LEVERAGING DISTORTIONS
Collin Rice

Reviewed by William D'Alessandro

Leveraging Distortions: Explanation, Idealization, and Universal Patterns in Science
Collin Rice
Cambridge, MA: MIT Press, 2021, £57.00
ISBN 9780262542616

Philosophy of science is undergoing an apparent sea change with respect to attitudes about explanation, causation, and representation. Collin Rice has been a major contributor to these trends for some years. His new book, *Leveraging Distortions*, carries his project further. It's an excellent piece of philosophy of science that deserves to be widely read and discussed.

The main ambition of *Leveraging Distortions* is to overturn received wisdom about the nature of models and their contributions to scientific explanation and understanding. The 'standard view' that Rice opposes is a package of claims including the following:

A causal theory of explanation: In order to genuinely explain some phenomenon, it's necessary (and perhaps sufficient) to identify its relevant or difference-making causes.

A 'decomposition thesis' about scientific models: The accurate parts of a model can be distinguished from its idealized or distorting parts in a principled way. Only the parts that accurately represent real, difference-making features contribute to a model's explanatory power.

A realist view of models and theories: Science contributes to our understanding of nature insofar as it veridically represents the world (and successful models and theories typically do so).

The standard view is broadly aligned with the ontic conception of Wesley Salmon and his successors. This tradition sees the identification of objective causal relations as central to scientific explanation. Its allies today include prominent philosophers of the 'mechanist' and 'manipulationist' schools (for example, Carl Craver, James Woodward, and Michael Strevens).

Rice rejects all three tenets of the standard view. In place of the causal theory of explanation, Rice endorses a counterfactual theory, in part because he thinks we can explain by appealing to mathematical results, optimality considerations, and other non-causal facts. (His version of the counterfactual approach is somewhat novel, in that it requires information about counterfactual irrelevance in addition to the usual relevance criterion.) Against the decomposition thesis, Rice argues that many models are 'holistically distorted', such that the accurate and inaccurate parts can't be neatly disentangled. And against realism, Rice suggests that non-veridical representations are an important source of scientific understanding.

Over the last decade or so, many philosophers of science have grown comfortable with non-causal explanation (Lange [2016]; Reutlinger and Saatsi [2018]). The counterfactual approach is also a perennial contender. So neither idea is revolutionary by itself (although it should be noted that Rice's earlier work on non-causal explanation helped contribute to its current mainstream status). What's most distinctive and important about *Leveraging Distortions* isn't these claims, but its insights about unrealistic models and the kinds of explanations they provide.

On Rice's view, many scientific models misrepresent the world in profound and pervasive ways, such that one can't meaningfully distinguish their accurate and inaccurate parts. Nevertheless, such models are often explanatory—not only in spite of their distortions, but often precisely because of them. In many cases this happens because a model system belongs to the same universality class as the real system under study. (A universality class is, roughly, a set of systems that exhibit reliably similar behaviours in spite of their different physical properties and causal dynamics.) Knowing a system's universality class can provide a wealth of modal information—positive information about which factors the phenomenon depends on, but also negative information about what isn't required for the phenomenon to occur. This kind of understanding would escape us if we focused only on the system's actual difference-making causes.

As Rice shows, universality plays a key role in many modelling practices, from 'minimal models' to those that invoke statistical results like the central limit theorem, to the jointly inconsistent sets of models often used by multi-scale modellers. Rice's treatment of these issues forms the heart of the book (Chapters 1–7). The last two chapters deal with applications and further questions, for example, about the nature of understanding, scientific progress, and the realism debate.

Leveraging Distortions is filled with interesting case studies. Describing one may help convey what Rice is up to. In Chapter 6, Rice discusses climate scientists' treatment of the formation of melt ponds on sea ice sheets. (This process is climatologically important because sea ice mostly reflects sunlight, while meltwater mostly absorbs it.) Studying these ponds via a realistic representation of their causal dynamics is a no-go, since the melting unfolds in a highly complex and hard-to-predict pattern. But it turns out that melt ponds belong to a universality class that's relatively well understood—a class of systems exhibiting rapid phase transitions between configurations with fractal dimensions of roughly 1 and 2:

> As a result of discovering that these melt ponds are in this universality class, [climate] modelers are able to justifiably use various idealized models within that universality class to extract information about the macroscale behavior of these real-world systems that are independent of their mechanisms, components, and causes. (p. 168)

What's more, according to Rice, the information gained from these unrealistic models allows us explain and understand melt pond development. For example, it tells us why ponds with an area of approximately 100 m^2 are unstable (namely, because this is the length scale associated with a fractal dimension of approximately 1.5, and systems in the universality class tend to evolve away from this transitional regime).

Other authors in the anti-standard view tradition have arrived at conclusions somewhat similar to Rice's. Alisa Bokulich may be the most prominent example: across a body of work (including Bokulich [2008], [2011], [2018]), she's argued forcefully against the ontic conception of explanation and defended the explanatory power of unrealistic models. (Others with allied views include,

for example, Robert Batterman, Catherine Elgin, Margaret Morrison, and Angela Potochnik.) It's worth highlighting some of the differences between Rice and others in this reformist cohort.

To start with, unlike Elgin and Potochnik, Rice accepts a broadly factive view of explanation and understanding, according to which an explanation is only acceptable if it provides true information about the explanandum. (Of course, for Rice, this need not be information about the explanandum's actual causes; in many cases, it will be facts about counterfactual relevance and irrelevance instead.)

Rice's disagreements with Bokulich are more subtle. Both are factivists, both think unrealistic models can explain without accurately representing causes, and both think counterfactuals are an important part of this story. Where the two views diverge most clearly is in their claims about how we gain counterfactual knowledge from models. For Bokulich, the relevant modal relationships must be represented within the model itself. For Rice, by contrast, our counterfactual knowledge often comes from the way we interpret and use models, not from the explicit contents of those models.

To see what this difference amounts to, consider the lattice gas automaton (LGA) model of fluid flow. This is a particularly simple 'minimal model' that represents fluids by point particles, arranged gridwise and subject to a few basic dynamical constraints (locality, conservation, and symmetry). Despite its lack of realistic causal information, scientists often appeal to the LGA model to explain flow patterns. This turns out to be another universality-based explanation: fluids with heterogeneous physical properties and causal dynamics exhibit similar behaviour because these fluids belong to the same universality class. Universality therefore lets us explain not just why some particular fluid flows as it does, but why all fluids in the class behave similarly.

As Rice emphasizes, however, this sort of insight can't simply be read off the model itself:

> Merely looking at the features represented by the LGA model fails to answer [such] explanatory questions. In order to answer them, a minimal model explanation is only provided by combining the model with various mathematical techniques that demonstrate the irrelevance of most of the features of real, possible and model fluids. (p. 77)

Contra Bokulich, then, not all model-based explanations work by directly representing patterns of dependence. We often need additional knowledge about which systems the model applies to and why.

A final feature of *Leveraging Distortions* worth highlighting is its account of scientific understanding. Pretty much everyone agrees that explanation and understanding are closely related. According to many authors (de Regt, Khalifa, Strevens, Trout), the relationship is one of dependence: possessing understanding requires (and perhaps just is) grasping an appropriate explanation.

Rice disagrees. On his view, it's possible to have objectual understanding without explanation. This state is often achieved with the help of unrealistic models, and Rice describes three scenarios. First, models can confer understanding by telling us what's necessary or unnecessary for a phenomenon to occur, without also accounting for the occurrence of the phenomenon in a way that rises to the level of genuine explanation. Second, we can obtain understanding from models of hypothetical scenarios, which show what kinds of behaviour may be produced by a given set of features (even if those features are never instantiated). Third, models can be used to explore modal space, enabling us to better understand observed phenomena by locating them within a wider system of possibilities.

I find much of this picture appealing and convincing, or at least worth taking quite seriously. But let me mention a few concerns.

First, Rice spends surprisingly little time presenting scientists' own views about the explanatory value of the models he discusses. The key chapters, 3 and 6, for example, are supposed to establish Rice's key claims about non-causal explanations in general and holistically distorted models in particular. Rice presents a variety of cases of both types. These examples are nicely set out, and it's clear from Rice's discussion that scientists find such models interesting and useful in various ways. But there are few places in either chapter where a relevant expert makes anything like a clear statement about explanation. Instead, Rice seems content to

rely on his own judgements and the opinions of other philosophers. This strikes me as a significant dialectical weakness. Friends of the standard view presumably have the opposite intuitions about these cases; without being able to show that scientific practice is on his side, it's unclear what reason Rice has given them to change their minds.

Second, I don't know how intra-mathematical explanation is supposed to fit into Rice's picture. This is an important phenomenon about which Rice is almost totally silent. Does he think there are explanations in pure mathematics, as many philosophers have recently (and I think correctly) argued (D'Alessandro [2019])? If so, is his counterfactual account supposed to somehow cover such cases? A counterfactual theory of mathematical explanation isn't the complete non-starter it might sound like (Baron *et al.* [2020]), but there are serious problems for such accounts to overcome (Kasirzadeh [2021]; Lange [2022]). On the other hand, if Rice's account isn't intended to cover mathematics, it would be nice to know why not. It's implausible that mathematical explanation and understanding are completely different creatures from their scientific counterparts. So we ought to expect the correct story about one set of phenomena to shed at least a little light on the other. I wasn't expecting Rice's book to resolve these issues once and for all, but it's a disappointing omission not to show any awareness of them.

Third, as mentioned above, Rice holds that models can confer understanding without providing explanations. I'm sympathetic to this view, but less sanguine about Rice's way of drawing the line. To illustrate his approach, Rice discusses Schelling's famous toy model of urban segregation. Schelling's model represents a city as a grid of housing units, and assigns to residents only a binary group identification (say, red or blue) and a preference to be surrounded by a certain proportion of neighbours from their own group. Schelling showed that segregation quickly results even when agents have relatively weak preferences for same-group neighbours.

Rice claims that Schelling's model enhances our understanding of segregation, but that it 'fails to provide anywhere close to a complete explanation of how any actual segregated city has arisen […] Merely knowing that it is possible for mild preferences for like neighbors to produce segregation falls short of being able to explain why cities are actually segregated' (p. 236).

So, what else would be needed to explain the segregation of a particular city? Presumably, the answer isn't the actual causal history of segregation in that city; this would be just the sort of causal constraint on explanation that Rice rejects throughout the book. Rice's stance seems to be that the Schelling model doesn't count as explanatory because it doesn't give us enough modal information: 'grasping the truth of [a] single counterfactual is woefully inadequate for providing a complete explanation of why cities are segregated' (p. 244).

I find this unsatisfying in a couple respects. For one, Rice is surely right that Schelling's model doesn't count as a 'complete explanation' of segregation. But does an explanation have to be complete in order to be useful or acceptable? Is there no such thing as an illuminating partial explanation? That's a strange view, but Rice seems to be suggesting something like this here.

In any case, Rice's analysis of the Schelling model seems to sit uncomfortably with his other commitments. Consider the LGA model of fluid flow discussed earlier. Rice considers the LGA model explanatory, which presumably means it provides a 'complete explanation'. But in what sense could this be true? The LGA model certainly doesn't explain every feature of fluid flow. Nor does it make completely accurate predictions even in its intended domain of application. (Indeed, the fluid dynamics textbook (Succi [2001]) has a whole section on 'lattice gas diseases', that is, defects of the LGA model.) I'm left wondering what sort of completeness is possessed by the models Rice considers explanatory, but not by other simple models like Schelling's.

For philosophers more invested in the standard view than this reviewer, it seems likely that much of the wrangling over Rice's claims will focus on another set of issues—for example, precisely what it takes for a model to count as veridical, what it means for models to represent causes, how we should think about identifying the contents of a model, and whether some form of the model decomposition thesis can be salvaged after the smoke clears. Rice anticipates several worries along these lines, and I think he's often successful at fending them off. But some lines of attack remain open, and it will be interesting to see how these debates play out in the coming years.

Wherever one stands along these fault lines, it's an encouraging sign of progress that we're no longer waging war by swinging crude clubs labelled 'ontic' and 'epistemic', but rather discussing tricky issues about the nature of representation through the lens of a broad array of unrealistic models. Rice has played a major role in bringing about this enlightened state, and we're all in his debt.

Acknowledgements

William D'Alessandro
LMU Munich
d.william@lmu.de

References

Baron, S., Colyvan, M. and Ripley, D. [2020]: 'A Counterfactual Approach to Explanation in Mathematics', *Philosophia Mathematica*, **28**, pp. 1–34.

Bokulich, A. [2008]: 'Can Classical Structures Explain Quantum Phenomena?', *British Journal for the Philosophy of Science*, **59**, pp. 217–35.

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

Bokulich, A. [2018]: 'Representing and Explaining: The Eikonic Conception of Scientific Explanation', *Philosophy of Science*, **85**, pp. 793–805.

D'Alessandro, W. [2019]: 'Explanation in Mathematics: Proofs and Practice', *Philosophy Compass*, **14**.

Kasirzadeh, A. [2021]: 'Counter Countermathematical Explanations', *Erkenntnis*, available at .

Lange, M. [2016]: *Because without Cause: Non-causal Explanations in Science and Mathematics*, New York: Oxford University Press.

Lange, M. [2022]: 'Challenges Facing Counterfactual Accounts of Explanation in Mathematics', *Philosophia Mathematica*, **30**, pp. 32–58.

Reutlinger, A. and Saatsi, J. [2018]: *Explanation beyond Causation: Philosophical Perspectives on Non-causal Explanations*, Oxford: Oxford University Press.

Succi, S. [2001]: *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*, Oxford: Clarendon Press.