact that it fails to accurately represent the causes of its target system(s).

6. Synchronic Mathematical Representations of Structural Features.

A third essential component of an optimality explanation is the mathematical representation of structural constraints and tradeoffs. The optimality model explains by showing how the equilibrium point of the system is counterfactually related to these constraints and tradeoffs; i.e. the model shows how changing the constraints or tradeoffs results in a change in the predicted outcome. These counterfactual relationships are key to the ability of these features to explain the target phenomenon. Given these counterfactual relationships, one might argue that these relationships are causal relationships after all since we can see how manipulating these structural features would change the equilibrium point of the system. Consequently, optimality models might be understood as providing a kind of (perhaps non-veridical) causal explanation. In response to this challenge, I will argue that—independent of their ability to veridically represent causes—when we look closer, the key counterfactual relationships within an optimality explanation are best interpreted as noncausal relationships.

To begin, it is often difficult to see how the tradeoffs represented within optimality models can be understood as causal relationships. For instance, in many biological optimality models, average energy intake and average predation risk will exhibit a tradeoff that is vital to the explanation of the observed phenotype, but it is unclear how we ought to understand the ontological claim that average energy intake is a *cause* of average predation risk (or vice versa).

Although there certainly is a "lower-level" causal story about why this tradeoff between population-level averages holds, the biological model does not reference these causes. Instead, the model focuses on relationships between multiply realizable features that aggregate over these causes. Moreover, within an optimality explanation it is really *the tradeoffs between variables* that are doing the key explanatory work. Indeed, as Eric Charnov claims, "the tradeoffs themselves are the fundamental objects of evolutionary interest, at least with respect to stabilizing or equilibrium selection" (Charnov 1989, 115). This is precisely why biologists have provided detailed analyses of how "evolutionary outcomes…*depend* on the shape and position of the tradeoff curves constraining the course of evolution" (de Mazancourt and Dieckmann 2004, 769).[29] In other words, it is in some sense the *syntactic* features (i.e. the relationships between variables), rather than the semantic features (i.e. the variables themselves) that do the explaining. However, even if two variables are causes, this does not entail that the tradeoff that exists between them is a cause. Indeed, it would be rather puzzling to claim, for example, that the population-level tradeoff between average energy intake and average predation risk is a cause of anything. In short, it is extremely difficult to see how our metaphysical intuitions about causes can be codified in the case of the tradeoffs that are central to optimality explanations.

In addition, many modelers within biology (and economics) appear to be uninterested in using their mathematical models to establish causal claims about the target system. While some contexts will certainly require reference to causes (or causal explanations), I think it is important to notice that scientific modelers often do not believe they require any metaphysical claims about causation in order for their mathematical models to aid in understanding the target phenomenon.[30] Indeed, I believe a welcome result of expanding beyond causal approaches to explanation is that scientists can provide explanations without having to make any ontological commitments about causal relationships between variables in their highly idealized models.

Still, many philosophers will remain unconvinced by these appeals to causal intuitions and the metaphysical agnosticism of some scientific modelers. A stronger argument for the noncausal interpretation of these relationships is that several of the key features common to causal explanations are absent in the instance of an optimality model's representation of tradeoffs.[31]

First, we can consider what Nancy Cartwright (2004) calls 'thick' causal concepts. After rejecting several attempts to provide a universal account of causation, Cartwright intends these concepts to help reveal the *productive relation* that is key to how causes relate to their effects.

These concepts include pushing, feeding, opening, compressing, unwinding, bonding, etc. However, the tradeoffs between statistical properties (e.g. averages) that often feature prominently in optimality explanations surely cannot be described in terms of compressing, pulling, or unwinding anything in the target system. Indeed, no such causal activity or causal concepts are employed in the optimality model's description of these structural features—nor anywhere else in the model's explanation.

Another reason to think that the constraints and tradeoffs of optimality explanations do not represent causal relationships is that the optimality model's mathematical representation of them *does not reference any processes, or events, that unfold prior to the explanandum.* A central feature of causal representations is that they are essentially *diachronic*—i.e. there is a temporal dimension to the representation that captures changes over time. This diachronic component is especially prominent in process and mechanistic theories of causal explanation (Craver 2006; Kaplan and Craver 2011; Machamer, Darden, and Craver 2000; Salmon 1984; Strevens 2009). The explanation provided by an optimality model, in contrast, merely identifies the optimal strategy by showing that the model's currency is optimized by a particular strategy, given the constraints and tradeoffs *synchronically* represented within the model. For instance, in Parker's model, the optimal strategy— i.e. the point at which the average rate of fertilization is maximized—is simply represented as the value of c at which the slope of a line that intersects the curve and passes through the origin is maximized. Nowhere does the model describe a causal process (or causal trajectory) that unfolds over time or any events that occur prior to the explanandum. Instead, the model merely identifies $c = 41$ minutes as the optimal strategy. Furthermore, none of the points along these mathematical curves need be instantiated in order for the model to explain the outcome. This is because, contrary to the standard way causes explain their effects, *optimality models do not provide a dynamical account of the processes or events that led to the explanandum*. Instead, these models identify optimal strategies by using synchronic representations of structural features of the system. So although dynamics are the main focus of causal explanations, this is precisely the kind of information that is (almost) entirely absent from an optimality explanation.

Finally, we can consider modularity and interventions. Interventionist accounts of causal explanation require that causes be *modular* in the sense that they can be manipulated independently of other causes within the system (Woodward 1997, 2003, 2010). Yet, in evolving

biological systems, the tradeoffs between (what are often statistical) higher-level properties of the system usually depend on the fitnesses of individuals and these variables in turn depend on a complex and integrated network of causal contributions to "fitness". Moreover, in many cases, these higher-level properties arise from complex systems whose dynamics are chaotic, non-linear, and involve feedback loops (Mitchell 2008, 2012).[32] Indeed, as Cartwright notes, in many cases "the causal laws are harnessed together and cannot be changed singly" (Cartwright 2004, 811).[33] Given the causal entanglement and complex integration of evolving biological systems, it is unlikely that one would (even in principle) be able to intervene in such a way that changed only a particular tradeoff's influence on the target phenomenon. Thus, it is extremely difficult to see how we could even in principle manipulate these tradeoffs' influence on the equilibrium point of the population independently of other causal factors. For instance, the key tradeoff in Parker's model is that time spent copulating is time that cannot be spent on other parts of the behavioral cycle. Intervening on this tradeoff would presumably require not only altering the causes that impinge on each individual dung fly, but also some kind of alteration to the basic principle that time spent on one task cannot be spent on other tasks. Precisely what this kind of (in principle) intervention would even look like is unclear.[34] Instead, I maintain that the apparent modularity of these relationships *within the idealized model* is merely an illusion created by the use of several idealizations and abstractions that eliminate the complexity of the causal networks of real-world biological systems. Modular relationships within an idealized mathematical model cannot establish modularity of causal relationships in the model's target system(s).[35]

Most importantly, however, optimality models are able to provide satisfactory explanations *without adding any claims about interventions or causation.* Nowhere in the description of the explanations above did we require any mentioning of these concepts. Indeed, additional claims about causation would be *otiose* in these cases (in the same way that claims about 'optimality' would be useless in many causal explanations). As a result, it appears that an optimality model need not tell us how things would have been different under an intervention in order to explain why we observe the explanandum. Although claims about interventions may be important for testing (or interpreting) causal claims, they are not required to establish that *the explanation* is sufficient. I will say more about the possibility of discharging this requirement from our account of explanation in the next section.

In sum, not only is it difficult to see how the relationships represented within optimality

models can be made to square with our metaphysical intuitions about causes, but also several of the key features of causal explanations are absent in the instance of an optimality model's representation of these structural features. Therefore, I conclude, *the key relationships represented within an optimality explanation are best interpreted as noncausal relationships.* Optimality models primarily focus on noncausal counterfactual relations between structural features and the system's equilibrium point. Moreover, these features can sometimes explain the target phenomenon without requiring any additional causal claims about the relationships represented in the model—i.e. *the explanatory claim and the causal claim are independent of one another.*[36]

Indeed, an explanatory strength of many optimality models is that their descriptions of these noncausal relationships are able to capture a range of systems that behave similarly despite being extremely heterogeneous in their causes. This explanatory goal is made explicit by some biological optimality modelers (Stephens and Krebs 1986).[37] This generality is yielded by the fact that the structural features represented within an optimality model are *multiply realizable*. Indeed, one reason optimality models are able to explain phenomena across extremely causally heterogeneous biological systems is because they focus on relationships that are invariant with respect to changes in the causes of the system. Put differently, moving away from causes is often what enables optimality models to provide the kind of explanation we seek.[38]

7. A Way Forward: Moving Beyond Causes.

7.1. Beyond The Causal Approach.

According to the causal approach, in order to explain an event a model must accurately represent the salient causes of, or causal mechanisms that gave rise to, the target explanandum (Craver 2007; Kaplan and Craver 2011; Lewis 1986; Railton 1981; Salmon 1984; Strevens 2004, 2009; Woodward 2003). This causal paradigm dominates the current discussion about explanation and modeling.

The above analysis has identified three key features of optimality explanations: equilibrium, idealization, and synchronic mathematical representation of structural features. I have argued that these features contribute to the optimality explanation by moving us away from

an accurate representation of the salient causes of the system(s). The three lines of argument that draw upon these features can be summarized as follows:

(1) Equilibrium explanations are often enlightening in virtue of showing us why causal dynamics are largely irrelevant.

(2) Highly idealized optimality models fail to meet the veridical requirements of the causal approach because they provide little (if any) accurate information about the causes within the model's target system(s).

(3) Instead of trying to accurately represent causal relationships, an optimality model focuses on synchronically representing noncausal (counterfactual) relationships.

In other words, the key features of an optimality explanation are ones that *move the model in the opposite direction suggested by the causal approach* by eliminating many of the target system's causally relevant details. Furthermore, the important work in these explanations is done by the structural relationships between a system's constraints, tradeoffs, and optimal strategy. These structural relationships, however, are best understood as noncausal relationships. By showing how the equilibrium state is counterfactually related to (i.e. depends on) these multiply realizable structural features, optimality models are able to provide an explanation that is independent of the causes of any particular system(s).

Moreover, in some contexts, a model that accurately represented the salient causes of the explanandum would actually provide a worse explanation. By moving away from causes, an optimality model will often: (1) apply to more possible systems, (2) provide explanatory information about what is not required for the explanandum, (3) provide explanatory information about why the phenomenon would occur in other causally heterogeneous systems, and (4) highlight the structural relationships essential to understanding why the explanandum occurred.

I conclude that optimality models provide a special kind of highly idealized equilibrium explanation despite the fact that they do not accurately represent the causes of the explanandum. Moreover, in many cases, it is in virtue of not accurately representing causes that optimality

models are able to provide a better explanation than competing causal models. Consequently, our account of explanation and modeling must expand beyond the causal approach.

7.2. A Way Forward: Batterman and Woodward.

This, of course, cannot be the end of the story. I am certainly not alone in suggesting that noncausal features of a system can be used to provide scientific explanations (Batterman 2002b, 2005, 2009; Bokulich 2011, 2012; Matthen and Ariew 2009; Pincock 2007a, 2012; Walsh et al. 2002; Walsh 2007; 2010). However, advocates of causal accounts of explanation have provided extended treatments of how causal explanations work. Given my conclusion, one could reasonably complain that I have created a problem without proposing a solution. Although providing a full account of (noncausal) explanation is beyond the scope of this paper, I do not think the revision will need to be as drastic as it might first appear. Importantly, many of the features of our existing accounts of explanation—e.g. counterfactuals and invariance (Woodward 2003), generality (Hitchock and Woodward 2003; Kitcher 1981; Strevens 2009), asymmetry (Salmon 1984), truth (Hempel 1965), expectability (Batterman 2002; Hempel 1965; Strevens 2009) etc.—are not restricted to a causal interpretation. Therefore, there is already a widely accepted set of features from which this alternative (and perhaps pluralistic) account of explanation can be constructed. In order to lay some groundwork for this project, I will provide some suggestions for how features of Batterman's (2002b) account of asymptotic explanations can be combined with features of Woodward's (2003) account of explanation.

To begin, my account of idealizations in optimality models agrees with many of the features emphasized by Batterman's account (Batterman 2002a, 2002b, 2005, 2009). Batterman identifies two features of what he calls 'universal behavior':

> 1. The details of the system (those that would feature in a complete causal-mechanical explanation of the system's behavior) are largely irrelevant for describing the behavior of interest.
> 2. Many different systems with completely different "micro" details will exhibit the identical behavior. (Batterman 2002, 73).

The key to explaining these universal behaviors, Batterman argues, is often to utilize idealizations (e.g. the thermodynamic limit) in order to show that the heterogeneous details of the systems are irrelevant for the occurrence of the repeatable phenomenon. For example, "[i]n order to understand why the thermodynamic laws and relationships hold for a variety of physically distinct types of systems, we require a method that allows for a systematic abstraction from the

details that distinguish the different systems—the realizers—from one another" (Batterman, 2002, 124). By showing how the use of limits is sometimes essential to this process, Batterman provides tools for understanding how idealizations can provide information about why the causal details of different systems are not important for understanding the phenomenon of interest. In addition, Batterman's account shows how idealizations are sometimes key to providing the generality we want. The idealized model explains the repeatable phenomenon by showing that the phenomenon depends on a few dominant features of the system—e.g. the critical exponent—and that this behavior is independent of various changes in the causal details of the system. Without these idealizations the explanation of this universality disappears. However, one thing that is key to add to Batterman's account of idealization is an account of why the resulting mathematical relationships are able to provide sufficient explanations.[39]

At this point, James Woodward's account of explanation is extremely useful. According to Woodward, an explanation can be understood as providing information about a pattern of counterfactual dependence between the explanans and the explanandum (Woodward 2003, 11). This involves answering questions about "what-if-things-had-been-different" or "w-questions". Accordingly, "[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, 11).

Although I like much of Woodward's account of explanation, I disagree with his requirement that these counterfactual relations be understood along strictly manipulationist or interventionist lines. The requirement that these counterfactuals must enable one to, in principle, *intervene* in the system restricts Woodward's account to specifically causal explanations.[40] However, I think it is a mistake to require that all scientific explanations must be causal. Indeed, if one looks at many of the explanations offered by scientific modelers, causes are not mentioned. Furthermore, as I have argued here, there are several cases in which a causal interpretation of the key explanatory relationships is inappropriate. One response is to try and expand our notion of causation in order to capture these cases. However, I suggest an alternative approach. Rather than expanding our account of causation far beyond our metaphysical intuitions, I suggest that we simply distinguish the question of what causation is from the question of what explains.[41] In line with this suggestion, Woodward himself appears to admit that perhaps not all explanations are causal and suggests a way that his account might be

expanded to include these additional cases:

> One natural way of accommodating these examples is as follows: the common element in many forms of explanation, both causal and noncausal, is that they must answer what-if-things-had-been-different questions. When a theory tells us how Y would change under interventions on X, we have (or have the material for constructing) a *causal* explanation. When a theory or derivation answers a what-if-things-had-been-different question but we cannot interpret this as an answer to a question about what would happen under an intervention, we may have a noncausal explanation of some sort. (Woodward 2003, 221)

In precisely this way, we can retain Woodward's emphasis on providing information about counterfactuals without requiring that these be *causal* counterfactual relations. That is, the counterfactual requirement on explanations is independent of the causal (i.e. interventionist) requirement. Indeed, in light of the previous discussion, I contend that *in some cases counterfactual information can be explanatory without tracking any relationships of causal dependence.*[42]

In addition, I maintain that explanations (of at least many explananda) ought to provide two kinds of counterfactual information by showing us both what *is* and *is not* important for understanding the target phenomenon.[43] This two-part requirement is why we require pieces from Batterman's account as well. In many cases, idealizations are essential to providing important counterfactual information about *why causes are irrelevant for describing the behavior of the system(s).* However, an explanation should also tell us about what features of the system(s) the explanandum does depend on. Here the model's representation of counterfactual relationships—that are sometimes noncausal—is the key to showing *what is important*.

Together these components are able to demonstrate a kind of *invariance* (another key feature of Woodward's account). In the case of noncausal explanations, the model (typically) shows how certain noncausal relationships between the explanans and the explanandum are invariant with respect to changes to the causes of the system(s). This kind of invariance is especially important for answering a particular type of why question—e.g., why do we see this phenomenon repeated across extremely causally heterogeneous systems? A satisfactory answer involves showing both what the phenomenon depends on as well as why it does not depend on the causes that are heterogeneous among those systems. So although there are plenty of why questions that require accurate representation of causes, in many cases we are interested in answering why questions about phenomena that are repeatable over extremely causally heterogeneous systems—and these why questions often require a different kind of answer.

8. Conclusion.

I have argued that many optimality models provide highly idealized equilibrium explanations despite the fact that they do not accurately represent causes. Moreover, in many cases, it is in virtue of not accurately representing causes that optimality models are able to provide a better explanation than competing causal models. Consequently, our account of explanation and modeling must expand beyond the causal approach. I have also suggested ways in which this expanded account might be constructed. Explanations should provide two kinds of counterfactual information by showing us both what *is* and *is not* important for understanding the target phenomenon. Idealizations are often essential to a model's ability to provide this information; e.g. when systems' causes are irrelevant to understanding the phenomenon. In addition, idealized models often explain by representing noncausal invariance relationships. Our revised account of explanation and modeling must incorporate these features if it is to provide an adequate account of how idealized mathematical models provide explanations—especially in biology.

Acknowledgements: Previous versions of this work were presented at the Philosophy of Biology at Madison Workshop and the University of Pittsburgh's Center for the Philosophy of Science. I would like to thank both audiences for their helpful feedback. I would also like to thank Robert Batterman, James Woodward, Denis Walsh, Christopher Pincock, André Ariew, Paul Weirich, and an anonymous referee for helpful comments and feedback on earlier versions of this work.

References:

Batterman, R. W. (2002a) 'Asymptotics and the role of minimal models,' *The British Journal for the Philosophy of Science, 53*(1), 21-38.

Batterman, R. W. (2002b) *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*, Oxford: Oxford University Press.

Batterman, R. W. (2009) 'Idealization and modeling,' *Synthese, 169*(3), 427-446.

Batterman, R. W. (2010) 'On the explanatory role of mathematics in empirical science,' *The British Journal for the Philosophy of Science, 61*, 1-25.

Beatty, J. (1980) 'Optimal-design models and the strategy of model building in evolutionary biology,' *Philosophy of Science*, 532-561.

Bokulich, A. (2011) 'How scientific models can explain,' *Synthese*, *180*, 33-45.

Bokulich, A. (2012) 'Distinguishing explanatory from nonexplanatory fictions,' *Philosophy of Science*, *79*, 725-737.

Bueno, O., and Colyvan, M. (2011) 'An inferential conception of the application of mathematics,' *Noûs, 45*(2), 345-374.

Cartwright, N. (2004) 'Causation: One word, many things,' *Philosophy of Science, 71*, 805-819.

Charnov, E. L. (1982) *The theory of sex allocation*, Princeton, NJ: Princeton University Press.

Charnov, E. L. (1989) 'Phenotypic evolution under Fisher's fundamental theorem of natural selection,' *Heredity*, *62*, 113-116.

Charnov, E. L., Warne, R. and Moses, M. (2007) 'Lifetime reproductive effort,' *The American Naturalist*, *170,* E129-E142.

Craver, C. F. (2006) 'When mechanistic models explain,' *Synthese, 153*(3), 355-376.

de Mazancourt, C. and Dieckmann, U. (2004) 'Trade-off geometries and frequency-dependent selection,' *The American Naturalist*, *164*(6), 765-778.

Fisher, R. A. (1930) *The genetical theory of natural selection,* Oxford: Clarendon Press.

Fodor, J. (1974) 'Special sciences; or, the disunity of science as a working hypothesis,' *Synthese, 28*(2), 97-115.

Garfinkel, A. (1981) *Forms of explanation: Rethinking the questions in social theory*, New Haven, CT: Yale University Press.

Hamilton, W. D. (1967) 'Extraordinary sex ratios,' *Science, 156*(3774), 477-488.

Hammerstein, P., and Selten, R. (1994) 'Game theory and evolutionary biology,' In *Handbook of game theory* (Vol. 2), Elsevier Science B.V.

Hempel, C. (1965) *Aspects of scientific explanation*, New York: Free Press.

Hitchcock, C. and Woodward, J. (2003) 'Explanatory generalizations, part II: Plumbing explanatory depth,' *Noûs, 37*, 181-199.

Houston, A., and McNamara, J. M. (1999) *Models of adaptive behavior*, Cambridge: Cambridge University Press.

Huneman, P. (2010) 'Topological explanations and robustness in biological sciences,' *Synthese*, *177*, 213-245.

Jackson, F., and Pettit, P. (1992) 'In defense of explanatory ecumenism,' *Economics and Philosophy, 8*(1), 1-21.

Kaplan, D. M. (2011) 'Explanation and description in computational neuroscience,' *Synthese*, *183,* 339-373.

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*, 601-627.

Kitcher, P. (1981) 'Explanatory unification,' *Philosophy of Science*, *48*(4), 507-531.

Kitcher, P. (1984) '1983 and all that: A tale of two sciences,' *Philosophical Review, 93*, 335-373.
Lewis, D. (1986) 'Causal explanation,' In *Philosophical papers* (Vol. II). Oxford: Oxford University Press.
Lewontin, R. C. (1961) 'Evolution and the theory of games,' *Journal of theoretical biology, 1*(3), 382-403.
Machamer, P. K., Darden, L., and Craver, C. (2000) 'Thinking about mechanisms,' *Philosophy of Science, 67*(1), 1-25.
Matthen, M., and Ariew, A. (2002) 'Two ways of thinking about fitness and natural selection,' *Journal of Philosophy, 99*, 55-83.
Matthen, M., and Ariew, A. (2009) 'Selection and causation,' *Philosophy of Science, 76*, 201-224.
Maynard Smith, J. (1978) 'Optimization theory in evolution,' *Annual Review of Ecology and Systematics, 9*(1), 31–56.
Maynard Smith, J. (1982) *Evolution and the theory of games,* Cambridge: Cambridge University Press.
Maynard Smith, J., and Price, G. A. (1976) 'The logic of animal conflict,' *Nature, 246*, 15-18.
Mitchell, S. D. (2008) 'Exporting causal knowledge in evolutionary and developmental biology,' *Philosophy of Science, 75, 697-706.*
Mitchell, S. D., (2012) 'Emergence: logical, functional and dynamical,' *Synthese*, *185*, 171-186.
Odenbaugh, J. (2005) 'Idealized, inaccurate but successful: a pragmatic approach to evaluating models in theoretical ecology,' *Biology and Philosophy, 20*, 231-255.
Parker, G. A. (1978) 'Searching for mates,' In *Behavioural ecology: an evolutionary approach,* Krebs, JR and Davies, NB (eds.), Blackwell Scientific Publications: Oxford, 214-244.
Pindyck, R. and Rubinfeld, D. (2009) *Microeconomics* (7th edition), New Jersey: Prentice Hall.
Pincock, C. (2007a) 'Mathematical idealization,' *Philosophy of Science, 74*(5), 957-967.
Pincock, C. (2007b) 'A role for mathematics in the physical sciences,' *Noûs*, *41*, 253-275.
Pincock, C. (2011) 'Modeling reality,' *Synthese, 180*, 19-32.
Pincock, C. (2012) *Mathematics and scientific representation*, Oxford: Oxford University Press.
Potochnik, A. (2007) 'Optimality modeling and explanatory generality,' *Philosophy of Science, 74*(5), 680-691.
Rice, C. (2012) 'Optimality explanations: A plea for an alternative approach,' *Biology and Philosophy, 27*(5), 685-703.
Salmon, W. C. (1984) *Scientific explanation and the causal structure of the world*, Princeton University Press: Princeton, NJ.
Seger, J., and Stubblefield, J. W. (1996) 'Optimization and adaptation,' In *Adaptation*, M. Rose and G. V. Lauder (eds.), Cambridge: Cambridge University Press, 93-123.
Sober, E. (1983) 'Equilibrium explanation,' *Philosophical Studies, 43*(2), 201-210.
Sober, E. (1999) 'The multiple realizability argument against reductionism,' *Philosophy of Science, 66*(4), 542.
Sober, E. (2000) *The philosophy of biology* (Second ed.), Boulder: Westview.
Stephens, D. W., and Krebs, J. R. (1986) *Foraging theory*, Princeton: Princeton University Press.
Strevens, M. (2004) 'The causal and unification approaches to explanation unified-causally,' *Noûs, 38*(1), 154-176.
Strevens, M. (2009) *Depth: An account of scientific explanation*, Cambridge: Harvard University Press.

Walsh, D. M. (2007) 'The pomp of superfluous causes: The interpretation of evolutionary theory,' *Philosophy of Science, 74*(3), 281-303.
Walsh, D. M. (2010) 'Not a sure thing: Fitness, probability, and causation,' *Philosophy of Science, 77*(2), 147-171.
Walsh, D. M., Lewens, T., and Ariew, A. (2002) 'Trials of life: Natural selection and random drift,' *Philosophy of Science, 72*, 311-333.
Weisberg, M. (2007) 'Three kinds of idealization,' *Journal of Philosophy, 104*(12), 639-659.
Weisberg, M. (2012) *Simulation and similarity: Using models to understand the world,* New York: Oxford University Press.
Weslake, B. (2010) 'Explanatory depth,' *Philosophy of Science, 77*, 273-294.
Williams, G. C. (1966) 'Natural selection, the cost of reproduction and a refinement of Lack's principle,' *American Naturalist*, *100*, 687-690.
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*, 287-318.

[1] Optimality modeling also includes game-theoretic modeling. Game-theoretic models are often referred to as "frequency dependent" optimality models since the payoffs to different strategies will depend on the frequency of strategies in the population.
[2] Other salient examples include Craver (2006) and Machamer, Darden and Craver (2000). In some places, Woodward (2003) allows that not all explanations must be causal. I will discuss the ability of Woodward's account to accommodate noncausal explanation in more detail in the final section of this paper.
[3] Precisely which causes must be included in an explanation is determined either by some account of causal difference making (Strevens 2009), or by the contextual interests of the investigator (Potochnik 2007).
[4] Of course, every model will be accurate with respect to certain features of its target system(s) and inaccurate with respect to others. However, on Strevens's account, the partial representation of an idealized model may still be an explanation if the model veridically represents core causal factors that make a difference (Strevens 2009, Ch. 8). Weisberg (2007, 2012) also includes this kind of veridical representation requirement in his account of minimalist idealization.
[5] Another salient example is Kaplan and Craver's claim that in order for a model to explain "will involve describing the underlying component parts, their relevant properties and activities, and how they are organized together causally, spatially, temporally and hierarchically" (Kaplan and Craver 2011, 605).
[6] Recently, causal approaches have taken more seriously the challenge of certain kinds of idealizations that eliminate irrelevant causal factors (Strevens 2004, 2009). These responses, however, fail to account for cases in which causes that *are* relevant to the explanandum are excluded or drastically distorted. Indeed, I will argue that in some cases a model that provides an explanation fails to accurately represent *any* causes.
[7] Italics in original.
[8] In biological contexts, fitness (or inclusive fitness) is the ideal currency, but often a more easily measured currency is used—e.g., average energy intake.
[9] These are often referred to as the strategies' "payoffs".
[10] Of course, many optimality models are not used to provide an explanation. However, these alternative uses, though important, are outside the focus of this paper.
[11] For instance, if most of the population is foraging in one area, it might be better for an individual to forage elsewhere where there will be less competition.

[12] There are, of course, several different ways that equilibrium and evolutionary stability get defined. However, I will not be concerned with those differences here since my arguments rely only on the general fact that these models explain by showing that a particular strategy (or set of strategies) is the system's equilibrium point.
[13] Importantly, these requirements preserve a version of the accuracy requirement.
[14] These examples are chosen with this goal in mind rather than for their status as well confirmed models.
[15] I emphasize this here because these models are often misleadingly presented, or at least motivated, in terms of decisions faced by individual organisms. However, within the model, individuals are only represented in the aggregate and the payoffs to a strategy are the *average* payoffs to individuals with that phenotypic trait.
[16] This kind of analysis was initially used in economics to determine the optimal strategy for investing limited resources.
[17] That is, additional copulation time brings smaller and smaller increases in the average number of eggs fertilized
[18] Indeed, Sober (2000) cites Parker's model as meeting a more rigorous standard for testing optimality models.
[19] Although I am unable to discuss these cases here, optimality models are also widely used in economics to explain the behavior of firms, markets, and other economic systems. In fact, economics relies so heavily on optimization models that some economists have claimed that microeconomics just is the study of optimization under constraints (or scarcity). For example, Pindyck and Rubinfeld begin their microeconomics textbook by claiming: "Microeconomics describes the tradeoffs that consumers, workers, and firms face, and shows how these tradeoffs are best made" (Pindyck and Rubinfeld 2009, 4). Optimality models are essential for this kind of analysis. There are, however, some important differences between economic optimality explanations and biological optimality explanations (thanks to an anonymous reviewer for helping me notice some more of them). As a result, a detailed analysis of economic optimality explanations will have to be provided in another paper.
[20] In addition, the male-female tradeoff remains essential even when various underlying assumptions of Fisher's model are changed (Hamilton 1967).
[21] Later on, Strevens makes it explicit that the key to explaining stable ecological patterns is to fill in the causal mechanisms (Strevens 2009, 168-75). According to Strevens, an ecological model that represents features that are multiply realizable without specifying the causal mechanisms of the system merely "functions as an explanatory template, to be filled out in different ways to obtain deep explanations of stability in different ecosystems" (Strevens 2009, 170).
[22] One problem is that measuring this kind of "p-generality" (Matthewson and Weisberg 2009) is notoriously difficult, but a comparison of p-generality is possible in many cases.
[23] Jackson and Pettit claim that this kind of modal information is given by a macrocausal explanation and tie the explanatoriness of the information to Lewis' (1986) causal account. I don't want to endorse this interpretation of the macro-level explanations, but I do agree with Jackson and Pettit about what the explanatorily relevant modal information is in this case.
[24] In this section I will usually refer to causal *mechanisms*, but the arguments apply equally well to considerations of causal processes, or causal relationships more generally. In other words, these arguments are not restricted to mechanistic accounts of causal explanation.
[25] These distortions are tolerable because the model is only intended to capture the basic constraints and tradeoffs between these variables *at the level of the population*. That is, they are intended to capture constraints, tradeoffs and fitness differences that range over *aggregates*—they are not meant to accurately describe the causal processes acting on individuals within in the population.
[26] Indeed, one difficultly in constructing biological optimality models is that there is often impossible to tell what phenotypic strategies were available in the past history of the population. This is often taken to be a serious obstacle to optimality modeling. However, perhaps by understanding how this idealization contributes to the explanation of the model, this objection can be mitigated somewhat.
[27] An interesting thing to note about the assumption of infinite population size is that drift is intimately related to natural selection. Indeed, the original interpretation of drift is as the statistical error term (Walsh, 2010) and many have defended a statistical interpretation of fitness that entails that selection and drift cannot be pulled apart as two separate causal processes (Matthen and Ariew 2002, 2009; Walsh et al. 2002; Walsh 2007, 2010). I will not commit myself one way or the other on this issue, but it is worth noting that if selection and drift are inseparable in the way suggested by the statistical view, then one cannot assume infinite population size without consequently *distorting the representation of the entire evolutionary process*. Importantly, recognizing this complication does not require endorsing the statistical view of selection and drift. The key insight is that the evolutionary processes that unfold within biological systems are often extremely complex and causally interdependent. Therefore, idealizations that are

introduced to eliminate specific parts of a process—e.g. assuming infinite population size in order to eliminate drift—will likely result in a distortion of the model's representation of other parts of the process as well.

[28] As before, this analysis is not intended to suggest that an increase in generality always makes optimality models objectively better than less general models (contrary to the kind of view defended in Potochnik 2007, 2010). Sometimes in science we want generality, other times we desire more detail. However, in many cases, our interests will dictate that the explanandum is best explained by a more general optimality explanation. In these cases, the pervasive use of idealizations that eliminate the causal details of real-world systems is part of what allows biological optimality explanations to provide the kind of explanation we seek.

[29] Italics added.

[30] Thanks to an anonymous reviewer for helping me clarify this point.

[31] One reason I choose this option rather than canvassing the details of the various extant causal accounts of explanation is because I largely agree with Cartwright (2004) that none of our current accounts are able to provide a universal analysis of causes. A more salient reason is that I do not have a particular account of causation (or causal explanation) to offer and so I hope not to bias my analysis towards considerations of one account or another.

[32] In addition, Mitchell points out that, "Many biological systems display features of dynamical complexity including bifurcation, amplification, and a type of phase change" (Mitchell 2012, 181).

[33] As a specific example, Jay Odenbaugh notes that in ecology, "[I]t is extremely difficult to manipulate ecological systems in systematic and controlled ways. There are multifarious factors at work and only some of them are recognized at any given time" (Odenbaugh 2005, 233).

[34] In addition, given that optimality models provide very limited accurate information about the causal dynamics that led to the explanandum, they will usually provide almost no information about how the system would behave under an intervention of this kind. Indeed, in order to establish this kind of modularity (or interventionist) claim requires a great deal of additional causal information about the system—information that is usually extremely difficult (if not impossible) to acquire for evolving biological systems.

[35] For some additional examples see (Pincock 2012). In Chapter 7, Pincock discusses how idealized mathematical models can lead us to mischaracterize the interconnectedness of different aspects of the real-world system.

[36] This result should be a liberating for modelers in biology and economics. Given the causal complexity of the target system(s) it is often extremely difficult to identify the causes that would enable us to effectively intervene on the system. Moreover, given the number of idealizations involved in biological and economic models it is rarely the case that we have an accurate representation of the causal relationships. Still, these highly idealized equilibrium models can *explain* a wide range of phenomena without having to accurately represent causal relationships.

[37] For example, Stephens and Krebs claim that: "Foraging theorists have tried to find general design principles that apply regardless of the mechanisms used to implement them. For example, the elementary principles of a device for getting traffic across a river—that is, a bridge—apply regardless of whether the bridge in question is built of rope, wood, concrete, or steel" (Stephens and Krebs 1986, 10).

[38] Philippe Huneman has recently used similar language to describe 'topological explanations', that "abstrac[t] away from causal relations and interactions in a system, in order to pick up some sort of 'topological' properties of that system and draw from those properties mathematical consequences that explain the features of the system" (Huneman 2010, 214).

[39] This is not to suggest that Batterman fails to answer this question. However, since this is not his main focus in analyzing asymptotic reasoning, filling in some additional details provides a fruitful expansion of his account.

[40] Woodward makes this restriction explicit: "[T]he theory I will be developing is restricted to *causal* explanations. To the extent that there are forms of explanation that are noncausal, …they will be outside the scope of my discussion" (Woodward 2003, 187).

[41] Another way to put the same point is that if we continue to expand the notion of *causal* explanation we arrive at a point where telling us that the explanation is causal is not informative.

[42] Indeed, as Bokulich (2011, 2012) suggests, I think we can distinguish different kinds of explanation by the kinds of counterfactual dependence involved. However, Bokulich's account requires 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, 39). In contrast, on my account, models can provide the relevant counterfactual *information* without meeting this further condition of isomorphism. One reason for this is that many models that are thought to provide explanations will fail to be isomorphic with the dependencies within their target system(s) (Weisberg 2012). Consequently, my view allows the dependence relations of an explanation to hold between variables or entities that are only "theoretical constructs", or can only be discerned within the fictional world of the highly idealized model—e.g. in an infinitely large population mating randomly. Bokulich (2011, 2012)

does allow "fictional" models to explain, but it is unclear how this can be made to work with her requirement that the structural dependencies be isomorphic to those in the real world. In short, I maintain that models can provide counterfactual information about what is relevant and irrelevant without having to meet any additional "isomorphic mapping" or "veridical representation" requirements between the dependencies in the model and the counterfactual structure of the phenomenon. In addition, Bokulich's "structural model explanations" require that the "dependence is a consequence of the structural features of the theory (or theories) employed in the model" (Bokulich 2011, 40). In contrast, in the kind of optimality model explanations that are my focus here, the structural features in the model are typically derived from observation, experimentation, or the imagination of the scientist; not from the theories employed in the model (e.g. the theory of natural selection). So another important difference between Bokulich's view and my own is that I think an explanation can depend on structural relations even when those relations are not structural features of the theory employed in a model.

[43] Woodward's later work echoes this requirement: "Good explanations should both *include* information about all the factors which are such that changes in them are associated with some change in the explanandum…and *not include* factors such that no changes in them are associated with changes in the *explanandum*" (Woodward 2010, 291).