**Title:**

*Leveraging Distortions: Explanation, Idealization, and Universality in Science* by Collin Rice: Reply by the Author

**Citation:** Rice, C. (2022). *Leveraging Distortions: Explanation, Idealization, and Universality in Science* by Collin Rice: Reply by the Author. *Studies in History and Philosophy of Science 95:* 233-235.

**Please Cite Published Version!**

Collin Rice

Colorado State University

Department of Philosophy

corice@colostate.edu

I’m extremely grateful to each of the contributors who took the time to respond to my book. Their comments have raised several useful critiques; given the limited space, I will do my best to say something in response to what I take to be their main points and criticisms.

**1. Reply to Odenbaugh**

Jay Odenbaugh suggests that my core examples of noncausal explanation can each be interpreted in such a way that there is some causal information provided by the explanation. Odenbaugh’s comments seem to suggest that finding some minimal causal information that can be gleaned from the explanation gives us reason to doubt that it is a genuine case of noncausal explanation. I disagree. Just because an explanation provides some causal information does not mean the explanation is a causal one. The onus is not on the defender of noncausal explanations to show that absolutely no causal information can be provided by the model, any more than a defender of causal explanation must show that absolutely no noncausal information plays a role in a causal explanation. As an alternative, I suggest that when noncausal features are the primary features cited as those on which the explanandum (counterfactually) depends, we have a noncausal explanation.

 In the optimality case, Odenbaugh suggests that the explanation is causal because the model shows us that “The singular causal sequences all causally produce the same explanandum” (Odenbaugh, p. 2). However, the reason this information plays a role in the equilibrium explanation is not to tell us about the importance of those causes, but is instead to illustrate the general *irrelevance* of the causes that led to the explanandum. Moreover, the features that *are* primarily responsible for the explanandum—i.e., the key features directly cited in the explanation—are noncausal structural constraints and tradeoffs.

 Odenbaugh correctly notes that, if my interpretation of statistical explanations required the statisticalist account of natural selection, then it would require begging the question against the causalist. But my account does not require this. Instead, my focus is on which features of the population *are referenced by the statistical models appealed to in these explanations*. Even if selection is a causal process, not all parameters and variables within biological models are intended to represent those causal processes.

 Similarly, while the LGA model does reference particles that move, collide, and jump, those minimal causal features play almost no role in the minimal model explanation of universality. Indeed, the whole point of the minimal model explanation is to show that the particular causal entities and interactions described by the LGA model are completely irrelevant for displaying the universal patterns of interest because radically different entities and interactions would have resulted in the same universal behaviors.

 In addition, I don’t think the deidealization processes Odenbaugh describes challenge the idea that the Hardy-Weinberg model is a holistic distortion for two reasons. First, as Odenbaugh notes, removing the assumptions of an infinitely large population or introducing parameters for selection violates the HW-equilibrium. But it is this equilibrium equation that is central to biologists’ explanations and understanding of various features of heredity (Morrison 2015). I don’t deny that we can introduce natural selection into the model *for other explanatory purposes*, but deidealizing to accomplish different explanatory aims is different from deidealizing and generating the same mathematical results. Second, I contend that introducing selection into this model is not really an instance of deidealization since the representation of the system’s selection processes will be, in most cases, highly idealized as well. Given that the model still pervasively distorts both the selective and non-selective features of the biological population, I see little reason to think the model will better approximate the truth or that we will be able to easily decompose it into its accurate and inaccurate parts.

**2. Reply to Jhun**

Jennifer Jhun proposes that the views defended in *Leveraging Distortions* might be applied to additional cases from economics. I agree. First, it seems quite plausible to me that the explanation of stable patterns via the construction of idealized agent-based models in econophysics will have much in common with the kinds of minimal model explanations I discuss in the book. Indeed, Jhun et al. (2018) have, correctly I think, argued that economists’ use of the JLS model has many of the features of what Robert Batterman and I call ‘minimal model explanations’: they involve the construction of minimal models and make use of infinite idealizations to apply renormalization. As a result, I think the justificatory story that involves showing that the idealized model is within the same universality class as the real system(s) will apply here in much the same way. Similarly, the dynamic stochastic general equilibrium (DSGE) model Jhun discusses is very similar to some biological cases. In the economics case, highly idealized ‘representative agents’—e.g. one big consumer—stand in for the aggregate of the actual individuals in the economy. This is similar to the use of ‘one giant leaf’ models in ecology that treat the whole canopy of a forest as a single giant leaf. Both models aim to capture key macroscale parameters of the system that can be realized (or disaggregated) in a variety of ways at the microscale. When the same macroscale parameters apply across systems that are very heterogeneous at their microscales, that is one way to give rise to a universal pattern. Moreover, in both cases, the models use idealized representations of microscale features (individual consumers or leaves) to stand in for macroscale features of the system that are stable across a number of changes to the microscale features. When this kind of modeling is successful, we can see why the real system(s) and the model system display similar universal macrobehaviors despite their having very different microscale features and despite the model drastically distorting the macroscale features that play important roles in how that pattern occurs in the real cases.

Jhun also raises some questions about how to differentiate universality classes. I agree that showing how the universality account applies to more straightforward cases of causal explanation would be useful. So I will try to do that more directly here. Jhun suggests two models:

M1. Lifetime Income = A\*(Years of Education)

M2. Lifetime Income = B\*(Years of Education) + C\*(Combined Years of Education of Parents)

She asks whether M1 and M2 are members of different universality classes or the same universality class. My answer is that this is an empirical question that depends on whether or not combined years of education of parents is a difference-making factor for the pattern of interest. For universal patterns where years of education plays an important role, but combined years of education of parents is insignificant, I suggest that these models will likely be in the same universality class (assuming they capture roughly the same patterns of counterfactual dependence between years of education and lifetime income). However, in cases where combined years of education of parents plays a significant (i.e., difference-making) role in whether or not the universal pattern of interest occurs, I suspect that M2 will be in the same universality class as the real cases, but M1 will not. In other words, since each system will display numerous patterns—that will depend on different features and be stable across different changes—whether two models are in the same universality class depends on contextual factors concerning which universal patterns are of interest and on empirical facts about the features of the system on which those patterns (counterfactually) depend. In short, given that these models describe different causal structures, whether they will be in the same universality class depends on whether those differences in causal structure make a difference to the occurrence of the universality pattern(s) that characterize that class.

**3. Reply to Elgin**

Catherine Elgin raises three main lines of critique involving the use of counterfactuals in my account of explanation, the use of universality classes to justify the use of idealized models, and my factive account of understanding.

 I readily admit that cases that appeal to mathematical necessities are some of the most challenging cases for counterfactual accounts of explanation. However, I think they are equally challenging for other difference-making accounts of explanation. For example, Woodward’s interventionist account would require us to evaluate what would have been different if an intervention had changed the mathematically necessary features of the system. Strevens’s difference-making account would require us to determine whether a veridical causal model without the mathematical necessity would still entail the explanandum. In both cases, we need to evaluate what would have been the case in the impossible situation in which the mathematical necessity fails to hold. This is difficult, but I think my counterfactual account has a slight advantage here because it does not add the requirement that we interpret such a change to a mathematically necessary feature of the system as a causal manipulation or intervention in order for it to be explanatory. Additionally, while my goal in discussing the cicada case was to directly address the challenges mathematical explanations raise for counterfactual accounts, I’m certainly open to there being other counterfactuals that might be used to explain in this case; e.g. the ones suggested by Elgin. Still, there are numerous other instances of scientific explanations that appeal to mathematically necessary facts that I think any counterfactual account will need to address.[[1]](#footnote-1)

 Next, what determines which universality classes are worthy of attention is the phenomenon of interest (i.e. the explanandum) and other features of the context of inquiry. As Elgin notes, much of the asymmetry of explanation can be captured by appealing to what the audience wants explained and how they would like it explained. But, more generally, which universal patterns are of interest—and which instances of universality we would like to explain the stability of—depends on the interests of scientists. Of course, this process will often involve picking out a set of difference-making features of interest as well. I certainly think of the counterfactual dependence requirement(s) involved in my account as specifying a noncausal conception of difference-making. However, I still think that claiming that the models used to explain in science are justified because “its divergences from and distortions of the target are not difference-makers” (Elgin, p. 5) is insufficient because the models *also drastically distort the difference-making features of their target systems.*

Similarly, while exemplification helps us accommodate some cases of using multiple conflicting models, it still struggles to handle cases of multiple conflicting models that each distort *the same* difference-makers for *the same* explanandum. If different features are relevant to the occurrence of different patterns, then we can maintain that different models are focused on different patterns and exemplify different features of the system. But this is importantly different from cases where we want to explain the *same* pattern via multiple models that provide conflicting representations of the *same* relevant features (Morrison 2011, p. 347). Here we can’t interpret the models as focused on different patterns, nor can we say that they each accurately represent, instantiate, or exemplify the features relevant to the occurrence of the pattern of interest. This is what Morrison (2011) calls a case of ‘genuinely inconsistent models’ of which, she argues, the multiple models of the nucleus is an example. There are several other cases throughout the sciences as well (Parker 2006, p. 350). I suggest that these cases of genuinely inconsistent models are quite different from exemplifying or representing different aspects of the system for different purposes or explanatory goals.

 Finally, I agree with Elgin that we need to have some justification for the claims involved in scientific understanding. However, I contend that there are ways to provide that justification that do not require scientific models to provide an accurate representation of their target systems. Indeed, much of the book focuses on showing that many scientific models are in the same universality classes as their target systems because this provides precisely the kind of alternative justification story about why the information extracted from scientific models ought to be believed (or accepted). Moreover, given my aim of allowing for groups to scientifically understand as well as individual scientists, I welcome the suggestion that ‘acceptance’ might be better suited for my purposes than ‘belief’. Nonetheless, I think scientists can (and do) justifiably accept the claims they extract from idealized models without accepting the models themselves. Thus, I resist the idea that moving to acceptance requires us to conclude that scientific understanding is embodied in the idealized models used in science.

**4. Reply to Pincock**

Christopher Pincock uses an example from economics to raise questions about two aspects of my account of explanation: (1) the requirement that the features on which the explanandum counterfactually depends ‘account for’ its occurrence and (2) the need to show that contextually salient irrelevant features are irrelevant. The case in question involves the following explanandum: “the monopoly will always be, in comparison with the price-taking firm, a price raiser and output restrictor” (Woodward 2003, p. 190). Pincock argues that the explanation here is given by noting that “As the marginal revenue curve for the monopolistic firm is strictly below the marginal revenue curve for the competitive firm, as long as the marginal cost curve is upward slowing, the monopolistic firm will maximize its profits by decreasing production.” (Pincock, p. 4) This entails that the price for a monopolistic firm will be higher than for a competitive firm.

 To Pincock’s first critique, I’d note that the explanation offered in this case *does* seem to predict that the price will depend on the number of firms in the market and tells us why that is the case. I suspect that the counterfactuals involved in the explanation of this causal generalization are things like: ‘If the marginal revenue curve for the monopolistic firm were strictly above the marginal revenue curve for the competitive firm, then the relationship between price and number of firms would be different’. Another counterfactual identified by Pincock appears to be, “If the monopoly firm did not act in a way to increase its profits, then it would not decrease production and so prices would not go up” (Pincock p. 4). More generally, I suspect there is much more counterfactual dependence information involved in this case than Pincock’s presentation suggests. My claim is that, when taken together, those features ought to make it the case that we could predict that, *when those features are present* (and perhaps that some ceteris paribus conditions are met), the occurrence of the explanandum can be accounted for—even if its probability of occurring in that situation is not very high. Of course, the explanation of this kind of conditional causal statement does not tell us when or where the antecedent of that conditional will be satisfied, but I think the explanatory counterfactuals identified by my account would still give us a clear sense of why, in cases where those conditions are met, the dependency holds.

 Concerning Pincock’s questions about explanations of universality, my general response is that this is clearly *not* an explanation of universality. This is because the explanation does not aim to tell us *why* the pattern is stable across a particular range of cases—only that it is. While I do think that explanations of universality need to say why a universal pattern is stable, this is not a requirement of my general account of explanation (nor do I think all patterns need to be explained by accounting for their stability). This is the crucial difference between showing *why* and showing *that.* My general account of explanation only requires the latter for irrelevant features that are contextually salient. Moreover, this information about counterfactual irrelevance can be provided in multiple ways and can play a crucial role in many other types of explanations. I contend that this is just what is happening in Pincock’s expansion of the case to make contextually salient the irrelevance of varying consumer income. If we want to explain why the causal generalization continues to hold despite differences in consumer income, then the explanation ought to show that consumer income is irrelevant. That isn’t the same as requiring the explanation to further explain *why* that feature is irrelevant. Showing why various heterogenous features are irrelevant is essential to explanations of universality, but not all explanations that cite the irrelevance of features to the explanandum are explanations of universality.

**References**

Jhun, J., Palacios, J. and Weatherall, J. O. 2018. Market crashes as critical phenomena? Explanation, idealization and universality in econophysics. *Synthese* *195*: 4477-4505.

Morrison, M. 2011. One phenomenon, many models: Inconsistency and complementarity. Studies in History and Philosophy of Science Part A 42(2): 342–351.

Parker, W. S. 2006. Understanding pluralism in climate modeling. Foundations of Science 11: 349–368.

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

1. I will also note that, given that rather drastic departures from the truth often deepen our scientific understanding, I’m less troubled by the distance of the possible worlds involved in claiming that ‘17 is prime’ is a contingent truth than Elgin is. But that discussion will have to be taken up elsewhere. [↑](#footnote-ref-1)