Counternomics

Peter Tan
(譚溥文)
Naperville, Illinois

Bachelor of Arts in Philosophy and History,
University of Illinois at Urbana-Champaign

A Dissertation Presented to
The Graduate Faculty of the University of Virginia
in Candidacy for the Degree of Doctor of Philosophy

Corcoran Department of Philosophy
University of Virginia
March 2019

For my forefathers
為我的祖先

Acknowledgements

I am deeply grateful to my dissertation director, Trenton Merricks, for his unwavering and tireless assistance with my project, and all the ways he has impacted it both visibly and invisibly. Here is a true story: in the very first conversation I ever had with Trenton, he called an idea of mine "very stupid." It was; and I would continue to show Trenton many very stupid ideas. He has read countless drafts, been patient with my stubbornness even more times, and importantly, has been a real and genuine human being in some times when I needed it very much. It is due to his example that I have learned to write with any degree of competence—another true story is that Trenton appraised the first draft of my qualifying paper for the en-route MA as "terribly written"—and that I have learned to appreciate the intellectual beauty, the complexity, and the surprisingness that can come in philosophical discourse. I can only hope that all those who write a dissertation can enjoy the privilege of having a director as helpful and supportive as him.

I am also deeply grateful to Paul Humphreys. Though he is not officially a part of my dissertation committee, his influence on every part of this dissertation is palpable. It is because of his course on Emergence and Reduction that I had a road-to-Emmaus conversion to viewing the metaphysics of science as deeply valuable and, indeed, essential to metaphysical inquiry in general. He and his work have taught me to be sensitive to the ways in which the contents and boundaries of scientific practice and reasoning provide limits on our inquiry in metaphysics and philosophy of science; and this influence is deeply felt throughout the science-driven methodology I employ throughout this dissertation. Lastly, the fruitfulness I have learned, from him, that the philosophy of science has for metaphysics and epistemology in general continues to shape my research agenda and philosophical interests. This dissertation would not have been possible without his influence.

I thank Elizabeth Barnes and Ross Cameron, who offered invaluable comments and discussion of various parts of my dissertation at its various stages, especially Chapter 1. Both of them have a uniquely deep insight into many of the areas my dissertation deals with both on the surface and underneath: philosophical methodology, the relationship between science and philosophy, and modality. I also thank Chad Wellmon for reading my dissertation and volunteering very last-minute to be my final reader.

I am grateful as well to Dan Korman, whose enthusiasm about philosophy was contagious during my time as an undergraduate, and without whom I would never have acquired an interest in metaphysics!

Thanks are also due to Juha Saatsi—a conversation with Juha during a summer workshop at the Durham Emergence Project was the impetus for my entire dissertation topic. I thank Jim Darcy for literally giving me the shirt off his back for my dissertation proposal defense, and for many helpful discussions. I also thank Adam Tiller, Anjan Chakravartty, Antonia LoLordo, Barbara Vetter, Daniel Nolan, Elyse Oakley, Harjit Bhogal, Jessica Wilson, Liz Jackson, Martin Glazier, Matthias Jenny, Sebastian Ramirez, and audiences at the PSA 2016, 2018, and the 2017 Society for the Metaphysics of Science for valuable feedback.

I am especially thankful for Jonathan Barker and Sam Lundquist, who have been truly invaluable friends in times of deep personal need. Jonathan and I have been throwing metaphysics back-and-forth at each other since the very first week of graduate school; I can only hope that what we have been slinging has increased substantially in its quality.

Lastly, I am most thankful for my mom, Li-su Tan, and my partner, Felicity Martin. I would not have made it this far without either of their support. And I have learned through my work that while philosophy can be deeply rewarding, there is nothing even close to being as rewarding or as worth pursuing as being with the ones you love.

Counternomics

Contents

Introduction

This dissertation is a series of essays on the laws of nature and on the metaphysics and epistemology of modality in science. My aim is to identify the ways in which the notions of *could, would,* and *must* are related to each other, and what their use in scientific reasoning can tell us about the world's fundamental inhabitants and structure. This dissertation focuses on the nature and relation of a particular kind of counterfactual conditional to the other modal notions, and how this sort of counterfactual can illuminate the other modal notions.

I focus on *counternomic* conditionals, which are those counterfactuals whose antecedents run contrary to the laws of nature. As we will see throughout my discussions here, counternomics are prevalent in every area of scientific practice. To identify just a few, counternomics are present in:

- *Scientific explanations*: "If diamond had not been covalently bonded, then it would have been a better electrical conductor," "If intermolecular hydrogen bonding did not occur, then water would be gaseous at room temperatures"
- *Reasoning about the laws of nature*: "If gravitation had been an inverse-cube law, instead of an inverse-square law, it still would have obeyed the conservation of energy," "If there were magnetic monopoles, then Ampère's circuital law still would have held"
- *Model-based reasoning*: "If krypton had been an ideal gas, then it would have obeyed $PV = nRT$ exactly," "If water were continuous and incompressible, then it would behave precisely as the Navier-Stokes equations describe"

- *Reasoning about superseded scientific theories*: "If Bohr's theory of the atom had been true, then an electron's angular momentum L in the ground state would have been observed at $L=\hbar$ (i.e. the Planck constant)," "If classical electromagnetism had been true, an alteration in the wavelengths of light shining on a sample of aluminum would have induced a change in the sample's rate of photoelectron emission."

All of these practices are in some way or another part of the core of our contemporary scientific practice. And one does not have to look very far to find examples of counternomics in these contexts. But despite how common counternomics are in scientific practice, there has been virtually no direct attention paid to them in the extant literature on counterfactuals, or the literature on modality, or even within the philosophy of science. (So far the only direct attention to counternomics has been in two areas. First, necessitarians about the laws of nature face an objection that they cannot account for counternomic truths. I advance my own version of this objection in Chapter 3. Second, Marc Lange has argued for a view of the laws of nature where lawhood is a matter of primitive "counterfactual resilience", and to illustrate that some laws are more counterfactually resilient than others, he uses the sorts of examples I have provided in the *Reasoning about the laws of nature* examples given above. These are the only two sources of extant discussion concerning counternomics.) This lacuna is surprising.

Counternomics raise, rather immediately, a number of intriguing philosophical questions. Their antecedents violate the laws of nature, yet they seem to be non-vacuously true: what does this tell us about the laws of nature? And given that we cannot observe law-violating states of affairs at all, since they obviously cannot obtain, how can we gain evidence for which counternomics are true? And irrespective of our *knowledge* of which counternomics are true, what is it in our ontology that makes counternomics true in the first place, again, given that there cannot be law-violating states of affairs?

This dissertation is the first and only one of its kind: it is the first and only sustained discussion of counternomic conditionals. Again, my aim is to identify what counternomics can tell us about theories of the counterfactual more generally, and also what they can tell us about the relationship between the counterfactual, physical and metaphysical possibility, and physical and metaphysical necessity.

Because this dissertation is a collection of papers, it is a little hard to find a single overarching conclusion I reach that I can easily point to. Nor is there the easiest common thread to find throughout all of these chapters. However, my contributions are still substantial. In what follows, I shall summarize each of these five chapters' main contributions, and whenever possible, identify a common thread that links one to another.

Of my dissertation's five chapters, **Chapter 2**, "Interventions and Counternomic Reasoning," is the most standalone. I examine two experimental results that seem to provide evidence for counternomic hypotheses. I show that our counterfactual epistemology in these cases resists characterization along the interventionist lines one might expect, since these experimental manipulations do not constitute cases of genuine non-confounded physical intervention. I then offer my own characterization of their evidential relationship as a kind of undefeated inductive inference that extrapolates from observed linearity in the scientific domains of interest. This is, to my knowledge, the first and only paper directly on the epistemology of counternomics.

I have said that one of my primary aims in this dissertation is to identify the relationship between the three modal notions. There is a long philosophical tradition, owing to Stalnaker (1968) and Lewis (1973), that analyzes every non-vacuous counterfactual in terms of what is metaphysically possible. 'If *p* were the case, *q* would obtain' is true iff the closest *p*-world are *q*-worlds; and of course, metaphysically possible worlds were in this view taken to be the only worlds there are. In recent years, there has been increasing push-back against this orthodoxy. **Chapter 1** provides a new argument against this orthodoxy. I identify a number of scientific practices in which counterfactuals play an indispensable explanatory or epistemic role, and show that science routinely treats counterpossibles in these contexts as being non-vacuously true or false. For as it turns out, many counternomics—many of those that I have given examples of above—have antecedents which are not just physically impossible, but metaphysically impossible. I argue that these cases constitute a conclusive reason to reject the orthodoxy on counterpossibles. This chapter represents the dissertation's largest contribution so far. If counterpossibles are non-vacuously true, the traditional "analysis" of counterfactuals in terms of metaphysical possibility fails.

My arguments in Chapter 1 inform the arguments in **Chapter 3**, which targets the view known as dispositionalism about modality. Dispositionalists hold that the *de re* modalities and the natural modalities (i.e., counterfactuals and the laws of nature) alike are grounded in the world's dispositional properties. It is rooted in dispositional *essentialism*, which takes there to be essentially dispositional properties, which have their causal profiles as a matter of metaphysical necessity. Dispositionalism implies that the laws of nature obtain with metaphysical necessity, which implies that every counternomic is a counterpossible. However, as I argue in Chapter 1, these counternomics are non-vacuous, and hence the dispositionalist must find some way to provide an ontological account of their truth-value. I argue that the dispositionalist cannot do so, since they cannot in principle countenance dispositions which fail to be law-abiding. On their view, after all, it is the laws that ultimately abide by the dispositions, not the other way around. Because dispositionalism cannot provide us with an ontological account of counternomics' truth, we must reject it as an account of the grounds of *de re* modality and the laws of nature.

Chapter 4 picks up this question, dealt with in Chapter 3, of what the laws of nature must be like, if there are non-vacuous counternomic truths. In this chapter, I focus in particular on one kind of law-relevant counternomic: those that feature and underpin *idealized* laws of nature. This chapter presents a unified argument against three widely-held contemporary analyses of lawhood: Humean reductionism about laws, the dispositionalist view of laws, and the view of laws as relations between universals. I argue that it is a datum about the idealized laws of nature that their lawhood would be "counterfactually preserved" if their idealizations had been true. But these three analyses of the laws all deliver the judgment that the ideal laws' lawhood is *not* counterfactually preserved in the event of their idealizations' truth. I close by considering some alternative views of the laws of nature that do not have this implication, all of which are *primitivist* views of laws in some way or another.

Finally, in **Chapter 5**, I examine *model transfer* in the sciences. Models of particular kinds of systems or phenomena, in particular scientific domains, are frequently applied to do explanatory work regarding the behavior of other kinds of systems or phenomena in distinct domains. Model transfer is a prima facie threat to the received, "ontic" conception of scientific explanation, because the ontology of the transferred models or their original domains is usually explicitly distinct from what the ontology of the transfer domain is taken to be. How can explanation be a matter of identifying ontological structure, if cases of model transfer show that successful explanations can be provided even by models that get their target's ontology wrong? I articulate a new conception of causal processes such that the same causal process can be exemplified by radically different kinds of systems. The presence of counternomics in helping us reason about and justify cases of model transfer supplements the intuitive case that there must be something shared, ontologically, between these disparate systems. I show how this new kind of higher-order, multiply-realizable causal process helps us make sense of explanatory model transfer in the sciences, and also carries surprising implications for theories of the natural laws, debates over scientific realism, and the unity of the sciences. One of the implications of this chapter is that at least some laws are ontologically primitive, which carries over the conclusion of Chapter 4.

To sum up, the contributions of my dissertation reach quite far. I undermine the established orthodoxy on counterfactuals' truth-conditions. I also provide two novel discussions of the laws of nature and their relation to counternomics. I also provide the first and only published discussion of the epistemology of counternomics, drawing from two surprising experimental cases. Lastly, I provide the first discussion of the metaphysical (rather than epistemic) implications of model transfer between sciences. All of this shows that the previous lacuna of discussion regarding counternomics contained the space for much fruitful discussion, and I feel fortunate to have been the one to advance these issues. This dissertation has been a pleasure to write, and my hope is that it will be a pleasure to read.

Chapter 1: Counterpossible Non-Vacuity in Scientific Practice

1 The Target: Counterpossible Vacuity

Part of the received wisdom regarding the counterfactual conditional is that counterfactuals whose antecedents are metaphysically impossible – commonly called "counterpossibles" – are indiscriminately vacuously true. This paper provides a new argument that this received wisdom is false. Drawing on a number of examples from across scientific practice, I argue that science routinely treats counterpossibles as non-vacuously true and treats other counterpossibles as false. In fact, the success of many central scientific endeavors requires that counterpossibles can be non-vacuously true or false. So the philosophical orthodoxy that counterpossibles are all vacuously true is inconsistent with scientific practice. I argue that this provides a conclusive reason to reject the orthodoxy.

My only aim in this paper will be to argue for the simple existential claim that there are non-vacuous counterpossible conditionals. As such, this paper will not deal at any length with the formal semantics or logic of counterfactuals. There are two reasons for this limitation in scope. First, as we shall soon see, the motivations for endorsing the vacuity of counterpossibles extend far beyond the implications of particular semantic models for counterfactuals. The strength and diversity of these motivations are such that one cannot avoid counterpossibles' vacuity by, e.g., simply modifying one's formal semantics. Secondly, my strategy for establishing that there are non-vacuous counterpossibles will be, rather than to identify defects in existing arguments in favor of vacuity, to argue that various usages of counterfactuals in scientific contexts have metaphysically impossible antecedents and must be interpreted non-vacuously. Excursions into the formal semantics would be tangential to a thorough pursuit of this strategy. And indeed, given the considerable pressures to endorse counterpossible vacuity, it would be a surprising conclusion by itself if there turned out to be clear and uncontestable examples of non-vacuous counterpossibles. I shall show exactly that.

The structure of the paper is as follows. Before introducing my examples of counterpossibles in science, it will be necessary to clarify what the orthodox position is supposed to be and exactly what motivates it.[1] It is worthwhile to recognize the diversity of the motivations behind counterpossible vacuity, and once we understand what vacuity amounts to, it will be easier to see what counterpossible *non*-vacuity would look like. In the following sections (§2-4), I argue that non-vacuous counterpossibles are indispensable to scientific explanation, to reasoning with idealized scientific models, and reasoning about superseded scientific theories. I close in §5 with some remarks about where my arguments direct us.

1.1 The Precedent in Favor of Vacuity

So, let us begin by clarifying what it would mean for a counterfactual to be vacuous. And why think that counter*possibles* are? The most familiar precedent in favor of counterpossibles' vacuous truth is set by the standard Lewis/Stalnaker semantics for counterfactuals, which judges a counterfactual, 'If *p* were the case, then *q* would obtain' to be true if and only if all of the closest

* For helpful discussion, I thank Elizabeth Barnes, Harjit Bhogal, Ross Cameron, Jim Darcy, Martin Glazier, Mike Hicks, Paul Humphreys, Liz Jackson, Matthias Jenny, Daniel Nolan, Juha Saatsi, Barbara Vetter, Jessica Wilson, and an audience at the 2017 Society for the Metaphysics of Science. I am especially thankful to Jonathan Barker, Anjan Chakravartty, and Trenton Merricks for their detailed comments on earlier drafts of this article.

[1] There has been some confusion regarding, e.g., what "vacuity" amounts to. See the disagreement between Brogaard and Salerno (2007) and Williamson (2018, forthcoming) on whether counterpossibles' vacuity implies their triviality, and on whether counterpossibles' consequents contribute anything to their meaning or truth-value.

worlds where *p* obtains are also worlds where *q* obtains.[2] Suppose that *p* is metaphysically impossible. Then, unless you are among the minority who countenance "impossible worlds", there will be no worlds where *p* obtains, and *a fortiori* no closest worlds.[3] So the generalization that the closest *p*-worlds are *q*-worlds will look to be vacuously true, and thus, the counterfactual alongside it.

Judging from this familiar semantic result, what characterizes counterpossibles' vacuous truth? It looks like the same kind of uninformativeness paradigmatically exemplified by vacuously true universal generalizations. Let me explain. It seems that 'All wizards are green,' for example, is true-but-uninformative in the following sense: it is true in virtue of nothing more than the fact that there are no wizards, and not because it describes a substantive fact about what the wizards are like or which properties they have. How could it? The wizards are not any way at all. Similarly, under the Lewis/Stalnaker semantics, a counterpossible conditional will come out true in virtue of nothing more than the fact that its antecedent is impossible, and not because it manages to describe a substantive fact about what would result from what in some possible goings-on.[4]

---

[2] Lewis (1973), Stalnaker (1968)

[3] In fact, one well-known reason for countenancing impossible worlds in addition to possible worlds in one's counterfactual semantics is that it aids in rendering counterpossible conditionals non-vacuously true (Nolan 1997). Jessica Wilson (2014) has argued that Humeans should interpret certain backtracking counterfactuals using impossible worlds. If so, Humeans may antecedently have reason to accept counterpossible non-vacuity. My arguments throughout this presuppose neither a Humean nor anti-Humean metaphysics. And I do not have anything to say about impossible worlds in this paper (regarding either their semantics or their metaphysics), but for a survey of that literature, see Nolan (2013).

[4] My characterization of counterpossibles' vacuous truth as uninformativeness resembles Brogaard and Salerno's (2013). They say: "Counterpossibles are…vacuously true and semantically uninformative…they are uninformative in the sense that the consequent of a counterpossible makes no contribution to the truth-value, meaning or our understanding of the whole" (p642).

The semantics of possible worlds supports characterizing counterpossibles in this way. But powerful pressure from considerations independent of any formal semantics leans in favor of this treatment as well. Here is the famous line from David Lewis on the issue, to start with:

> There is at least some intuitive justification for the decision to make a 'would' counterfactual with an impossible antecedent come out vacuously true. Confronted by an antecedent that is not really an entertainable supposition, one might react by saying, with a shrug: If that were so, than anything you like would be true! Further, it seems that a counterfactual in which the antecedent logically implies the consequent ought always to be true; and one sort of impossible antecedent, a self-contradictory one, logically implies any consequent. (1973, p24)

I take it that Lewis's remarks here are meant to float free of any particular semantic model of counterfactuals. And his verdict is clear. When a counterfactual has an impossible antecedent, it is the impossibility of the antecedent that does all of the legwork in explaining why the conditional is true. That is, it is because an impossibility implies anything that a counterpossible comes out true. It is not as if it needs to (or even manages to) describe any substantive facts about what counterfactually depends on or results from what else.

In fact, Lewis seems to suggest that counterpossibles could *not* be true because they provide such information. They are by nature uninformative regarding counterfactual facts. For by his lights,

> Nothing can depend counterfactually on non-contingent matters. For instance nothing can depend counterfactually on what mathematical objects there are, or on what possibilities there are. Nothing sensible can be said about how our opinions would be different if there were no number seventeen, or if there were no possibility for dragons and unicorns to coexist in a single world. All counterfactuals with impossible antecedents

> may indeed be vacuously true. But even so, it is seldom sensible to affirm them. (1986, p111)

This excerpt can be understood in two ways. The first is as a broadly metaphysical thesis: modal space contains no facts about what would happen under some impossible state of affairs. This thought can seem quite natural, especially given Lewis's conviction that counterfactuals at bottom concern possible goings-on (1986, p11). An impossible counterfactual supposition fails to correspond to any way that the world *could* be; so how can there be any facts about the ways the world *would* be if it occurred? If this thought is right, it is not just that counterpossibles are true for reasons other than their conveying some substantive counterfactual information; there is no such information for them to capture or convey at all. Indeed, on the second way of interpreting this excerpt, counterpossibles lack content entirely. There is "nothing sensible," Lewis claims, to be made of even *saying* that such-and-such would happen in an impossible state of affairs.[5] Then, counterpossibles are not apt for truth at all, much less informative, non-vacuous truth. On either way of understanding Lewis's remarks, counterpossibles fail to say anything informative about the world; they are not even apt to do so.

There are arguments for the orthodoxy independent of and more recent than Lewis's, which again do not rely on any particular semantics for counterfactuals. Timothy Williamson (2007, 2018, forthcoming), the orthodoxy's most vocal contemporary defender, defends it as follows. It is a plausible hypothesis that strict implication is sufficient for counterfactual implication: "Suppose that A could not have held without B holding too; then if A had held, B would also have held" (2007, p156). So, suppose that '$\Box(A \rightarrow B)$' implies '$A \Box\!\!\rightarrow B$'. But if $\Box\neg A$, then '$\Box$(A

[5] Although see Williamson (forthcoming) on resisting the claim that counterpossibles lack content or meaning.

$\rightarrow$ B)' is vacuously true for any B. And following the hypothesis that strict implication is sufficient for counterfactual implication, A $\Box\!\!\rightarrow$ B for all B. So, that hypothesis yields—independent of any semantic model of counterfactuals—the conclusion that all counterpossibles are uniformly vacuously true.

Williamson's hypothesis about strict implication is *prima facie* plausible. But it can also be used to define metaphysical possibility and necessity in terms of counterfactuals (Williamson 2007, pp156-177), and it features prominently in an attractive counterfactual-based epistemology of modality (Williamson 2007). He therefore notes,

> Strong theoretical pressures push towards orthodoxy about counterpossibles. It is required by the simple and natural approach to the semantics of counterfactuals, and it contributes to a simple and natural picture of how counterfactuals and metaphysical modality fit together. (2018, p359)

If this argument is sound, then a counterpossible's vacuous truth will be inherited from the vacuous truth of a strict implication. Thus by Williamson's lights, too, counterpossibles' vacuous truth is characterized by their being true but not because they provide any substantive modal information. For just like a vacuously true universal generalization is by nature uninformative with regard to what, e.g., the wizards are like, presumably when '$\Box$(A $\rightarrow$ B)' is vacuously true because $\Box\neg$A, it is by nature uninformative with regard to whether any implication holds from A to B. The counterpossible that inherits this vacuous truth will accordingly be uninformative about whether there is any substantive counterfactual relationship between A and B.

There is one final argument to recount in favor of the orthodoxy. Williamson argues that the *logic* of counterfactuals supports counterpossibles' vacuous truth (2007, p174). For counterfactuals seem to allow for the intersubstitution *salva veritate* of coreferential names. From 'Had the rocket continued on that course, it would have hit Hesperus,' it seems that one can infer '…it would have hit Phosphorus'. But suppose, contra the orthodoxy, that some counterpossibles can be false, such as: 'Had Hesperus not been Phosphorus, then Phosphorus would not have been Phosphorus'. Even so, the following is presumably still trivially true: 'Had Hesperus not been Phosphorus, then Hesperus would not have been Phosphorus'. Accordingly, Williamson observes: the falsity of some counterpossibles, contra the orthodoxy, seems to require alongside it the falsity of plausible principles about counterfactuals' logical behavior. He thus forewarns the orthodoxy's detractors, saying,

> The logic of quantifiers was confused and retarded for centuries by unwillingness to recognize vacuously true universal generalizations; we should not allow the logic of counterfactuals to be similarly confused by unwillingness to recognize vacuously true counterpossibles. (2007, p175)

Thus we can see one final precedent from the logic of counterfactuals in favor of counterpossibles' vacuous truth. It is not only the familiar semantic results that exert pressure to endorse the orthodox view. Its motivations are diverse and powerful indeed, spanning from the logic of counterfactuals and other conditionals, to the metaphysics and epistemology of modality.

I shall argue that the orthodoxy on counterpossibles' vacuous truth is false. Some commentators have already argued against it in a vein similar to what I will pursue; they have noted the presence of apparently non-vacuously true or false counterpossibles in various contexts. Some metaphysicians have argued that philosophical discourse engages in counterpossible reasoning

(Brogaard and Salerno 2007, Merricks 2001, p5, Sider 1999, p339). For example, if utilitarianism is false, the conditional antecedent 'Had utilitarianism been true…' appears to be the antecedent of a counterpossible, since philosophical theories are normally taken to be necessarily false if false at all. Others have noted that ordinary usage of counterfactuals does not prejudge that the ones with impossible antecedents are vacuously true (Nolan 1997): 'Had Hobbes squared the circle,…' does not immediately seem to be vacuously true. If you believe in these examples, you will think that counterpossible vacuity must be rejected, since it threatens to undermine philosophical practice and ordinary discourse. But if, like Williamson (2007, 2018, forthcoming), you are skeptical of the veracity of our judgments about contexts like ordinary counterfactual discourse and counterfactual discourse within philosophy, these preexisting examples of counterpossibles may not suffice to convince you of counterpossibles' non-vacuity.

This paper provides an entirely new set of examples of non-vacuously true or false counterpossible conditionals. I shall argue that the success of numerous scientific endeavors—scientific explanation, model-based reasoning, and reasoning about superseded theories—requires using counterfactuals which frequently turn out to have metaphysically impossible antecedents. We shall see that by the lights of scientific practice, counterfactuals can be informatively, non-vacuously true even if they happen to have metaphysically impossible antecedents; they are, in other words, at least *apt* for non-vacuous truth or falsity. The philosophical orthodoxy on counterpossible conditionals is therefore inconsistent with science. I shall argue that this provides a conclusive reason to reject the orthodoxy. I deny, *pace* Lewis, that counterfactuals with impossible antecedents say "nothing sensible" about the world. For science is not engaged in uninformative nonsense. Nor, *pace* Williamson, will taking counterpossibles to be non-vacuously true or false plunge us into a Dark Age. Taking scientific practice seriously will not undermine our understanding of such phenomena as the counterfactual conditional.

2 Counterpossibles in Scientific Explanations

In this section, I shall provide my first example of non-vacuous counterpossibles in science: those that appear in scientific explanations. Counterfactuals in general are a frequent sight in scientific explanation. In explaining why a wine glass is fragile, you might note that it has a particular kind of needlelike crystalline microstructure that contributes to fragility. We agree that the glass's microstructure explains its fragility in part because we see that the latter counterfactually depends on the former: if the glass had a different microstructure, it would not have been as fragile. Consider another example. An engineer might explain why your electronics project overheated by noting that you mistakenly wired the batteries in series, rather than parallel. She might explain, "In a series circuit, voltage is summed, and that is why your circuit overheated—too much power. If you had wired it in parallel instead, then it would not have." Just as with the wine glass, this counterfactual helps provide the explanation of why your circuit overheated because it specifies how the explanadum's occurrence depends on the explanans. Indeed, in general, when the sort of relation that backs scientific explanation obtains – relations like causation, microphysical realization, mechanistic realization, etc. – there is a corresponding counterfactual that describes that relation. So there are countless more examples just like the wine glass and the circuit that one could give.

The role of the counterfactual conditional in scientific explanations is not limited merely to specifying that some state of affairs physically depends on another. Counterfactuals also help identify what is *relevant*, explanatorily, for one state of affairs to bring about another. This, too, is a task essential to scientific explanation. As Fred Dretske says:

> A causal explanation is, I assume, more than a specification, under some description or other, of the event's cause. An explanation requires some indication of which of the properties of the cause…underlie the cause's efficacy in producing that effect. Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is surely irrelevant to their having this effect. The glass would shatter if sounds meant something completely different or if they meant nothing at all. (1989, p1–2)

Why is the meaning of a shrill operatic verse irrelevant to explaining why it broke a glass? Simple: the breaking would have still occurred, had the meaning been different or absent; the breaking does not counterfactually depend on the meaning in the appropriate way. The pitch and amplitude of the sound waves, on the other hand, *are* relevant, and for exactly the opposite reason. This familiar counterfactual test for explanatory relevance is another way in which counterfactual facts are indispensable to scientific explanation.[6]

Finally, some philosophers of science, noting the ubiquity of counterfactual reasoning in explanation across the sciences, have developed detailed accounts of scientific explanation that are explicitly counterfactual. Here I have in mind James Woodward's highly influential "manipulability" account of causal explanation:

> An explanation ought to be such that [it enables] us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been

[6] Although for a recent alternative to the standard counterfactual litmus test, see Strevens (2008).

> different in various possible ways. We can think of this as information about a pattern of counterfactual dependence between the explanandum and explanans, provided the counterfactuals in question are understood appropriately. (2003, p11)

Woodward's view holds that what constitutes an explanation—that is, the source of any derivation or argument's explanatory power—is counterfactual information, specifically, answers to "what-if-things-had-been-different" questions. Giving an explanation of some event y by citing the occurrence of another, x, requires you to answer the counterfactual question of what the facts regarding y would have been like, if x had been different in certain ways. On this view, "explanation is a matter of exhibiting systematic relations of counterfactual dependence" (Woodward 2003, p191); the explanatory information about a phenomenon *is* counterfactual information.[7]

So, counterfactual conditionals are indispensable parts of scientific explanations. In some cases, a counterfactual that features in a scientific explanation or that describes an explanation-backing relation will have a metaphysically impossible antecedent. The most straightforward examples are cases where some substance's properties at a macroscopic scale are explained by—and hence counterfactually depend on—its structural properties at a microscopic scale.

Here is just one such example. Diamond is an exceptionally poor electrical conductor. Yet some other allotropes of carbon (graphene, for instance) are exceptionally *good* electrical conductors. Diamond and graphene alike are made of carbon. Thus there is a physical fact that seems to call

[7] Woodward's own account focuses only on causal explanation, for a variety of technical reasons. But a growing number of authors have argued that his account generalizes to a wide variety of non-causal explanations in science (such as the sort of microstructural example I give below). For discussion, see—among others—Bokulich (2008), Reutlinger (forthcoming), Saatsi and Pexton (2013), Jansson and Saatsi (2017).

out for explanation: why is diamond in particular an exceptionally poor conductor? The answer is that diamond is covalently bonded. A substance's conductivity at a macroscopic scale is explained by its constituent atoms having free electrons that can move from one atom to another, carrying electric current. Covalent bonding between atoms, which forms diamond, leaves those constituent atoms with no free electrons; other bonds, like ionic bonds, are more generous. Hence, diamond is an exceptionally poor conductor. For

> (D) If diamond had not been covalently bonded, then it would have been a better electrical conductor.

The macroscopic property of diamond's poor electrical conductivity is brought about by, and hence depends on its microstructural bonding properties. This counterfactual conditional, (D), is indispensable to providing a scientific explanation of why diamond has that macroscopic property, since it identifies and describes that dependence relation.

Now, the fact that (D) features in an explanation of a particular substance's properties, and describes the relationship between two of that substance's properties, suggests that a *de re* reading of its apparent referring terms is its correct interpretation. And considered *de re*, diamond has its actual physical microstructure (a kind of cubic lattice) as a matter of metaphysical necessity. Nothing could qualify as diamond, i.e., that very substance, without having that microstructure. In other words, it is metaphysically impossible for diamond to be differently chemically bonded.

Moreover, within materials science and chemistry, diamond is individuated from other allotropes of carbon by its bonding structure, and individuated from other hard substances by its chemical composition. So, even if one considers the *de dicto* content of 'diamond', as a theoretical term, it is not the case that some other substance could qualify as diamond and fill its theoretical roles without having its actual composition and microstructure. Indeed, it would be straightforwardly contradictory to suppose that the following theoretical role—"…is an allotrope of carbon, covalently bonded in a cubic lattice"—could be filled by any non-covalently-bonded substance, carbon-composed or otherwise. To reiterate, it is metaphysically impossible to suppose that diamond could have been differently bonded.

The antecedent of the counterfactual that features in the explanation of diamond's poor conductivity—'If diamond had not been covalently bonded…'—is metaphysically impossible. So, I conclude that that counterfactual is, in fact, a counterpossible. But let me forestall one possible misunderstanding. Because it is metaphysically impossible for diamond to have a different bonding structure than it in fact has, you might suspect that what makes the above counterfactual true is not diamond possibly being such-and-such a way. Instead, some merely-possible superficial substitute ("schmiamond") does so.[8] If so, you might deny that (D) is a counterpossible, since it is not impossible for some alien substance, schmiamond, to be ionically bonded. However, this "schmiamond strategy" is actually no objection to what I have said. My only aim was to show that that counterfactual, (D), has an impossible antecedent. But the primary reason you might adopt this "schmiamond strategy" is if you agreed that diamond could not be non-covalently bonded, i.e., if you agreed with me that (D)'s antecedent is metaphysically impossible. Supporters of the schmiamond strategy are thus already supporters of my conclusion thus far in this section: that (D)'s antecedent is in fact metaphysically impossible.[9]

[8] This sort of strategy is familiar from how Kripkeans about natural kinds or essentialists about causal profile will interpret the apparent conceivability of natural kinds or dispositional properties instantiating different properties. See, e.g., Bird (2007); Kripke (1980); Shoemaker (1980).
[9] Thanks to Alison Fernandes and Jessica Wilson for, on separate occasions, flagging to me the importance of discussing the "schiamond strategy".

I have argued that in explaining why diamond is a poor conductor, scientific practice uses a counterpossible conditional: (D) provides the explanation of diamond's insulativity by describing that substance and the dependence relations its explanandum-property stands in. And (D) is clearly non-vacuously true, if true at all. It identifies and describes an empirical fact, namely, the physical dependence of diamond's macroscopic insulativity on its microstructural bonding properties.[10] So, there are non-vacuous counterpossibles in at least one scientific context: in scientific explanations.

But there is room for my argument to be a bit more careful here. I have said that (D) is indispensable to explaining why diamond is a poor conductor. And you might think I have drawn this conclusion too quickly. You might ask, "Isn't the explanation's success really due to a general, lawlike principle that relates a substance's conductivity to its microstructure?" Maybe the explanation requires subsumption under a general law or regularity, and goes like this: "Diamond is a poor conductor because it is covalently bonded, and *any* covalently bonded substance is such that if it were not bonded that way, it would be a better conductor." If so, (D) might be eliminable as part of the explanation. Then, I would be mistaken in identifying (D) as a core part of the scientific explanation of diamond's conductivity—instead of, e.g., a "lawlike" counterfactual that describes the dependence-relation at a general level. And I would not be entitled to my conclusion that counterpossibles are sometimes indispensable to scientific explanations.

[10] Here it may be unavoidable to suggest, given the discussion of 'worlds' in the previous paragraph, how counterfactuals about these dependence-relations should be interpreted. Recent literature (Jansson & Saatsi (2017); Kistler (2013)) suggests that these kinds of counterfactual relations can be interpreted in interventionist terms, despite being non-causal relations.

My reply is as follows. First, it is not clear that this response actually helps avoid my conclusion that counterpossibles sometimes feature indispensably in scientific explanations. For that general principle, too, is plausibly a counterpossible conditional. Remember that it says of every covalently bonded substance that if it had been bonded differently, it would have been a better conductor. But won't the molecular microstructure of many covalently bonded substances – water, methane, ammonia, etc. – be essential to those substances? Look at it this way. If a molecule that is actually covalently bonded were to be differently bonded, wouldn't it just be a different *kind* of molecule? Seems so. Then, this objection simply trades one impossible antecedent for another, since it will turn out to be metaphysically impossible that any substance which is covalently bonded could have failed to be covalently bonded. And this objection will not successfully avoid my claim that counterpossible conditionals feature indispensably in some scientific explanations.

At any rate, I deny that (D) can be eliminated from a successful explanation of diamond's poor conductivity. To see why, note that the particular phenomenon being explained is one of diamond's properties. And plausibly, any scientific explanation of that fact is incomplete until one specifies how the facts that are said to explain it really do make a difference to whether it occurs. That is, following Woodward (2003), the provision of any explanation of some fact is incomplete until one specifies, counterfactually, how that very fact would have been different under various circumstances. So the general principle—that *any* covalently bonded substance is such that were it differently bonded, it would be a better conductor—cannot fully replace (D) in an explanation of diamond's poor conductivity, even if it is paired with the observation that diamond is covalently bonded. For that other observation says nothing at all about a dependence or difference-making relation, or for that matter, about the targeted explanandum. And the general statement only describes covalent bonding and conductivity as general property types. (After all, *ex hypothesi* it is a general statement that says nothing in particular about diamond or any other particular substance's properties.) So it says nothing about the target explanandum or how it would have differed under various circumstances. Of course, together with the other observation, it derives a statement that does, namely, (D). So giving the explanation might require that generalization for that and other explanation-completing reasons. But it cannot replace (D) or do the same explanatory work (D) does. Only (D) manages to fill in the relevant explanatory details with regard to diamond's poor conductivity.

(D)'s antecedent is metaphysically impossible. So any complete explanation of diamond's poor conductivity relies on what is in fact a counterpossible conditional describing its properties. The explanation shows how diamond is such that if had not been covalently bonded—even though in fact it *must* be—it would have had different conductive properties. According to the orthodoxy on counterpossibles, this cannot be non-vacuously true about diamond unless diamond really could have been differently bonded. Therefore, the orthodoxy is mistaken.

3 Counterpossibles in Idealized Scientific Models

In this section, I shall argue that non-vacuous counterpossibles are prevalent in another scientific context: in reasoning with and testing idealized models. It is a familiar observation that some classical mechanical models of motion represent inclined planes as frictionless; other models represent pendulums as suspended by massless strings, gases as ideal "billiard balls", planets as point masses, and so on. There obviously are no such idealized objects. Nearly all of these idealizations' truth would comprise massive violations of natural law; and as I shall argue, one can find idealizing assumptions that are also metaphysically impossible. Yet idealization is indispensable to scientific practice because it makes immensely difficult physical problems computationally tractable and calculable to close approximations of their actual values.

A central philosophical issue surrounding scientific idealization concerns the source of the epistemic utility of idealized models. Why can we learn anything at all about pendulums from idealized models of pendulums with massless strings, given that actual pendulums definitely do not have massless strings? And more generally, under what conditions is an idealization informative with regard to actual, physical goings-on?

There is widespread agreement that idealized models are informative when the observed behavior of their target objects closely approximates, at a counterfactual level, the idealized objects' behavior.[11] Why can we manage to learn anything from an ideal, massless-string model of a pendulum? Well, it is because the trajectory of an actual pendulum is quite close to what its trajectory would be like, if its string *were* massless. Thus we can reason successfully about the actual object using the model.[12] Ditto for frictionless planes, planetary point masses, and the like. Here is what Stathis Psillos has to say about the issue:

[11] There are many more supporters of this widespread view not directly quoted below. It is articulated and defended in a recent book by Michael Shaffer (2012). David Lewis (1986, p26-27) takes the content of idealized models to be counterfactual in the sense that the counterparts of the target systems are their truthmakers. Peter Godfrey-Smith (2009) and others who treat scientific idealizations as fictions agree that the information provided by idealized models is modal. Alisa Bokulich (2011) defends a view of model-based explanation according to which idealized models are explanatory when their counterfactual structure bijects to the counterfactual structure of the target system. There are still others not mentioned here who defend this modal view of idealization.

[12] See my (2017) for an account of the link between the informativeness of successful idealizations and their counterfactual approximation of the target physical systems.

> How can models give causally relevant information about the world, if they are not themselves part of the causal order of the world? Most models are such that had there been concrete entities that matched them exactly, these entities would have been part of the causal order of the world. So models can provide causally relevant information about worldly systems because on their basis certain counterfactual conditionals can be settled. (2011, p16)

Likewise, Ernest Adams writes:

> The orbits of the planets are often calculated on the counterfactual assumption that these bodies are mass points between which forces of gravitation act. It is plausible that a counterfactual conditional is implicit in this reasoning because it might be explained as: "The orbits of the planets are similar to those of mass points and *if they were mass points* then their orbits would be such-and-such; therefore their orbits are similar to such-and-such." (1993, p5)

In other words, an idealized scientific model (say, of krypton as an ideal, "billiard-ball" gas) represents its target system in part by representing the physical system *as if* it were ideal, i.e., counterfactually:

> If krypton were an ideal gas, then it would obey $PV = nRT$ perfectly.

When we reason using ideal gas models to predict or explain krypton's behavior, we appeal to the truth of this counterfactual. Likewise for all sorts of other common idealizations.[13]

[13] I say "common" idealizations because there are some idealizations that do not represent their targets in this counterfactual, *as-if* way. If cases of "minimal modeling" in science (e.g., the Lotka-Volterra model, the Ising model) are understood of cases of abstraction rather than idealization, they will fall outside of the scope of this modal interpretation of idealization. The familiar cases of (so-called "Galilean") idealization are very plausibly understood in these counterfactual terms. For some discussion of the various types of idealization in science, see Weisberg (2007).

Some counterfactuals of this sort have metaphysically impossible antecedents.[14] Consider one case. Certain models of water useful for modeling wave propagation represent it as a continuous, incompressible medium (e.g., the Navier-Stokes equations for fluid flow). These models represent water *not* as being composed of discrete molecules, but instead, as if it were continuous. This is obviously an idealization. But it is a useful and informative idealization because

> If water were a continuous, incompressible medium, then it would behave as the Navier-Stokes equations describe.

In other words, the wave-behavior of *actual* water closely approximates how water would behave, if it really were continuous. And when put to their proper purpose, these models are quite successful.[15]

The antecedent of the above counterfactual conditional about water is metaphysically impossible. As is widely held, necessarily, water is identical to H_2O. If so, it is metaphysically impossible for water to be a continuous and incompressible medium. But even if you are not convinced of that *de re* necessary identity, there are still powerful reasons to think that that counterfactual's antecedent is metaphysically impossible. It is metaphysically impossible for the functional and theoretical roles that water fills, which comprise the *de dicto* content of 'water', to be filled by some substance which is not composed of discrete particles, i.e., which is continuous and incompressible. For instance, nothing continuous, non-particulate and incompressible could act as a solvent for particulate solids, or could freeze into the many phases of ice (which are all non-continuous and instantiate some particular molecular bonding

[14] In fact, you might suspect that *all* of the idealizations I have mentioned—the ideal gases, the massless strings, etc.—are metaphysically impossible. David Lewis (1986, p26-27) seems to disagree; he takes them to be possibilia.
[15] See Morrison (1999, pp53-60) for discussion of the history and physics in these cases.

structure). Any such medium would fail to fill some of the theoretical roles associated with water. And that is simply to say that it is metaphysically impossible even for anything that functionally or causally qualifies as water to have been continuous and incompressible. So, the counterfactual supposition that water is continuous and incompressible, which is built into those models of wave behavior, is metaphysically impossible. This is my second example of a counterpossible conditional in a practice central to our mature sciences, i.e. scientific modeling.[16]

Taking a look at what scientific modeling practices are like leans strongly in favor of taking some of these counterpossibles to be non-vacuously true and others to be false. For scientific modeling practices include the testing of various models against the physical systems they represent and eventually confirming or disconfirming a model. This means taking certain "idealizing counterfactuals" as non-vacuously true, and taking others to be false. In cases where the idealizing counterfactual is a counterpossible, this *a fortiori* means taking some counterpossibles to be non-vacuously true, and others to be false.

Let me provide an example. Suppose a group of scientists constructs a pair of mathematical models of some water, M_1 and M_2, that each represent it as a continuous, ideal fluid, but differ with regard to the viscosity they ascribe to it. These scientists are, let us suppose, interested in how waves behave when they crash up against a certain kind of rough surface. In testing these models, what they will do is closely observe how water *actually* behaves when its waves crash against such a surface, and check the models' predictions against those observations. Imagine

[16] Thanks to Anjan Chakravartty for discussion here and in this whole section.

that M_1 is found to closely approximate the observed behavior, while M_2's predictions are a bit farther off. In rejecting M_2 in favor of (tentatively) accepting M_1, the scientists will thus take

> "If water were a continuous, incompressible medium, then it would behave as M_2 predicts"

to be false, while simultaneously taking

> "If water were a continuous, incompressible medium, then it would behave as M_1 predicts"

to be true. Just as before, both of these counterfactuals are counterpossibles. Yet they differ in their truth-value. If the orthodoxy on counterpossibles were correct, then both conditionals would be true, and vacuously, too. Thus there is evidence that the orthodox position is mistaken. Some counterpossibles are false, while others are non-vacuously true. Indeed, the mere fact that it seems like a worthwhile scientific endeavor to test these models against reality despite the fact that they rest on counterpossibles is evidence against the orthodoxy. The worthwhileness of testing these models (and hence the truth-value of their idealizing counterfactuals) against reality suggests that counterpossibles are at least apt for falsity and non-vacuous truth.

4 Counterpossible Reasoning about Superseded Scientific Theories

I shall argue, in this section, for a third and final set of examples of non-vacuous counterpossible conditionals in science. I shall argue that scientific reasoning employs non-vacuous counterpossibles when reasoning about the contents of superseded scientific theories. This is an extension of the observation that counterfactuals help us make sense of the content and success of idealized models. For just like reasoning with idealized scientific models consists at least partially in counterfactual reasoning ('If Jupiter were a point mass…'), so too does evaluating the content and truth-value of a theory about some phenomena or another ('If classical mechanics had been true…'). Often, a theory of some phenomenon is simply a representational model or set of representational models of that phenomenon, together with the assumptions that were used in constructing or deriving the model(s). Hence, in many cases, 'If such-and-such a theory had been true…' is equivalent to 'If these objects were as such-and-such a model describes…'.

Now, counterfactual reasoning about a theory's truth can take place in simple contexts, such as simply describing a theory's empirical content. For instance, you might note that if the ancient Ptolemaic theory of celestial spheres had been true, the celestial spheres themselves would have been unobservable entities (Chakravartty 2007, p3). But the most widespread use of counterfactual reasoning in evaluating the truth of a theory is in explaining why we know a particular false theory in the history of science to be false. Consider one such example:

> (B1) If Bohr's theory of the atom had been true, then an electron's angular momentum L in the ground state would have been observed at $L = \hbar$ (i.e. the Planck constant).[17]
>
> (B2) It is not the case that an electron's angular momentum L in the ground state is observed at $L = \hbar$.
>
> (B3) Therefore, Bohr's theory of the atom is false.

It is known from experiment that the angular momentum of an electron is 0 in the ground state. Thus Bohr's theory (whose core is the familiar "planetary" model of the atom) is known to be false. Bohr's theory has many other shortcomings, and our reasoning about them can be reconstructed just as above.

[17] As just mentioned, this conditional is equivalent to the "idealizing conditional", 'If atoms were Bohr atoms…', which is a counterfactual, not a strict or material implication. I argue in more detail below (§4.1) that conditionals like (B1) should be read as counterfactuals.

This famous example exemplifies the sort of counterfactual reasoning commonplace when explaining why a certain false theory is known to be false: If it had had been true, we would have observed thus-and-so; we do not, so we know that it is false. Yet there is an important observation to make about this commonplace pattern of reasoning. I take it that "theory-evaluating" counterfactuals used in these sorts of contexts are non-vacuously true, if true at all. For unless (B1) managed to convey some actual information about what would result from Bohr's theory being true, we would not be able to use an empirical observation like (B2) to rule out that theory's truth and thereby learn something substantive about the world. That is, more generally: in order for this commonplace pattern of reasoning to be epistemically fruitful, theory-evaluating conditionals must describe genuine relations of counterfactual dependence and implication. They must, in other words, be non-vacuously true. Accordingly, if one can show that some theory-evaluating conditional is a counterpossible conditional, that will suffice to show that some counterpossibles are non-vacuously true.

As it turns out, (B1) is a counterpossible. It is logically impossible—and *a fortiori* metaphysically impossible—that Bohr's theory could have been true. Bohr's theory of the atom is a canonical example of an inconsistent scientific theory. Because it rests on both classical and quantum physical assumptions, certain parts of the theory represent orbiting electrons as radiating energy as they move about, while other parts represent orbiting electrons as non-radiative. But (B1)'s antecedent – 'If Bohr's theory had been true…' – concerns the truth of the theory as a whole. Taken as a whole, Bohr's theory contains a contradiction; so its truth is logically impossible. That is, it is logically impossible that there could have been Bohr atoms. Evaluating the truth of Bohr's theory thus requires taking on not just a counterfactual supposition, but a counter*possible* one. And Bohr's theory is only one such example. There are many other examples of logical inconsistency in scientific theories—Lorentz's early theory of the electron (Vickers 2013, §7.4), the old quantum blackbody radiation theory (Norton 1987), Kirchhoff's diffraction theory (Saatsi and Vickers 2011), etc. In all of these cases, the theory-evaluating conditionals like (B1) will turn out to have logically impossible antecedents. So they will all be examples of non-vacuously true counterpossible conditionals in scientific practice.

4.1 Defending the Counterfactual Reading

Maybe you are not convinced. Unlike my previous two examples of counterpossibles in science, this example admits of an objection regarding the nature of the conditional in question. Why take theory-evaluating conditionals to be counterfactuals rather than, say, strict or material conditionals, or as expressing logical entailment?

Before launching into a defense of why we should read these conditionals as counterfactuals, I want to note that reading them as non-counterfactual conditionals will not help explain why they appear to be apt for non-vacuous truth and falsity. Suppose that (B1) actually expresses material or strict implication. In either case, it will still be vacuously true, and anything that shares its antecedent will also be. For its antecedent is, again, logically impossible, and so necessarily false. But ordinarily we judge that conditionals expressing material and strict implication are vacuously true if their antecedents are, respectively, false or necessarily false. (Defenders of counterpossibles' vacuous truth, e.g., Williamson (2007, p156), will certainly agree.) So a reading of (B1) as a material or strict conditional will also render it vacuously true.

So, reading (B1) as another kind of conditional cannot explain why it appears to be non-vacuously true.

At any rate, theory-evaluating conditionals like (B1) deserve to be read as counterfactuals. I have already given one preliminary reason: evaluating the truth of a theory is often tantamount to evaluating the truth of its central representational models, and it is *counterfactual* conditionals that explain the epistemic utility of idealized models. Another tentative reason in favor of the counterfactual reading is that these history-oriented conditionals only appear in contexts when the speaker knows full well that the antecedent is false. A physics instructor, for instance, might teach her students, "If the classical mechanics of the atom had been true, then the Stern-Gerlach experiment wouldn't have detected any abnormalities in angular momentum caused by spin." Since the instructor knows that the classical theory is false and aims to explain to her students *why* we know so, the utterance of this conditional's antecedent is plausibly interpreted as her entertaining a counterfactual supposition.

Here is a final, substantial defense in favor of the counterfactual reading. It is well-known that transitivity and strengthening the antecedent fail for counterfactuals, while they succeed for strict and material implication as well as logical entailment. If theory-evaluating conditionals are not counterfactuals, but instead express implication or entailment, then transitivity and strengthening the antecedent should be valid for them. But they are not valid.

Consider first the failure of transitivity. Start with:

(i) If classical electromagnetic theory had been true, then the Fresnel equations for reflection and refraction would be highly empirically successful.

Augustin Fresnel contributed to late-19th-century optics a set of equations that describe how light is reflected and refracted when moving between media of differing refractive indices (e.g., glass and air). They predate Maxwell and Faraday's classical electromagnetic theory, but they survive intact in that later theory (cf. Worrall 1989). Thus the truth of Maxwell's theory would have been sufficient for the predictive success of those equations; hence, (i). Now consider the following:

(ii) If the Fresnel equations for reflection and refraction are highly empirically successful, then we have compelling empirical evidence for Fresnel's wave-aether theory of light.

This is also true. The Fresnel equations are at the core of an obsolete theory of light, Fresnel's own wave-aether theory. But Fresnel's equations for reflection and refraction, and his theory of diffraction (Saatsi and Vickers 2011), were highly empirically successful. The remarkable success of Fresnel's equations provided much confirming evidence in favor of his aether theory of light; hence, (ii). Now, with both (i) and (ii)'s truth in mind, recall that if theory-evaluating conditionals are not counterfactuals, then transitivity will be valid for them. And if transitivity is valid for (i) and (ii), then the following is true:

(iii) If classical electromagnetic theory had been true, then we would have compelling empirical evidence for Fresnel's wave-aether theory of light.

But there is good reason to think that (iii) is false. Classical electromagnetic theory and the earlier wave-aether theories of light, despite both being false, disagree substantially regarding the nature of light. Wave-aether theories took light to be a wave that propagated through an aether; Maxwell's theory takes light to be a kind of electromagnetic radiation. It is simply not obvious that the truth of Maxwell's theory would provide evidence for the existence of a luminiferous aether. In fact, I would think that if it really had been true, we would have empirical evidence *against* the aether theory. So transitivity seems to fail for theory-evaluating conditionals.

Strengthening the antecedent also fails. The following inference form is valid for strict and material conditionals. And logical entailment, being monotonic, is analogous.

$(P \rightarrow Q)$

Therefore, $((P \wedge R) \rightarrow Q)$

This inference fails to be valid for theory-evaluating conditionals. Consider:

(iv) If classical electromagnetic theory had been true, then alterations to the wavelengths of light shining on a sample of aluminum would have induced a change in the sample's rate of photoelectron emission.

This conditional is true. The classical theory predicts that the rate of photoelectron emission is affected by the *frequency* of the incident light. The frequency of a beam of light is inversely correlated with its wavelength; hence, (iv): the classical theory predicts that the rate of photoelectron emission would be affected by changes to the incident light's wavelength. (We know the classical theory to be false because there is in fact no difference to photoelectron emission under that intervention—Einstein won his Nobel Prize for developing an early quantum theory that explained why.) Now, if this conditional, (iv), is not a counterfactual but instead expresses implication or entailment, then strengthening its antecedent should yield another true conditional:

(v) If classical electromagnetic theory had been true (but alterations to a beam of light's wavelength had no effect on its frequency), then alterations to the wavelengths of light shining on a sample of aluminum would have induced a change in the sample's rate of photoelectron emission.

As mentioned, the classical theory generates that prediction in part because a beam of light's frequency can be manipulated through its wavelength. But if frequency and wavelength were unrelated, then even by the lights of the classical theory, changes to wavelength would not affect photoelectron emission. Therefore, this conditional, (v), is false. But if it were a strict or material conditional, it would be true, given the truth of (iv) above. So both transitivity and antecedent-strengthening fail for theory-evaluating conditionals. This is further evidence that they are counterfactuals. Together with my earlier observations in favor of the counterfactual reading, I thus conclude that theory-evaluating conditionals like (B1) are counterfactual conditionals. My third example of counterpossibles in science is therefore vindicated. Reasoning about a theory's truth-value requires using non-vacuously true counterpossible conditionals.

4.2 Additional Examples of Theory-Evaluating Counterpossibles

In fact, I have not just vindicated my earlier example regarding Bohr's theory of the atom, I have supplemented it. As I shall now argue, all of these example counterfactuals I have just given regarding the truth of Maxwell's electromagnetism also have metaphysically impossible antecedents. The classical theory represents light as a special kind of purely continuous wave, i.e. electromagnetic radiation, whereas light is *not* a continuous wave, but instead is a wave-particle. So, their shared counterfactual antecedent, 'If classical EM theory had been true…' is plausibly read as representing light *de re* as a fundamentally different sort of thing than what it actually is. If so, then it is metaphysically impossible (for familiar reasons: that wave-particle stuff could not have been a radically different sort of thing). Then, all of these examples of counterfactuals regarding Maxwell's theory – the true (i) and (iv) and the false (iii) – are counterpossibles with metaphysically impossible antecedents.

Here, however, a more thorough argument is needed. There are, again, alternate readings of ‘If classical EM theory had been true…’ under which it is not obviously metaphysically impossible. A *de dicto* reading interprets ‘light’ in the classical theory as a theoretical term, such that ‘If classical EM theory had been true…’ effectively expresses:

> “If the causal and theoretical roles associated with light had been filled by the classical, pure waves described in Maxwell’s theory, then…”

This conditional’s antecedent is not obviously metaphysically impossible. Compare it to the *de re*, “referring” reading. It is presumably metaphysically impossible that what *in fact* plays the causal and theoretical roles of light—i.e. *that very stuff*, light—could have been different in radical and fundamental ways. But prima facie it is not impossible that something else—say, pure waves of some variety—could have filled those same causal and theoretical roles.

Nevertheless, reading ‘If classical EM theory had been true…’ this way still renders it metaphysically impossible. For the desired interpretation does not describe a state of affairs in which *some* merely possible pure wave fills light’s causal and theoretical roles. It describes a state of affairs in which the pure waves of *Maxwell’s theory* fill that role. And the pure waves of Maxwell’s theory have their own causal and theoretical roles, many of which are are incompatible with the causal and theoretical roles light actually fills. (For example, the energy transmitted by light is quantized; the classical theory describes it as continuous. Refraction and reflection, as actually observed and accurately described by quantum electrodynamics, admit of multiple solutions; the equations of the classical theory have unique solutions.) Because of these inconsistencies between the causal and theoretical roles that light actually plays and the theoretical roles Maxwell’s theory ascribes to it, this reading of the counterfactual renders its antecedent metaphysically impossible, maybe even logically impossible. So, under this reading of ‘If classical EM theory had been true…’, it is still a counterpossible conditional.

At this point, one might respond that the theoretical content of ‘light’ in the suggested reading above is not as detailed as I have suggested. The thought might go as follows. When we are reasoning counterfactually about the truth of Maxwell’s theory, we do not mean to include in the meaning of ‘light’ *all* of the theoretical and causal roles that light plays. We instead mean the familiar *observational* roles of light. That is, by ‘the theoretical roles associated with light’, we mean something like, “…reflects off of metals, creates rainbows, and illuminates dark hallways”. Thus we have a final suggested reading of ‘If classical EM theory had been true…’ (and equivalently, ‘If light had been a Maxwellian pure wave…’). On the suggested reading, it expresses,

> “If the *observational* roles of light had been filled by Maxwellian pure waves…”

And since it is not obviously false that a Maxwellian pure wave could have filled light’s ordinary, observational roles, this reading ostensibly avoids rendering the antecedent of ‘If classical EM theory had been true…’ metaphysically impossible in the way that the previous reading did.

I do not think this is an admissible reading. There are two reasons. First, it seems objectionably arbitrary to take the theoretical meaning of ‘light’ to include only light’s observational roles. Here this is a mere instance of a more familiar problem: it is hard to draw a principled distinction between something’s observational and non-observational, theoretical roles. I mean, on the suggested reading, what is the observational content of ‘light’? Just its directly observable, “folk” content, like reflections and rainbows, right? If so, observations like those gained from the double-slit experiment will not be included in the observational roles of light, since they support a hypothesis about light that seems to be quite at odds with the folk concept of light (viz., the hypothesis that it is a wave-particle). But ruling out those observations from contributing to the theoretical meaning of ‘light’ simply because they are taken from experiment or confirm a strange-sounding hypothesis seems objectionably arbitrary and indeed, unscientific. It seems demonstrably false that the concept of light that appears in scientific practice is never more detailed than nor ever diverges from the content of the folk concept. So it seems false to suggest that the content of ‘light’ is limited to its observational roles, especially when we are engaged in a piece of substantive scientific reasoning, e.g., counterfactually evaluating the content of a theory like Maxwell’s electromagnetism.

Secondly, even if this “observational role” reading worked for ‘light’, it seems to fail at a more general level. The reading can sound quite natural regarding ‘light’, where the phenomenon in question is associated with directly observable behaviors (e.g. reflections and rainbows). But if it is the correct way to read theoretical terms, then it should generalize to other cases. It should equally apply to, for instance, the term ‘atom’ in ‘Had atoms been Rutherfordian…’, or ‘space’ in ‘Had space been Newtonian…’. This interpretation would yield, respectively, “If the observational roles of atoms had been filled by Rutherfordian atoms…” and “If the observational roles of space had been filled by Newtonian space…”. But this cannot be an adequate reading of those theory-evaluating counterfactuals. For neither the structure of atoms nor the fundamental geometry of space are directly observable in anything resembling the way that light is. There are no direct observational roles associated with those phenomena in the way that the suggested reading requires. Thus it seems false that interpreting theoretical terms in terms of their *observational* roles could generalize to ‘atom’ or ‘space’. So I deny that the “observational role” interpretation of theoretical terms will work at any general level. Interpreting ‘light’, ‘atoms’, or ‘space’ as theoretical terms, likely more closely resembles the usual, previously-suggested functional reading. And I have already shown that interpreting those terms as theoretical terms in that way does not avoid rendering their antecedents metaphysically impossible.

But that is not to say that the observational roles of light fail to feature in the meaning of ‘light’ in claims like, ‘If light had been a Maxwellian pure wave…’ at all. Those coarse-grained observational roles can contribute to the meaning of such a conditional by helping fix the term’s referent. To see why this is important, notice that contexts where one is explaining why we know Maxwell’s theory to be false, it would not make much sense to say something like, ‘If light had been Maxwellian, then we would have observed…’ unless Maxwell’s theory was supposed to have been a theory of the very same thing that ‘light’ actually picks out. The same goes for any other historical theory. Referential continuity between the central terms in our theories is an important necessary condition on scientific progress and our being able to evaluate differences between older and newer theories.[18] I take it, then, that the observational roles of light can help in understanding what we mean by such claims as, ‘If Maxwell’s theory had been true…’, or ‘If light had been an aether-wave…’ by establishing exactly what those theories were meant to be theories *of* in the first place. The observational roles can, after all, fix the referent of ‘light’ in a way that we can grasp pretheoretically. For by using them to fill in some details, we can understand a theory-evaluating conditional as follows:

> “If the correct theory of whatever it is that causes rainbows, reflections, and the like had been Maxwell’s theory, then we would have observed such-and-such.”

[18] See, among many others, Psillos (1999, §11)

And this seems to be exactly what we mean in uttering claims like, ‘If classical EM theory had been true…’. So, insofar as the observational roles of light contribute to the meaning of a theory-evaluating conditional about an historical theory of light, we should think that they do so by fixing the reference of that conditional’s central terms. Now, this point is important. It means that those theory-evaluating counterfactuals are in fact counterpossibles, exactly as I have argued. For on this reading, the term ‘light’ in a claim like, ‘If light had been a Maxwellian pure wave…’, manages to refer *de re* to whatever in fact plays the theoretical and observational roles of light. But it is metaphysically impossible for whatever in fact plays the theoretical and observational roles of light to be a fundamentally different sort of thing (i.e. a pure wave rather than a wave-particle). In other words, such a theory-evaluating counterfactual says that if—*per impossibile*—Maxwell’s theory had been true of light, then we would have observed light doing certain things that we do not actually observe. And of course, it is only thanks to counterfactual conditionals like that one that we can explain why we know an otherwise-successful historical theory, like Maxwell’s electromagnetism, to be false. But that conditional is a counterpossible. So it is only thanks to counterpossible conditionals that we can complete that important explanatory task.

I have just argued that ‘If classical EM theory had been true…’ is a counterpossible conditional under all of its available readings. But presumably, there is nothing special about the example of classical electromagnetic theory in particular. There are any number of false theories in the history of science that represent the role-players of physical phenomena as being fundamentally different entities than they possibly are. Newton’s theory of space and time, the many false theories of the atom, and the many false theories of chemical elements immediately come to mind as plausible examples of metaphysically impossible theories. And my arguments that ‘If classical EM theory had been true…' is a counterpossible will equally apply to the counterfactuals used to evaluate *those* theories. So, the presence of counterpossible conditionals in theory-evaluating contexts is not limited merely to those theories that turn out to be logically inconsistent (e.g., my first example of Bohr's theory of the atom). Non-vacuous counterpossibles will turn out to be virtually ubiquitous when evaluating false scientific theories.

5 Concluding Remarks

This concludes my arguments for the prevalence of non-vacuous counterpossibles in science. I can think of few practices more central to science than providing scientific explanations, constructing and evaluating models, and reasoning about the content of scientific theories. Counterfactuals are indispensable to each of these practices. I have just argued that these counterfactuals frequently turn out to have metaphysically impossible antecedents. So, many scientific endeavors require the non-vacuous truth or falsity of counterpossibles for their success.

Indeed, within scientific practice, the metaphysical possibility or impossibility of a counterfactual's antecedent clearly makes no difference to whether it is non-vacuously true or false. It is not as if one needs to check whether the antecedent of a counterfactual is metaphysically possible before thinking that, e.g., a scientific explanation that uses it manages to say something informative, or even amounts to more than nonsense. Yet the orthodoxy on the counterfactual conditional suggests otherwise. The orthodoxy suggests—to recall David Lewis—that "nothing sensible" can be made of a counterfactual if its antecedent is metaphysically impossible. The orthodoxy even suggests that it is part and parcel of the very phenomenon of the counterfactual conditional that counterfactuals with impossible antecedents are uninformatively, vacuously true. [19] So I think that the orthodoxy must be rejected. It is incompatible with our best scientific practice, and delivers the implausible verdict that our best science is routinely and ineliminably engaged in nonsense.

It is not simply because I respect the authority of our mature sciences that I think the orthodoxy must be rejected. The orthodoxy's incompatibility with the way that science uses counterfactuals means that it fails to accurately capture the phenomenon it was meant to capture, viz., the counterfactual conditional. That is the real reason why the orthodoxy must be rejected. Look at it this way. The counterfactual itself, as a "phenomenon", is interesting and worth theorizing about in large part because of its indispensability to explanatory reasoning (Kment 2014, Woodward 2003), causal reasoning (Lewis 1973, Paul and Hall 2004), constructing and evaluating models, theories, the like (Shaffer 2012), and many other ordinary and scientific contexts. That is what the phenomenon *is*. The counterfactual is not merely a toy of philosophers and linguists. Hence, if one's philosophical or linguistic theory of how the counterfactual works is incompatible with the way it is used in those core contexts, that is a sufficient reason for the theory to be rejected. For that would mean that that theory fails to accurately capture the very phenomenon it was built to account for. And it is the phenomena that must be saved, not the theories built to account for them. A theory of the counterfactual fails to save the phenomenon if it is inconsistent with the way the counterfactual is used in those contexts, or worse, if it declares that those uses constitute an an incorrect way of using the counterfactual.

[19] Cf. Williamson (2007, p158, emphasis added): "...B $\square\!\!\rightarrow$ C is vacuously true if and only if B is impossible (this is *almost a definition* of "vacuously" for counterfactuals)..."

Science plausibly contains the most paradigmatic and successful examples of counterfactuals in explanatory reasoning, in causal reasoning, and in evaluating theories and models. In other words, if there is anywhere that the "phenomenon" of the counterfactual can be found, plausibly it is within scientific practice. Now, I am not suggesting that scientists or scientific practice should be taken as authoritative on, e.g., the semantics or the metaphysics behind the counterfactual conditional. But the scientific community are experts at actually giving scientific explanations, and evaluating and constructing theories and models. So, plausibly, they are authoritative on actually *using* the counterfactual in some of its paradigmatic use-cases. Yet the philosophical orthodoxy on the counterfactual, which takes counterpossible conditionals to be indiscriminately vacuously true, declares that science routinely uses the counterfactual incorrectly in those very cases. This is an absurd conclusion. If a core feature of the phenomenon of the counterfactual is its role in providing explanations, and counterfactuals in scientific explanations sometimes have metaphysically impossible antecedents, then it is simply a datum about the phenomenon of the counterfactual that counterfactuals are sometimes informatively true even if their antecedents are impossible. If a core feature of the phenomenon of the counterfactual is its role in constructing and evaluating theories, models, and the like, and the counterfactuals used to evaluate scientific theories and models routinely have metaphysically impossible antecedents, then that provides the same datum about the phenomenon. The philosophical orthodoxy on the counterfactual conditional simply fails to save the phenomenon. That is a failure sufficient to warrant its rejection.

I said at the beginning of the paper that my only goal was to establish, with new examples, the non-vacuity of some counterpossible conditionals. But this is hardly the end of the line. Moving forward, what is needed is a thoroughgoing philosophical treatment of the counterfactual conditional that accommodates the data that counterpossibles can be non-vacuously true or false. There were a diverse and powerful set of motivations that led to the orthodoxy, and they will all require careful attention. The most familiar motivation in favor of counterpossible vacuity was, of course, the semantics for the counterfactual in terms of possible worlds. So, naturally, the first place that commentators have looked to begin building a counterpossible-friendly treatment of counterfactuals is their semantics. Adding a set of *impossible* worlds to the preexisting possible worlds has proven to be the most popular option (Brogaard and Salerno 2013, Nolan 1997, Jenny 2016, Vander Laan 2004, and many others). Another strategy might be to de-absolutize the distinction between what is metaphysically possible and impossible (Hellie, Murray, & Wilson (forthcoming); Murray & Wilson (2012), Clarke-Doane (forthcoming)).[20] But if you are persuaded that scientific practice provides some of the clearest paradigm cases of counterfactual reasoning, then you may be inclined to disavow "worlds" altogether and adopt a different semantics, e.g., structural equations (Woodward 2003, Pearl 2009, Kistler 2013, Wilson 2015, Baron et al 2017), one that may better cohere with scientific uses of counterfactual reasoning.

Lastly, it is important to note that it is not simply the received semantics for counterfactuals that must be revised. As we saw at the beginning of the paper, there are *prima facie* plausible metaphysical, epistemological, and logical principles that support the orthodoxy (Williamson 2007, 2018, forthcoming) independent of any semantic model for counterfactuals. In the nascent literature on non-vacuous counterpossibles, it is counterfactual semantics that have enjoyed the lion's share of philosophical attention, and these other non-semantic motivations for the orthodoxy have thus far received nearly no discussion. I am sure they will raise interesting

[20] Thanks to Jessica Wilson for this suggestion.

problems all on their own. I regret that I cannot answer here any of these important issues, beyond simply providing new data on non-vacuous counterpossibles in science for future work to draw on.

Chapter 2: Interventions and Counternomic Reasoning

1 Introduction

"Counternomics" are counterfactuals whose antecedents run contrary to the laws of nature.[21] Counternomics appear in many areas of scientific discourse. Explaining why we know some theory in the history of science to be false frequently requires making counternomic claims. For example, we know classical electromagnetic theory to be false because if it had been true, then a sample of metal's rate of photoelectron emission would have varied as the wavelengths of light shining on it were altered; and photoelectron emission is unchanged under that intervention. That counterfactual statement – 'If classical electromagnetic theory had been true…' – is a counternomic.[22] Additionally, counternomics often appear in scientific explanations, especially those that invoke laws or lawlike regularities to do explanatory work. For instance, diamond's lack of free electrons explains why it is an exceptionally poor electrical conductor. Accordingly, one is inclined to affirm the following counterfactual: "If diamond's chemical structure had been such that it had more free electrons, then it would've been a better conductor." But because it is at least a matter of nomological necessity that diamond has its actual chemical structure, that too is a counternomic.

---

[21] Also called 'counterlegals'.

[22] One might wonder why I identify this as a subjunctive, not a strict or material conditional. One reason (beyond surface grammar) is that it doesn't behave logically like strict/material conditionals, e.g. strengthening the antecedent seems to be invalid (whereas it is valid for strict/material conditionals). Observe: "Had classical EM theory been true (but we *didn't* observe any changes in the photoelectric effect), then we would've observed such-and-such a change in the photoelectric effect" seems false, while the unstrengthened conditional is true.

Despite how common counternomics are in scientific discourse, surprisingly little philosophical attention has been paid to them.[23] This lack of attention is surprising because counternomics straightforwardly raise a number of interesting epistemological and metaphysical questions. For example, what counts as evidence regarding what would have happened in some nomologically impossible circumstances, and why? Which counterfactual semantics best models how scientists reason counternomically? Relatedly, since some counternomic antecedents seem to be logically or metaphysically impossible – e.g., the logically impossible, 'Had Bohr's theory of the atom been true, …'[24] – do our theories of modality and modal language somehow need to accomodate impossible worlds and counterpossible conditionals?

My aim in this paper is to get started answering some of these questions. I want to address a question concerning our knowledge of counternomics – in particular, about how observations of actual physical systems might support counternomic hypotheses. Again, counternomics describe circumstances in which the laws of nature are different from what they actually are. So we have limited epistemic access (to start, we have no empirical access) to any of the circumstances they describe. Yet in some cases, real-world observations seem to support our beliefs about what would have happened under those different laws. The question I'm concerned with is: How? How can the way things behave under the *actual* laws support claims about how they would have behaved under *non*-actual laws?

[23] Counternomics have mostly been discussed as a problem for necessitarianism about laws (Handfield 2004, Hendry and Rowbottom 2009, Lange 2004). They've also been mentioned regarding the laws' role in practice (Lange 2000, Roberts 2008: 6) and recent accounts of non-causal explanation (Lange 2016: 2, Jansson and Saatsi forthcoming).

[24] Bohr's theory of the atom is a canonical example of a logically inconsistent scientific theory.

I'll begin by narrowing the scope of the question to a particular kind of interesting experimental case and provide two representative examples (§2). Because these are cases of experimental manipulation, you might think that how we reason counternomically using them is a kind of intervention-based counterfactual inference. I show that this characterization faces serious difficulties (§3), and suggest another characterization (§4).

2 Narrowing the Question

This paper concerns the following question: how can facts about how things behave under the actual laws support claims about how they would have behaved under a *different* set of laws? I can put the question a more precise way. When we have some actual evidence E and some counternomic conditional C, what inference or evidential relationship moves us from E to C? I think it's still important to narrow the scope of *this* question, since at least some counternomic circumstances are such that our reasoning about them is fairly transparent. I will now distinguish between these "easy" cases and those that I will focus on.

Here is one example of our reasoning about a counternomic circumstance where the structure and content of our reasoning using "real-world" data is easy to understand. It's widely known that certain mathematical models of motion in classical mechanics represent concrete objects as moving across frictionless planes. This is obviously an idealization. But this idealization relies in part on a counternomic: it's nomologically impossible that there *could have been* frictionless planes. Now, there's a whole range of nomologically possible planes ranging from very frictive (e.g. rubber) to not very frictive at all (e.g. wet ice). And it is easy to graph how the acceleration of some object differs between more frictive and less frictive surfaces, and as the coefficient of friction approaches zero. So how do we know how objects would have moved across a frictionless plane? Simply extend the observed curve to the point where the coefficient of friction is zero. However an object's acceleration is characterized there, that's how it would have behaved if there had been frictionless planes.

A second, similar kind of case where the content and structure of our counternomic reasoning is transparent involves objects whose behavior is governed by well-known mathematical laws. For example, while Newton's law of gravitation is an inverse square law, it's possible to construct simulations of how planetary interactions would have looked if gravitation had been an an inverse cube law. One need only formulate the non-actual law mathematically, and make the behavior of the simulated objects fit it. Thus it is easy to know how things would have behaved in that counternomic circumstance. The reasoning that leads us to these bits of counternomic knowledge is, as above, quite transparent.

The two types of cases I've just mentioned share the following feature: they involve an extrapolation of mathematical descriptions of physical systems. The sort of cases I am concerned with are those where our counternomic reasoning is based not on mathematical descriptions, but instead on observations of experimental systems. These cases are puzzling for a number of reasons. Let me briefly describe their puzzling features before providing examples. These cases comprise observations of experimental settings in which an ordinarily law-governed phenomenon is "absent". Accordingly, those observations can and do support counternomic claims about what the world would have been like, if some law had not obtained, or if some kind of entity had not been excluded by the laws. But in each of these experimental settings, the state of the experimental system is significantly distorted from the sorts of conditions described in the targeted counternomic circumstances. The presence of the distortion in these cases makes it puzzling exactly how these experimental observations manage to provide evidence for the counternomic claims of interest.

2.1 Water and hydrogen bonding

To understand my first example, some background will help. Water has an abnormally high boiling point for a molecule of its mass. Other molecules of water's approximate mass are gaseous at the temperature ranges at which water is a liquid. Based only on its mass, one expects that water would be a gas at room temperature. The boiling points of other group 16 hydrides (e.g. H_2S, H_2Se) obey a nearly-linear relationship with regard to their molecular mass:

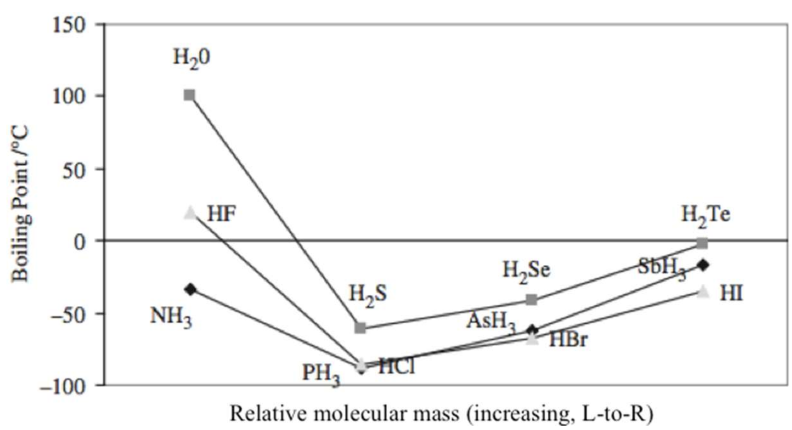

Figure 1. Boiling points of various hydrides (Hendry and Rowbottom 2009).

Extrapolating from this relationship, one expects that water's boiling point will be at about -70 °C.[25] Obviously, it is not. What explains water's abnormally high boiling point is its tendency to

[25] HF and NH_3 also form intermolecular hydrogen bond networks; note the similar spike in their boiling points.

form intermolecular hydrogen bond networks.[26] Hydrogen bonds are much stronger than the weaker van der Waal bonds the other group 16 hydrides form; as a result, water's boiling point is much higher than expected. Because of hydrogen bonding's role in explaining why water's boiling point is so high, one is inclined to affirm the following counternomic:

H-BONDS If there had been no intermolecular hydrogen bonding, then water would have been a gas at room temperature.

This theoretical background provides ample evidence for H-BONDS. And there is no mystery as to how *this* evidential relationship works; indeed, finding H-BONDS by extending the almost-linear relationship between molecular mass and boiling point mirrors the sort of case I set aside as "easy". Thus consider, as promised, additional, experiment-based evidence for H-BONDS.

When one introduces a large amount of energy into a sample of liquid water (e.g., by heating it), one can observe its intermolecular hydrogen bonds breaking.[27] Mass in the liquid sample is lost as the bonds break and portions of liquid evaporate. Suppose that one observes all this and, quite reasonably, says, "Here's a situation in which we know that the water's hydrogen bonds aren't there. And look, it's a gas! Shouldn't this count as some evidence for what would happen if there had been no hydrogen bonding?" Indeed, this observation *does* allow us to reason veridically about what water would've been like, had hydrogen bonding not occurred. It is an additional evidential basis for H-BONDS.

[26] In fact, hydrogen bonding explains other of water's peculiar properties, including its high surface tension and the fact that its liquid phase is denser than its solid phase; see Chaplin (2007).
[27] See Steinel et al. (2004) for how, using high-speed spectroscopy, we can observe hydrogen bonds breaking when energy is introduced via vibrational relaxation.

The puzzle I'm interested in is understanding how this observation serves in that evidential capacity. This case is puzzling for two reasons. First, even though water's hydrogen bonds break when it is boiled, it's nevertheless still a law that intermolecular hydrogen bonding occurs. All you've done is break some hydrogen bonds by boiling the water; surely even that bond-breaking is governed by the laws. So, to say that observations of boiling water show us what things would have been like if there had been no hydrogen bonds seems to mischaracterize what the observation actually is.

The fact that observing this situation in which—here's an infelicitous description—hydrogen bonding "doesn't occur" requires us to first boil the water we're observing poses a second, more worrying problem. For the thing that we took ourselves to be reasoning about is what water would have been like *at room temperature* in the absence of hydrogen bonding. Of course, heating some water means it's no longer at room temperature. So, on the face of things, that observation shouldn't allow us to reason about the very thing that we wanted to reason about.

2.2 Magnetic monopoles

Here is my second example. Maxwell's second equation – a law – ensures that there cannot be any magnetic monopoles, i.e. a magnet with only a south pole or a north pole. Nevertheless, the following counternomic seems true:

MONOPOLE If there had been magnetic monopoles (i.e. had Maxwell's second equation not held), then they would still have obeyed Maxwell's first and fourth equations.

Maxwell's "first" and "fourth" equations are Gauss's law and Ampère's circuital law; together they govern the fact that electric charge, flux, current strength, and the magnitude of fields always stand in particular proportions to each other. So, qualitatively, MONOPOLE says that if there had been more kinds of magnetic objects (e.g. monopoles), they would have behaved in basically the same ways that magnetic objects already behave. Put this way, this should seem plausible. Quantitatively, the only change to Maxwell's equations (i.e. other than the second equation itself; from $\nabla \cdot \mathbf{B} = 0$ to $\nabla \cdot \mathbf{B} = \rho_{\mathrm{m}}$) would be to the Maxwell-Faraday equation (from $-\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t}$ to $-\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t} + \mathbf{j}_{\boldsymbol{m}}$). This change adds a variable $\mathbf{j}_{\mathrm{m}}$ for the contribution of the monopole's magnetic charge density, so that magnetic and electric monopoles end up being analogous.

Some recent experimental results happen to confirm that if monopoles had existed, they would have behaved in this way. In other words, these results confirm MONOPOLE. Now, there are no genuine, naturally-occurring magnetic monopoles. Any genuine monopoles would be elementary particles that carry a net magnetic charge; there are none. But in experimental settings, materials known as "spin ices" behave, when supercooled (<1K), as if they contained magnetic monopoles. In these low-energy systems, coordinated groups of particles will jointly—i.e., as "quasiparticles"—carry a net magnetic charge. And while the magnetic charge in these cases is not attached to a single particle, it behaves mathematically as one would expect the magnetic charge of genuine monopoles to behave (Castelnovo et al. 2008, Bramwell et al. 2009).

Spin ices are solids composed of tetrahedra arranged in a “corner-touching” lattice (tetrahedron vertices touching other tetrahedron vertices), with electrons sitting at the vertices of each tetrahedron. Each electron’s spin is constrained to point either directly into or out of its tetrahedron. When supercooled, the spins tend to align one spin in for every one spin out; this is the system’s lowest-energy configuration. However, due to thermal fluctuations, some individual spins flip, so that some regions end up containing net magnetic charge. A simplified model for understanding this is presented below:

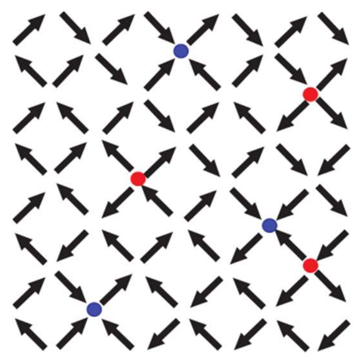

Figure 2. A two-dimensional representation of monopoles in spin ice (Aldus et al. 2013).

In this figure, each arrow represents a spin; red/blue dots are the centers of sites of magnetic monopole quasiparticle behavior. Notice that at each normal site (without a dot), the spins are aligned “two in, two out”; but at the monopole sites, one spin is flipped so as to induce a region of net magnetic charge.

What I'm interested in is how these results support MONOPOLE. To restate it as before: what inference or evidential relationship moves us from these experimental observations to MONOPOLE ('Had there been magnetic monopoles, then…')? In this example, just like the first example, the values of other variables in the experimental system are distorted from normal conditions in important ways. The extremely low energy of these condensed matter systems is a distorted variable, together with the fact that the "monopoles" in question are not genuine monopoles. They are quasiparticles. Yet MONOPOLE presumably describes circumstances in which there are genuine, fundamental-particle-monopoles. How, then, do these experimental observations manage to provide evidence for it?

3 Intervention-based Counterfactual Reasoning?

Both of these examples involve experimental manipulation of some sort or another and an observation of what happens under that manipulation. Accordingly, you might suspect that the most straightforward characterization of how these cases support some counternomic conditional is as a kind of counterfactual reasoning from intervention.

Interventions on the variables in a physical system often help us differentiate between various causal structures we might think the system has.[28] If you manipulate whether some event A occurs in a system, and nothing whatsoever happens to B, then you ought not to think that A causes B. But if you manipulate A, and B changes the same way every time you manipulate A, then you ought to conclude that A causes B. What's relevant for current purposes is that

[28] Much recent work in philosophy of science has focused on intervention-based causal and counterfactual reasoning, largely spurred by Woodward (2003) and Woodward and Hitchcock (2003).

interventions support *counterfactual* reasoning in the same way they can support causal reasoning. If one observes the occurrence of both A and B in some physical system, then interventions can help you reason counterfactually about what would happen if either one had been absent. Suppose you know that A causes B because of interventions on A. Then, you should infer on the basis of those interventions that if A had been absent, B would also have been.

The current suggestion is that the structure of the counternomic reasoning in my problem cases is the same kind of counterfactual reasoning that interventions normally facilitate. The thought seems simple enough: one observes that net magnetic charge is always zero in unmanipulated electromagnetic systems, and that Gauss's and Ampère's law hold. Then, thanks to an intervention that results in some net magnetic charge in an electromagnetic system, one learns that monopole charge makes no difference to whether either of those laws hold. So you reason counterfactually and conclude that MONOPOLE is true.

This suggestion is clearly quite tempting. But I don't think it succeeds. The reason stems from something I've already mentioned. It's not possible to intervene on the variables whose absence comprises the counternomic circumstances of interest without distorting other important variables: one can't manipulate whether hydrogen bonding occurs, or whether there's net magnetic charge in a region without interfering with some other behavior in the experimental system. Our counternomic reasoning in the cases I'm interested in can't be characterized as intervention-based counterfactual reasoning simply because their manipulations don't qualify as the sorts of manipulations that can properly facilitate counterfactual reasoning.

To see why I say this, notice that when one uses interventions to reason counterfactually, it is important that all features of a system other than those that are manipulated (and those that the manipulated variables cause) be kept intact. Indeed, the basic idea behind an intervention is that one manipulates *only* the variable whose causal and counterfactual relationships are being investigated. Otherwise, interventions would be of little use in causal and counterfactual reasoning. Suppose you're evaluating whether raising the voltage of a circuit causes a lightbulb plugged into it to burn out faster. In order to properly reason about that counterfactual circumstance, you need to hold fixed *everything* except for the voltage of the circuit. You especially don't want to change anything else that might affect the bulb's burn rate (e.g., filament thickness, gas pressure). The reason is simply that if the relationship you're investigating is actually there, you want the intervention to help you isolate it. Thus in order for a manipulation of a target system to properly qualify as an intervention, it had better isolate only the variables of interest.[29]

With this in mind, consider the following case. Consider some ecosystem in which a species R is subject to predation by species D, and suppose that R has very few natural predators of their own. Imagine you're interested in evaluating the following counterfactual: 'If D's rate of predation were lower, then R's reproduction rate would increase by *n* individuals/year.' Now suppose it's possible to manipulate this ecosystem by introducing a species H that preys only on D but not R. However, even though H doesn't prey on R, H competes for the same resources that R does, peaking during R's mating period. Now, even though introducing Hs is a manipulation of that system that lowers D's rate of predation, I don't think it allows you to properly evaluate

[29] Cf. Woodward (2003, pp46—48).

that counterfactual. The reason should be pretty clear. Bringing in Hs distorts the very features of the ecosystem required for evaluating whether that counterfactual is true, since bringing in Hs interferes with the Rs independently of how it affects the Ds. Accordingly, that "intervention" (so-called) can't be an appropriate basis for evaluating that counterfactual.

The same problem affects the experimental-to-counternomic cases. The targeted circumstances in H-BONDS and MONOPOLE implicitly or explicitly require specific background conditions. H-BONDS describes temperatures as normal, MONOPOLE describes genuine monopoles. However, the only manipulations available to us directly affect the experimental systems in *exactly* those required respects. That is, those manipulations distort the very background conditions required to evaluate those counterfactuals. So, just as introducing Hs can't be a basis for intervention-based counterfactual reasoning about R's reproduction rate, neither can the experimental observations of interest by themselves provide evidence for their counternomic conditionals. They don't properly qualify as interventions (in the technical sense).[30] So our reasoning about H-BONDS and MONOPOLE, from the experimental results I've mentioned, can't be characterized as intervention-based counterfactual reasoning.

4 Counterfactuals and Inductive Inferences

Putting aside the obstacles faced by the interventionist characterization, the central philosophical issue raised by these examples remains the same as before. How can observations

[30] Op. cit.

of how things behave under the actual laws serve as evidence for claims about how things would have behaved under a different set of laws incompatible with the actual ones?

One of the cases I described as "easy" will prove to be helpful here. Remember that one can use a graph that tracks an object's deceleration on smoother and smoother surfaces as evidence for a counternomic claim about how it would behave if it had been sliding on a frictionless plane: simply extend the relationship to the point of frictionlessness. We can easily suppose that such a graph is based on data observed in actual cases (for example: take a piece of plastic, slide it on some wet ice; sand it down a little smoother; do it again and again) and hence taken from physically realizable systems. But frictionless planes are not physically realizable. Friction arises in part from electromagnetic attraction and Pauli exclusion at the microscopic level, so the familiar idealization of objects and planes as frictionless is counternomic with respect to at least those law-governed phenomena. So, despite how easy it is to understand the structure of our counternomic reasoning in this case, it resembles the experimental cases in this respect: the evidence used to support a claim about the counternomic is based in observations of actual, physically realizable systems. Then, we can ask the same question of this case that's asked of the experimental cases: how can observations of physically realizable systems, governed by the actual laws, support claims about the behavior of systems governed by non-actual laws?

The question is, of course, relatively easy to answer with this case. But a careful look at how things work here can shed some light on how the evidential relation in the experimental cases works. If analogous features are present in the experimental cases, then there will be a way of understanding how they work, too.

The evidential relationship in the case of the frictionless planes is an inductive inference from observed cases to counterfactual frictionless cases. In that case, it's an extrapolation of a mathematically-described relationship, so the induction is quite transparent. Yet it's important to note that inductive inferences of that variety sometimes don't work. For example, one can't indefinitely extend the relationship that tracks the amount of energy needed to accelerate an object to some speed, since the speed of light imposes a hard threshold on reachable speeds. Nevertheless, the induction regarding frictionlessness is easy to accept because nothing suggests that there is a similar hard threshold. Nothing suggests that the otherwise smooth mathematical curve would spike or drop very suddenly after some point; and nothing in our physical theory suggests that things would behave unpredictably if they were to become smoother and smoother. In other words, there's no reason to doubt that the unobserved, counterfactual frictionless planes would have resembled the actual, even though that counterfactual turns out to be counternomic. I take it that's why an induction from the observed relationship to the counternomic frictionless cases seems unproblematic, despite its using behaviors governed by the actual laws as evidence for a claim about behaviors under *non*-actual laws. This point is important. It means that if you're using some observations of actual physical systems to help support a counternomic, the fact that the target describes a counternomic situation doesn't *by itself* provide a defeater for the inference or evidential relationship.

I take it that the problematic experimental cases also involve an inductive inference. They, too, use observations as evidence for claims about some unobserved, *viz.* counterfactual, cases. To be sure, their inductive evidence is of a different sort than the induction regarding frictionless planes; there is no mathematical relationship that they can extrapolate from. Nevertheless, it seems just as before that there are no defeaters for thinking that the counterfactual would have resembled the actual. So *prima facie* those experimental observations can support inductive inferences about their counternomic analogues.

Let's begin by considering the monopole quasiparticles. They carry a net magnetic charge, which still obeys Gauss's and Ampère's laws, and which behaves as if it had been carried by a single object. Those observations are used to support a claim about how a genuine monopole, i.e. a fundamental particle, would have behaved. What would be a defeater against using this observed behavior as evidence for a counterfactual about the behavior of genuine monopoles? The most straightforward defeater would be an indication that electromagnetism at non-fundamental scales (including quasiparticles) fails to resemble the electromagnetic behavior of fundamental particles. As far as I'm aware, we have no such defeater. In fact, electromagnetic behavior *does* scale above the fundamental; and we antecedently take macroscopic electromagnetic behavior to resemble the microscopic. So in fact there is some (undefeated) reason to take the electromagnetic behavior of a quasiparticle as inductive evidence for what a fundamental particle with analogous properties would have looked like.

Now, you might note here that the resemblance to the fundamental particles is exactly what's at issue: the fundamental particle analogues of the quasiparticles would have been genuine monopoles, which are *incompatible* with the actual laws. But it is unclear that this is a defeater all by itself. Nothing in electromagnetic theory predicts that if monopoles had existed, they would have behaved unusually or unpredictably in ways that can't be approximated by what is actually physically realizable, i.e., ways unusual beyond simply being monopoles. (Compare: special relativity predicts that if there had been lightspeed-exceeders, they would have behaved in radically different ways from what's physically realizable, e.g. they'd have infinite mass). If so, then the known fact that monopoles are incompatible with the laws doesn't *by itself* provide a defeater against the inductive inference. Thus the presumption that the counterfactual would have resembled the actual seems to be undefeated, even though in this case the counterfactual is also counternomic. In other words, an inductive inference that the behavior of genuine monopoles would resemble the behavior of observed monopole quasiparticles seems undefeated, even if one knows that genuine monopoles aren't physically realizable.

So there is a *prima facie* admissible inductive inference from the observations of monopole quasiparticles to MONOPOLE. I think a similar story can be told for H-BONDS. We observe that water's hydrogen bond networks break when it is boiled, and simultaneously observe that it's a gas. Is there a defeater for believing that things would have resembled the actual (e.g., actual water without hydrogen bonds is gaseous) in the counterfactual absence of hydrogen bonding? Nothing that I'm aware of suggests that other kinds of intermolecular bonding would have behaved radically differently. And nothing suggests that, e.g., how gases react to temperature would have been different, either. The judgment that the counterfactual would resemble the actual again seems to face no defeaters, even though the counterfactual supposition here—"Had there been no hydrogen bonding…"—is also counternomic. If so, there is a *prima facie* undefeated inductive inference that uses those observations, governed by the actual laws, to support H-BONDS, a counterfactual about non-actual laws.

Tentatively, then, it seems plausible to characterize our counternomic reasoning in my two experimental cases as a certain kind of defeasible inductive inference from the observed to the counterfactual. I've just briefly argued that in these two cases, there is no reason to *doubt* that such an induction is justified.

In at least the quasiparticle case, there may be additional, positive reasons for believing that the counterfactual would've resembled the actual. I can introduce the idea by calling on the "easy" case again. Idealizations like frictionlessness are frequently indispensable because they make calculations computationally tractable. That familiar idealization obviously produces results within close approximations of objects' observed behaviors. It seems like its success can support an inductive inference that those counternomic objects, i.e., frictionless planes, would have appropriately resembled actual objects. For the fact that that idealization yields approximately the right results seems to suggest that if frictionless things really had existed, their behavior would not have diverged significantly from the patterns established by the frictive stuff we actually observe.[31] If so, then given the success of the idealization, we have some reason to believe: if there had been frictionless planes, they would have behaved as expected, i.e. approximately similarly to the objects we actually have. In other words, the success of the idealization seems to *prima facie* justify using actual observations as an approximation of, or as inductive evidence for the behavior described in the counternomic.

Quasiparticles also feature in a frequent idealization. In condensed matter physics, treating coordinated groups of particles ("quasiparticles") as if they were unitary makes tractable many calculations that would otherwise be near-impossible. Now, following my preceding remarks, it seems like the success of an idealization can support an evidential relationship between observed cases and claims about the ideal or counternomic. So consider how the observations of magnetic

---

[31] This view is familiar in the literature on idealization. Ernest Adams, for instance, says: "The orbits of the planets are often calculated on the counterfactual assumption that these bodies are mass points between which forces of gravitation act. It is plausible that a counterfactual conditional is implicit in this reasoning because it might be explained as: 'The orbits of the planets are similar to those of mass points and if they were mass points then their orbits would be such-and-such; therefore their orbits are similar to such-and-such.'" (1993, p5)

monopole quasiparticles seem to support MONOPOLE: how does it work there? Well, idealizations of quasiparticles as single objects are highly successful in various computational contexts. We might reasonably infer, then, that if those objects had existed, their behavior would have resembled or approximated the behavior of the quasiparticles. The same principle should apply when it comes to experimentally observed quasiparticles, even though they're not being idealized for computational purposes. When one observes a quasiparticle, one can infer that if that quasiparticle had genuinely been a single object, then that object's behavior would have closely resembled the observed quasiparticle's behavior. And if so, the observed behavior of the experimentally observed objects—i.e. the monopole quasiparticles—can serve as *prima facie* evidence for a claim about how genuine monopoles would have behaved. So the observations of monopole quasiparticles can serve as *prima facie* evidence for MONOPOLE.

To conclude, reasoning about the counternomic using behaviors governed by the actual laws seem justifiable if there are no defeaters for thinking that the counterfactual would have resembled the actual. I've argued that the fact that the counterfactual happens to be counternomic does not by itself provide a defeater for such an inference. At any rate, these experimental cases are only two isolated examples of counternomic reasoning in scientific practice. I am sure that many other examples of counternomics in science can equally benefit from careful philosophical discussion.

Chapter 3: Dispositionalism and the Grounds of Counternomic Conditionals

1 Introduction

Dispositionalism about modality (or dispositionalism for short, throughout this paper) is the view that the *de re* modalities and the natural modalities alike are grounded in the world's dispositional properties. As a rough first pass, we might give the following sort of example: it seems metaphysically possible for me to be a bus driver. What is it that grounds this *de re* modal fact about me? According to dispositionalists, it is that I instantiate a number of dispositional properties (biological, psychological, and so on) that might be summarized as follows: I have the property of *being such that if I were to train to be a bus driver, I would be a bus driver*. Dispositionalism is, importantly, alleged to be able to ground the natural modalities as well. For the scientific world is replete with dispositional properties and laws of nature that seem to describe what behaviors various sorts of objects are disposed to engage in. For instance, the property of *being charged* seems clearly to be dispositional, since it seems that what it is for an object to have that property is to be such that if another charge is nearby, the object experiences an acceleration. This dispositional property—that is, *being charged*—thus seems apt not only to explain why there are counterfactual facts regarding the behavior of charged objects, but also, importantly, why there is a law of nature (Coulomb's Law) that describes that behavior.

In this paper, I argue that dispositionalism is false. For there are a class of *de re* modal, natural, and counterfactual facts whose reality which dispositionalism in principle lacks the resources to countenance. These are the class of *counternomic* facts: those that describe what things would have been like if certain laws of nature had been different. (For example, if gravitation had been an inverse-cube law instead of an inverse square law, it still would have conserved energy.) I can summarize my argument as follows. Dispositionalists hold that the laws of nature and counterfactuals alike are grounded in essentially dispositional properties, so, every dispositional property is law-abiding; but if every disposition is law-abiding, then there are no dispositions that ground (non-vacuously) true counternomics; the dispositionalist thus judges that there are none. Dispositionalism must be rejected on these grounds, since there *are* counternomic facts, and they form an important part of the scientific modalities.

The paper runs as follows. First, I introduce dispositionalism in more detail. Then, I show the conflict between that arises when dispositionalism faces counternomic discourse in science; and along the way I respond to possible recourses they may take. I conclude that none of these are satisfactory and thus that dispositionalism must still be rejected.

2 Dispositionalism Introduced

To understand what motivates dispositionalism, we must take a step back from the idea of grounding the modalities and begin with a discussion of the identity-conditions of dispositional properties. Consider, for instance, the property of *being charged*. There is an old thought—a Humean thought—that it is metaphysically possible for charged particles not to behave in the ways that charged particles actually do. For instance, so the thought goes, there are metaphysically possible worlds in which some particles genuinely instantiate the property *being charged* but, unlike charged particles in the actual world, just sit there when you stick another charge next to them.

There is an attractive sort of thought according to which some merely-possible property that behaves like this just would not be the very same property as charge. For consider how, following Kripke (1979), many metaphysicians tend to think about the individuation of the natural kinds. We can conceive of metaphysically possible worlds in which some very tiger-like animals, and some very gold-like substances, fail to be (let us say) mammalian and of atomic weight 79, respectively. If we put aside the issue of whether it would be consistent with the laws of nature for there to be non-mammalian tigers or non-79 gold, it is clear that the tigers and gold that exist in the actual world are of univocal types that "carve nature at the joints". If so, then clearly whatever tiger-like creatures or gold-like substances that world contains are not of the same sort that the kinds in our world clearly are. They are not genuinely tigers (e.g.), for 'tiger' picks out a class identical to the kind striped, furry, teethy mammals we have. And if tigers are identical to that sort of animal, then they *essentially* are that sort of animal; that is, it is metaphysically necessary that they are.

This Kripkean argument is well-rehearsed, but it is worth rehearsing it once again here because the idea can easily be extended to include properties (Shoemaker 1980). There are properties which have, in the actual world, certain causal profiles (i.e., dispositional profiles), and *being charged* is one of them. If there is a conceivable possible world that contains superficially charge-like properties, but they do not confer on their bearers the same causal roles that the substantially charge-like property of *being charged* in fact confers on its bearers, then it will turn out that those properties are simply not the same as *being charged*. In other words, so this thought goes, the property of *being charged* essentially confers on its bearers the causal roles that it in fact confers; the supposedly possible Humean worlds just contain a superficially similar, merely-possible substitute for charge. Put it in slogal form: charge is essentially dispositional, or alternatively, it has a dispositional essence (Shoemaker 1980; Mumford 2004; Bird 2005, 2007; Chakravartty 2007; Whittle 2008; Tugby 2013, 2014). And of course, the property of *being charged* is only one of many here; although there is some disagreement among dispositional essentialists (those who endorse this argument) as to how many properties we should countenance as essentially dispositional.

Dispositional essentialists end up endorsing a view of the world's actual ontology which has *de re* modality built into it. For as Armstrong (1996) notes, if there are genuinely essentially dispositional properties, they always "refer" to manifestations which need not occur. If a fragile glass does not break, the glass is still such that if it were to be suitably thrown, it *would* break: that is simply *what it is* for a glass to be fragile. Now suppose that you believe in dispositional essences; the world has *de re* modality built into it. Nature then contains hidden counterfactual facts that came prepackaged with it. Indeed, if nature contains essentially dispositional properties, it also contains *necessity* that is built in: it is metaphysically necessary how certain objects and properties behave.

If so, it is a natural thought that these "built in" modalities are just what ground *de re* modalities, counterfactuals, and the laws of nature. There is no need to countenance primitive possible worlds, as some have argued (Jacobs 2010). Why bring in extra ontological tools to do the grounding for the world's modal dimensions, when there already are modal facts built into your metaphysics, and they are of a familiar essentialist sort that many metaphysicians already find plausible? Thus these authors write:

- "[Dispositions] provide natural necessity and possibility and are fit to be the truthmakers of modal truths. They are not the truthmakers of all modal truths: only the natural or *de re* modal truths." (Mumford 2004, p170)
- "DE promises an account of natural modalities that is more 'local' or 'direct' than the Neo-Humean approach, enables explanations seemingly unavailable on the latter doctrine and is at the same time metaphysically more parsimonious than several other non-Humean views." (Jaag 2014, p2)
- "[Dispositionalism] is a thoroughly actualist view that locates modality in what we intuitively take to be the constituents of reality: individual objects and their properties." (Vetter 2016, p2684)

So dispositionalism about modality (again, 'dispositionalism' for short from here on out) has considerable attractiveness (see Borghini and Williams 2008; Martin 2007; Jacobs 2010; Vetter 2015, for defenses).[32] It promises to be naturalistic. There is a natural and intuitive sense in which *de re* modalities look like they can be grounded in intrinsic dispositional properties. And it promises to unify the otherwise-slippery notion of "natural" or "physical" modality with the more ordinary *de re* modalities. So, we can sum up the dispositionalist's core commitments: first, there are essentially dispositional properties. And these dispositional essences are what ground *de re* modalities, counterfactuals, and the laws of nature. All of these features of the world—the "modal package"—are due to the "built in", metaphysically necessary essences of dispositional properties.

2 Counternomics Introduced

[32] E.g., "A state of affairs S is possible iff there is some actual disposition d, the manifestation of which is (or includes) S," Borghini and Williams 2008, p26

In this section, I show how dispositionalism encounters serious difficulties in acccounting for the *full* modal package. Let me introduce the impetus for these difficulties. For the reasons we have just seen, dispositional essentialism and dispositionalism (about modality) turn out to be views on which the laws of nature obtain with metaphysical necessity. To see this, consider (e.g.) Coulomb's Law, which describes how charged particles interact. According to dispositional essentialists, the Law is grounded in the dispositional essence of *being charged*. That is, charged particles have a particular behavior—which ends up being described in Coulomb's Law—of metaphysical necessity. (In general, most dispositionalists hold that the "Law" is unreal as such, i.e., as an additional thing, it only describes the dispositional behavior of charged objects.[33]) If so, in order for it to be metaphysically possible for Coulomb's Law to be different, it must be metaphysically possible for the property of *being charged* to confer on its bearers different behaviors and causal roles than it actually confers. But it is not. So, the Law is metaphysically necessary.

If the laws of nature are metaphysically necessary, then any counterfactual whose antecedent is contrary to the laws of nature will have a metaphysically impossible antecedent, in other words, will count as a *counterpossible* conditional. Let us now introduce the class of counterfactuals known as counternomics: those with antecedents that violate the laws of nature. Counternomics, as I have argued in recent work (Tan 2017, 2019: **Chapters 1 and 2 here**), are a staple in many areas of science. They help us understand how the laws of nature are related to one another: for instance, if gravitation had been an inverse-cube law instead of an inverse-square law, it still

[33] Swoyer 1982, Ellis 2001, Mumford 2004, cf. Hildebrand 2014, although see Bird 2007: 9 for disagreement

would have conserved total energy (Lange 2000, 2009). It is a common exercise in introductory electromagnetism classes to identify which of Maxwell's equations, which state the laws for electromagnetism in classical contexts, are preserved under the counternomic supposition that magnetic monopoles exist.[34] Likewise, counternomics are a frequent sight in scientific explanations: we can some macroscopic properties are explained by appeal to the differences laws make, e.g., if there had been no intermolecular hydrogen bonding, water would have been a gas at room temperature (Hendry and Rowbottom 2009, Tan 2017). Finally, counternomics are even taken to be empirically confirmed by the behaviors of certain experimental systems, e.g., magnetic monopole quasiparticles in "spin ices".

If dispositionalism is true, then the laws are metaphysically necessary; and if the laws are metaphysically necessary, then every counternomic is a counterpossible. The orthodoxy on counterpossibles is to treat them as indiscriminately vacuously true, so if dispositionalists wish to follow the orthodoxy, they must do so too. However, I have argued that we have powerful reasons to treat some counterpossibles, and especially those that appear in science (many of which appear as counternomics), as non-vacuously true or false (Tan 2019). The considerations for counterpossible non-vacuity are drawn from the observation that all of these sorts of uses of counternomics, mentioned above, clearly correspond non-vacuously to physical relations and processes of physical reasoning.

So perhaps dispositionalists will want to agree that there are some non-vacuous counternomics, but instead find some sort of error theory that explains how they are no trouble for their theory.

---

[34] See, for example, Schwinger 1998: §3, #8 and Jackson 1975, pp254-256.

There have been two such strategies. First, Toby Handfield (2002) has argued that counternomics all implicitly contain an additional clause before the "real" counternomic antecedent, which states that, "If it is (or were) metaphysically possible for the laws of nature to be different…". On this strategy, a counternomic like:

> "If diamond had not been covalently bonded, then it would have been a better electrical conductor"

Instead is interpreted as,

> "If it is (or were) metaphysically possible for the laws of nature to be different, then, if diamond had not been covalently bonded, then it would have been a better electrical conductor."

This strategy helps in two ways. First, if we read the hidden clause as a non-counterfactual conditional, e.g., a strict or material conditional, then it might still turn out to be vacuously true, but this would not violate arguments about the non-vacuous truth of counterfactuals. Secondly, and more importantly, if the first counternomic really does contain this hidden clause, we can read the conditional overall in an epistemic tone. For dispositionalists can say: we are not really *certain* that the laws of nature are metaphysically necessary; our reasons for thinking so are fallible. So, for all we know, the laws of nature could turn out to be metaphysically contingent, and if they turned out to be (for all we know) then we might be inclined to accept this other counternomic fact. The conditional, read this way, has a distinctly epistemic tone. And hence it poses no worries for the dispositionalist's accounting of the modal package, even if its antecedent turns out to be metaphysically impossible. For if the counterfactual is an epistemic one, then it is not even a part of the modal package, properly speaking.

Barbara Vetter has also adopted, in broad strokes, a strategy to read non-vacuous counterpossibles as epistemic. She writes,

> "...the dispositionalist is not concerned with epistemic modality...Modality, to her, is after all about the modal properties of objects, and 'disposed to' is best understood as a predicate operator. Epistemic modality, if it is about (say) the compatibility of a proposition with our knowledge or evidence. [...] Non-vacuous counterpossibles are never circumstantial (but rather epistemic), and hence they are not within the scope of a properly understood dispositionalist semantics anyway." (2016, p2696-7)

Vetter concludes that dispositionalists can avoid the threat that non-vacuous counterpossibles face for their theory. And again, this is because the strategy is simply to avoid reading these counterfactuals as being about metaphysical modalities.

I do not think that either Handfield's or Vetter's strategies succeed. The reason is simply that it seems clear that some counternomics and counterpossibles describe non-epistemic modalities. To begin with, the implications of Handfield's strategy, to identify a "hidden" counterfactual antecedent, seem false. It simply seems not to be the case that with many counternomics, we hide away or attest to a hidden-away antecedent regarding the metaphysical possibility of differences in the laws. Inded, with many counternomics, e.g., the diamond one just mentioned, the counterfactual seems complete as stated, because what it targets is a known physical dependence relation.

Similarly, Vetter's strategy will not succeed, either. For even if some counterpossibles appear to have epistemic readings, that does not preclude them describing genuine facts about how the world is. To give an example, consider:

(B1) If Bohr's theory of the atom had been true, then an electron's angular momentum L in the ground state would have been observed at $L=\hbar$ (i.e. the Planck constant)

I have argued (2019, p48) that this is a counterpossible. But it must be non-vacuous. For unless (B1) managed to convey some actual information about what would result from Bohr's theory being true, we would not be able to use empirical observations, like the value of L (observed at 0), to rule out that theory's truth and thereby learn something substantive about the world. The real reading of (B1) must be "metaphysically" modal and non-vacuous. So, neither Handfield's nor Vetter's strategies for avoiding the threat of counternomics (or, on the dispositionalist's view, counterpossibles) succeed. It cannot be avoided that there are some non-vacuous counternomics.

3 No Dispositionalist Grounds for Counternomics

3.1 The Argument

The dispositionalist is forced to agree that there are non-vacuous counternomic conditionals. In this section, I argue that they can in principle provide no grounds for these non-vacuous counterfactuals. The upshot is, of course, that dispositionalism must be rejected.

To begin my argument, let us note that dispositionalists countenance a tight connection between which dispositional properties there are and which counterfactuals are true. Borghini and Williams write, "If anything, counterfactual conditionals will be true…in virtue of dispositions" (2008, p24). Vetter is also explicit about this, giving us two passes at how the connection between counterfactuals and dispositions might be accounted:

> ‘If *p* were the case, *q* would be the case’ is true iff all contextually relevant systems have a disposition whose stimulus consists in *p* and whose manifestation consists in *q*.

And

> ‘If *p* were the case, *q* would be the case’ is true iff all contextually relevant systems are such that if they have any disposition whose stimulus consists in *p*, then they have a disposition whose stimulus consists in *p* and whose manifestation consists in *q* (2016, p2683)

She notes that neither of these need to be the dispositionalist’s last word on the matter, since the broad interaction that there is between counterpossibles and dispositions does not concern minutiae in how the two are connected. There are non-vacuous counternomics. And as we can see, according to dispositionalism, any non-vacuous counterfactual is accounted for in terms of there being an underlying dispositional property. So, can the dispositionalist agree that there are corresponding dispositional properties?

No. The thought is pretty simple: dispositions necessitate the laws of nature, so all dispositions must be law-abiding. If all dispositions are law-abiding, then there are no counternomic dispositions. But let us be a little more careful here. Let’s first choose an example of a counternomic: if there were magnetic monopoles, then Ampère’s circuital law still would have held (Tan 2017). If this counternomic is true, then it must be that there is some physical system that has a disposition whose stimulus condition is to instantiate a net magnetic charge, and whose manifestation consists in the behaviors associated with Ampère’s circuital law (that a certain proportion of charge strength, density, and flux is maintained through all dynamic changes).

Here is an explicit way of putting the argument above, about the law-abidingness of dispositions:

(1) If there is a dispositional property, where object O bears the property, of *being such that if O has net magnetic charge, then it behaves as Ampère's circuital law will describe*, then it is a law of nature that 'All magnetic monopoles behave as Ampère's circuital law describes.'

(2) 'All magnetic monopoles behave as Ampère's circuital law describes' is not a law of nature.

(3) Therefore, there is no dispositional property of *being such that that if O has net magnetic charge, then it behaves as Ampère's circuital law will describe*.

(This argument is put as if it were the dispositionalist speaking.) Here there is nothing special about Ampère's circuital law, or electromagnetic properties in particular. We could run this same argument about the counternomic that if gravitation had been an inverse-cube law, it would have still conserved energy: there is no corresponding law of nature, so that is empirical evidence (indeed, it is *proof*, by the dispositionalist's lights) that there is no underlying disposition. This argument helps us see exactly why the dispositionalist should reject law-violating dispositions, i.e., those with nomologically impossible stimuli and manifestations. According to their view, the laws of nature are determined entirely by which dispositions there are, so *of course* there only ever are "law-abiding" dispositions: it is ultimately not the dispositions that abide by the laws, but rather the laws that abide by the dispositions. And here is a datum: there is no law that describes the behavior of magnetic monopoles. That example law, given above, is totally vacuous; there is nothing for it to describe. Now, you might object and say that there are after all vacuous laws of nature, for example, idealized laws. But appeal to these will not help to defend the lawhood of 'All magnetic monopoles…'. For vacuous laws earn their lawhood by approximately describing, or by being simpler descriptions of, behaviors that there actually are:

the ideal gas law earns its keep because, though vacuous, it is *approximately* non-vacuous. There is simply no physical objects or behaviors for 'All magnetic monopoles…' to even approximate. It is a datum that there is no such law. So, by the lights of the dispositionalist, we thus have *proof* that there is no underlying, law-violating disposition. And if not, there is nothing that can render the corresponding counternomic conditional non-vacuous. Sayonara, dispositionalism!

3.2 Possible Dispositionalist Recourses

Perhaps you think there are other ways of somehow getting to the point where the dispositionalist agrees that there are nomologically impossible dispositional properties. But the fact that it is the laws that abide by the dispositions, rather than the other way around, explains why the dispositionalist cannot in principle countenance law-violating dispositional properties. To consider just one strategy, Daniel Nolan and C. S. Jenkins (2012) suggest that cases of experimental measurement setups (e.g., the apparatus used to detect the unit charge in Millikan's oil drop experiment) as well as idealized scientific models give us evidence that there are unmanifested—indeed, unmanifest*able*—counternomic dispositions.

Note that Nolan and Jenkins's arguments leave a lacuna open through which dedicated dispositionalists may escape.[35] Suppose I think that the experimental apparatus simply is disposed to detect the value of the unit charge; and of course it is such that *on the condition* that the unit charge is different, then, if it were stimulated in the right way, it would have detected that different value. It is no problem to accommodate this whole conditional if you are a

[35] For a different strategy, which rejects wholesale the non-vacuity of counterpossibles, see Vetter (2016).

dispositionalist, since the antecedent materially implies the embedded conditional in the consequent. In fact, Nolan and Jenkins themselves acknowledge this lacuna (p747-8), and do not suggest much by way to defeat it. Indeed, they write,

> "*If* we had independent reasons for thinking that there are no unmanifestable dispositions, maybe the conditional substitutes would be worth considering as a way to salvage something of the underlying intuitions withoutdoing (what we consider to be) full justice to them…for those whoo think they have such reasons, we recommend this as a strategy to pursue to explain the plausibility of objects having unmanifestable dispositions" (p748)

And dispositionalists about modality are clearly one party who take themselves to have good independent reasons for thinking that there are no counternomic dispositions. I have provided an additional argument that shows why dispositionalists must reject the reality of counternomic dispositions. So they clearly will not be persuaded by Nolan and Jenkins's arguments, especially given that those authors have taken the burden of finding an escape route.

Another strategy by which you might think the dispositionalist can find counternomic, unmanifestable dispositions in their ontology is by appeal to alien properties. An alien property is simply a merely-possible property, i.e., one that does not exist in the actual world at all. So the thought goes, if there are merely-possible dispositions to behave in particular ways if you instantiate a net magnetic charge, then *those* merely-possible dispositions can ground counternomic truths. So they could. But the dispositionalist cannot countenance their reality. For dispositionalists, remember, what is possible is determined by which dispositions there actually are. So, if there are merely-possible, "alien" counternomic dispositions, then there be some grounding for that possibility in the stimulus-conditions of an actual-world disposition.

But if there were, then the laws of nature would be different than the laws actually are. Dispositionalists must agree (as, e.g., Contessa 2010) does, that there are no alien properties. All of the properties there possibly could be are exhausted by the ones there actually are.

Now, a nearby version of this argument might succeed. As Matthew Tugby (2013, 2014, 2015) points out, if dispositionalists endorse Platonism, then they can agree that there are some uninstantiated properties. Maybe these Platonic, uninstantiated dispositions are not only unmanifested, but some of them are counternomic. This would be a surefire way of getting counternomic dispositions into your ontology; and it indeed would succeed at providing a grounds for counternomic truths. However, while I personally have no truck with this strategy, I do not think that dispositionalists will find it appealing.

If properties are Platonic, then they are abstract objects which exist necessarily. If so, there are *de re* modal facts about them that are grounded still in facts about dispositional properties, but not about the material realizations of any dispositional properties. This seems to undermine one of the core attractions of dispositionalism, which—as Vetter puts it—is that

> "[Dispositionalism] is a thoroughly actualist view that locates modality in what we intuitively take to be the constituents of reality: individual objects and their properties." (Vetter 2016, p2684)

Vetter here does not specify that that the constituents of reality must be concrete. But it is worth asking: what do we "intuitively" take to be the constituents of reality? An ontology that includes Platonic universals seems rather unparsimonious, and also one in which the modal facts are out of epistemic reach. And as Vetter (2015: Ch.1) points out, a core attraction of dispositionalism is that it renders the epistemology of *de re* modalities non-mysterious. How do we know what is possible and necessary? By way of the dispositions! (Remember that the dispositional essences obtain with metaphysical necessity.) Endorsing Platonism means undermining one of the selling points of dispositionalism: we get to ground the modalities in the stuff that we can see *around us.*

There is one last sort of strategy that the dispositionalist might take. Robin Hendry and Darrell Rowbottom suggest the following view as an amendment to dispositional essentialism, when faced with the exact sort of counternomic-based challenge I have raised throughout this paper:

> "Our move is to suggest that having a property *P* may involve manifesting *M* in response to *S* in some contexts, but manifesting M_1 in response to *S* in other contexts, where such contexts are possible worlds. Alternatively, we may explain the position by saying that to have a property is to have a single set of actual and possible (but non-actual) dispositions, rather than a set of actual dispositions only…the *actual* dispositions associated with any given property may not be the only dispositions associated with that property" (2009, p673-4)

This strategy builds in the extra modality required for counternomic truthmaking into the profile of the property itself: it is possible for there to be (minor) changes of this-or-that sort to the causal profile of a dispositional property. One first objection to this strategy is that it will not generate a response to all kinds of counternomics. For something like, e.g., "If there were no hydrogen bonding…' (their example) it will work, and it will also work for something like, "If the electromagnetic constant were 4% weaker…". These involve changes to properties that there already are. However, counternomics like, "If there had been an entirely new kind of force and interaction, but which were still time-translation invariant, then it still would have conserved energy" concern totally uninstantiated properties. And as such, there is nowhere in the causal profile of what already exists to include these; a better strategy would be the Platonic one we have already discussed above.

A deeper objection to this view of Hendry and Rowbottom's, however, is that it seems to undermine dispositionalism. For their strategy consists in making a *de re* modal claim about the causal profiles of essentially dispositional properties. Following the general strategy that dispositionalism about modality prescribes, these *de re* modal claims—e.g., property *P* could have survived this-or-that sort of change to its causal profile, but not this other change—must be grounded in dispositional properties. But because the *de re* modal facts here ascribe modal properties to other ordinary properties, any disposition that grounds them would have to be a dispositional property that is instantiated *by a dispositional property*. It must be second-order properties that ground them.

Three problems arise here. First is that this suggestion of Hendry and Rowbottom's begins to seem *ad hoc*. Presumably, the dispositional essentialist was perfectly happy to say that the modal profiles of dispositional properties are of the essential sort *until* this challenge about counternomics appears, at which point we are impelled to believe that properties like the electromagnetic constant have more robust modal profiles. But what about another dispositional property, like *being squeaky*? I can see no independent reason to accept a fully general version of Hendry and Rowbottom's proposal, and it seems that the only reason to accept it at all must be targeted at specifically scientific properties that are apt feature in counternomics. This gives the entire proposal the feeling of being *ad hoc*, and it seems worth rejecting on these grounds. Secondly, even if we are inclined to take the suggestion at hand and believe that second-order dispositional properties ground the possibility of slight differences in the causal profile of, e.g., *being charged*, we run into the same epistemic problem that was raised about Platonism. These second-order properties are not *materially* manifested. And indeed, the only reasons we have to believe in them start to look *ad hoc*. So there end up being *de re* possibilities that are grounded not in what we intuitively take to be the constituents of reality, instead, there are *de re* possibilities grounded in totally epistemically inaccessible properties.

Third, and lastly, if there are second-order dispositional properties that ground the possibility of slight differences in ordinary dispositional properties, what is to stop there from being *third*-order dispositional properties, that give rise to possible differences in the second-order dispositions? There seems to be nothing in principle that theoretically prohibits this. And if so, it seems open theoretically to say that there is an infinite regress of higher-order dispositions instantiated by *being charged*, which eventually make it such that it is metaphysically possible for an object to be charged but just sit still when another charge is put next to it. To summarize all of this, the suggestion that it is metaphysically possible for the causal profile of *being charged* to change a bit must be, for the dispositionalist, rooted in higher-order dispositions. But if there are higher-order dispositional properties, it is hard to see what could rule out there being higher-order dispositions that make the original property, i.e. *being charged*, not essentially dispositional at all. So the suggestion that there are possible differences in dispositional profile ultimately undermines dispositional essentialism, which was the underlying motivation for dispositionalism about modality. It must therefore be rejected.

We have seen that there are a number of ways that the dispositionalist might try to include counternomic dispositions in her ontology. But none of them succeed. Dispositional essentialism implies that there cannot be counternomic dispositions; so the dispositionalist about modality must agree that there are none. But by the lights of dispositionalism, if there are no counternomic dispositions, then there is nothing to render counternomics non-vacuous. And counternomics, as we have seen, *are* non-vacuous. So dispositionalism lacks the resources to deliver on its promise of grounding the modal package in the dispositions. There is an important part of the modal package which, in principle, it must leave out.

Chapter 4: Ideal Laws, Counterfactual Preservation, and the Analyses of Lawhood

1 Introduction

Consider three observations about the laws of nature. First, many statements of natural law are idealized. Newton's law of universal gravitation, for instance, represents the relata of the gravitational interaction as point masses subject to no other forces except the gravitational force exerted by one another. The ideal gas law represents gas particles as perfectly hard point masses which are not subject to electromagnetic interactions and which transfer kinetic energy perfectly elastically. The formula for a pendulum's period both takes the pendulum's string to be massless and the entire mechanical system to be two-dimensional. The list goes on and on. The second observation is related to the first. Because of their idealized content, ideal laws are vacuous. The objects, properties, and relations they represent the world as containing do not in fact—cannot possibly—obtain. (But they are laws nonetheless.) Consider the third and final observation, which is a general hallmark of the laws of nature, including—*a fortiori*—the ideal laws. The laws of nature are *counterfactually preserved.* They would have continued to hold under an incredibly wide variety of counterfactual states of affairs; indeed, each law of nature would have held under any counterfactual supposition logically consistent with it. The laws are, after all, not mere physical accidents. And they enjoy counterfactual robustness in ways that no mere accident does.

These observations are quite familiar. But I shall show that reflection on them yields surprising conclusions for contemporary theories of the laws of nature. For these observations provide us arguments against some of the most popular contemporary views of lawhood. I will be concerned with, on the one hand, contemporary Humeans (e.g., Lewis 1973, 1983) about the laws of nature, who maintain that lawhood is analyzed as membership within the best axiomatic description of the physical facts. And on the other hand, my argument also targets contemporary anti-Humeans about the laws who maintain that the facts about lawhood are analyzed in metaphysical terms (either in terms of relations between universals, e.g., Dretske 1977, Tooley 1977, Armstrong 1983, or in terms of essentially dispositional properties, e.g., Ellis 2001, Mumford 2004, Bird 2007). These two families of views on the nature of lawhood share something in common. Each view holds that lawhood is in some sense analyzable. That is, by each of these views' lights, our metaphysics can ultimately dispense with this additional notion, *lawhood*, in favor of some simpler or more primitive notions. The fact that a physical regularity or relation is a law will turn out to be entirely constituted by the fact that such-and-such another property or set of properties is predicated of it.

My claim will be that these popular views, by virtue of thus being analyses of lawhood, are susceptible to a shared weakness. They are unable to account for the ideal laws' counterfactual preservation. Ideal laws are, being laws of nature, of course counterfactually preserved under events which are logically consistent with them. Throughout this paper, I shall focus in particular on the fact that the ideal laws are preserved under the counterfactual supposition that their idealizations obtain. The problem is as follows: whatever properties, relations, or regularities in fact constitute some ideal law or another differ from whatever would constitute it in the counterfactual event of its idealizations' truth. After all, we live in a non-ideal world, populated by non-ideal properties and regularities; and if the world had been physically ideal, then those very properties and relations would not have obtained. But since those properties are relations are meant to be the analysantia of those laws—that is, what those laws *amount to*—these views of the laws deliver the judgment that the ideal laws are not thus preserved.

The next section of the paper clarifies and further articulates what is involved in the three observations we started with. I finish with the idea that the ideal laws are preserved in the event of their idealizations' truth. Next, I argue in more detail that these popular analyses of the laws imply that the ideal laws are not thus preserved. I finish by showing that a variety of minority views on the laws of nature are able to avoid this argument, and suggest that this may be reason to endorse them.

2 Three Observations Refined

2.1 Idealization and Vacuity

Let's begin by clarifying our three data about the laws of nature. The first observation was that the laws of nature often have idealized content. One does not need to look very far to find examples of this. Indeed, virtually all statements of natural law in mathematical language are idealized (Cartwright 1983). For in order even for any mathematical formalisms, e.g, 'f = ma', to be considered as expressing any law of nature, they must be assigned a physical interpretation, supplied by theoretical or derivational context and in virtually every case heavily idealized. Cartwright's famous example was of Newton's law of universal gravitation, mentioned above. The ideal gas law, *PV=nRT*, is another common example. The formula for a pendulum's period takes both the pendulum's string to be massless and the entire mechanical system to be two-dimensional. Other natural laws idealize physical planes as frictionless, liquids as continuous and incompressible, and so on.[36] There is no shortage of idealization within the laws of nature, even within fundamental physical theories.[37]

It is well-established that idealization in statements of natural law renders those laws vacuous. This was our second observation. For the objects and relations ideal laws represent the world as containing are nowhere to be found. And there is a sense in which there could not ever have been those sorts of objects and relations. After all, it is at least physically impossible for point masses, or a property like frictionlessness or perfect elasticity to be realized. Nancy Cartwright (1983) famously takes this to be evidence that the laws of nature fail to state the facts. Anjan Chakravartty agrees to a similar characterization:

> "In-principle vacuity occurs when law statements describe relations between causal properties that cannot exist in the actual world. This is not to say that such properties could not have existed if the world had been inhabited by different causal properties… Idealization is a rich source of in-principle vacuity. Consider, for example, the ideal gas law, which relates properties of pressure, volume, and temperature of gases. Since no gas is ideal, the ideal gas law is vacuous in the sense that it maps properties and relations which, strictly speaking, do not exist as described" (2007, p143)

---

[36] Cartwright (1983) is the *locus classicus* regarding idealization in statements of natural law; she discusses the above issue regarding the interpretation of statements of law in depth. Also, here and throughout, my examples will be of so-called "Galilean" idealizations, rather than, e.g., abstraction in minimal modeling. See Weisberg (2007) for discussion of kinds of idealization in science.

[37] For example, consider what Sean Carroll says in his exposition of some basic concepts in relativistic physics:

> "The primary usefulness of geodesics in general relativity is that they are the paths followed by unaccelerated test particles. A test particle is a body that does not itself influence the geometry through which it moves—never perfectly true, but often an excellent approximation…the geodesic equation can be thought of as the generalization of Newton's law $f = ma$, for the case $f = 0$, to curved spacetime." (1994, pp108-109)

In other words, the relativistic equivalent of Newton's first law is that an unaccelerated test particle never leaves its geodesic. But this law, as well as the very idea of a geodesic ("a line in spacetime") which is crucial to so many parts of relativistic physics, is defined in terms of the idealized concept of a test particle. And as Bryan Roberts (ms.) agrees, this idealization renders these parts of the theory strictly-speaking vacuous.

The sorts of behaviors that an ideal law describes do not obtain. Yet ideal laws can be laws despite being vacuous.[38] This tells us, importantly, that whatever (if anything) is responsible for their status as laws, it is not the properties, events, and behaviors they represent the world as containing. This point will play a crucial role in the following sections of the paper, where we shall see how it creates problems for contemporary Humean and anti-Humean analyses of the laws alike.

2.2 Counterfactual Preservation

The third and final observation which forms the focus of this paper is that the laws of nature are *counterfactually preserved.* To begin to understand what this bit of jargon means, we need only take a simple step back and recall a distinguishing feature of the laws of nature: that they are not mere accidents.

The standard sort of example runs as follows. Compare the fact that all the people in this room are sitting, and the fact that all ravens are black, One of the most salient differences between these two universal generalizations is that the first pattern of the world is merely accidental or coincidental in a way that the second is not. This difference can be illuminated in large part by observing how each generalization or "pattern" interacts with counterfactual suppositions of a particular kind. 'All the people in this room are sitting' could have failed to have been true if more people had entered the room. But 'All ravens are black' would not have failed to be true even if we introduced several billion more ravens into the world (or if God were to enraven

[38] In case you are worried, Humeans certainly agree to all of this. Barry Loewer, defending Humean supervenience, writes: "[There are] vacuous generalizations some of which seem to be laws; e.g., the ideal gas." (1996, p111).

actual non-ravens). It is, as John Roberts (2008, §5) puts it, "inevitably true" of ravens that they are black. This counterfactual robustness the laws enjoy—their "inevitability", "stability", or "preservation", as it has variously been called—is the very feature of the laws that distinguishes them from merely accidental universal generalizations. I will stick with "preservation."

To put the idea of counterfactual preservation more explicitly, the laws are such that they would still have obtained under any counterfactual supposition consistent with them. The laws jointly (Roberts 2008; Lange 2000, 2009) are counterfactually preserved. But obviously the starting point is that individual laws of nature enjoy this modal resiliency; it is part of the very content of ascribing to some generalization non-accidental status to say that it is counterfactually preserved. Indeed, John Carroll (1994, §1) identifies counterfactual preservation as a conceptually necessary truth linking the notion of a law to that of the counterfactual.

So the laws of nature are preserved under those counterfactual suppositions logically consistent with them. There are some idealized laws of nature. So, the idealized laws of nature are preserved under those counterfactual suppositions logically consistent with them. What I want to call attention to is the following counterfactual supposition and its relation to the ideal gas law:

> "If gases had been ideal, then…"

It seems saliently true that this counterfactual supposition is one under which the ideal gas law is preserved. In other words, it seems saliently true that

> "If gases had been ideal, then it would still be a law of nature regarding the behavior of the appropriate range of gases that their $PV=nRT$"

I can say a bit about why this seems so saliently true. The supposition that gases are ideal is, after all, logically consistent with the truth of the ideal gas law. The equation that states the ideal gas law only manages to do so in the first place because of content supplied by its various background idealizations. So *of course* that mathematical statement of the law, suitably interpreted, is logically consistent with its idealizations. So of course the truth of the law's idealizations would be the sort of event that preserves the law.

The point here can be refined by temporarily dispensing with the vocabulary of 'law' and focusing instead on the closely related underlying notion of physical patterns of behaviors. Consider a mundane case of counterfactual preservation. If all of the objects in my neighborhood were to suddenly become purple, the gravitational interaction—that physical pattern—would be unbroken. That counterfactual event makes no difference to whether the pattern obtains; *that* is why the "law" is unbroken. Likewise, imagine that some ideal gases were to suddenly pop into existence. It seems saliently clear that the pattern of behavior described by the equation 'PV=nRT' would remain unbroken.[39] That behavior, after all, is *exactly* what these new objects instantiate. (Hence the particular phrasing in the consequent above: '…it would be a law that *their PV=nRT*'.) And just as before, it is because the pattern of behavior remains unbroken, and not only because the equation remains true, that it seems appropriate to say that the law is preserved in that counterfactual state of affairs. So in sum there are powerful reasons—stemming from the concept of lawhood itself—to take the ideal laws of nature to be preserved in the counterfactual event of their idealizations' truth.

[39] In fact, you might think that the pattern becomes *better* than being unbroken. For while it is actually vacuous, and is unbroken in virtue of having no disconfirming instances, the pattern would (if ideal gases were to pop into existence) gain positive instances. But this difference hardly implies that the pattern or law would not be preserved – it simply means that its range of application would have broadened. A useful analogy here is to *ceteris paribus* clauses, since sometimes ideal laws are mistaken for or treated as *ceteris paribus* laws. If the conditions required for a *ceteris paribus* clause to be met were to always obtain, surely the *ceteris paribus* law itself would obtain as well. Those are, after all, the exact conditions under which those lawful patterns of behavior are realized. Ditto for ideal laws and their idealizations.

2.3 Two Objections Considered

Let me pause to consider two objections. First, you might be concerned that the idea of the ideal laws' preservation in the event of their idealizations' truth is at odds with the observation that they are in the first place vacuous. After all, if their idealizations had been true—if, for instance, all gases had been ideal—then they would not have been vacuous. Is this reason to doubt what I have said regarding their counterfactual preservation? No. For scientific treatments about the laws themselves provide ample evidence that the same physical pattern—that is, the same law of nature—can be instantiated both with and without any positive cases. The chief example of this is explicit vacuum solutions to equations which express laws of nature, the most famous being Minkowski spacetime as a vacuum solution to the field equations of general relativity. The fact that it is a worthwhile exercise to find whether how the Einstein field equations can be satisfied in a vacuum spacetime is evidence that, by scientific lights, the physical structures and patterns expressed by the Einstein field equations—i.e. the same laws of nature—are preserved even in a counterfactual difference regarding vacuity or non-vacuity.

Second, while it is obviously true that the idealizations involved in, e.g., the ideal gas law, are logically consistent with it, one might still well note that those idealizations are physically impossible. They violate a massive array of other laws. Is this reason to doubt what I have said regarding the ideal laws' counterfactual preservation? No. Indeed those idealizations are physically impossible; and the world would have to be *vastly* different in numerous ways for the truth of the ideal gas law's idealizations to obtain. But there is ample evidence that many laws of nature are preserved even under counterfactual antecedents which violate physical necessity (perhaps even metaphysical necessity), again with the necessary condition that those antecedents do not contradict those laws themselves. In other words, many laws of nature are preserved under a selection of counter*nomic* antecedents. To begin with, recent philosophical literature has discussed the observation that conservation laws are counternomically stable in ways that other laws are not (Lange 2016). If, for instance, gravitation had been an inverse-cube law (instead of an inverse-square law), it still would have conserved total energy—so long as it still were time-translation symmetric. (This is because of Noether's Theorem, according to which every differentiable symmetry corresponds to a conserved quantity.) Another example is as follows. It is a common exercise in introductory electromagnetism classes to identify which of Maxwell's equations, which state the laws for electromagnetism in classical contexts, are preserved under the counternomic supposition that magnetic monopoles exist.[40] All of this suggests that laws of nature can be and often are preserved under counterfactual suppositions that happen to be physically impossible. It is simply up to science and its mathematics to tell us when.

So, there is, as we have seen, good reason to think that the ideal laws of nature would be preserved in the counterfactual event of their idealizations' truth. A counterfactual antecedent like, 'If all gases had been ideal…' is physically impossible, but there is no reason to think that it thereby fails to be law-preserving with regard to the ideal gas law. It is perspicuously consistent with that law. And to be sure, the truth of an ideal law's idealizations would make it non-vacuous, but that difference is not enough to doubt that the law would be preserved if that were to obtain.

[40] See, for example, Schwinger 1998: §3, #8 and Jackson 1975, pp254-256.

3 Counterfactual Preservation and the Analyses of Lawhood

3.1 Introduction

In this section, I turn to the main argument of the paper. I shall argue that three of the most popular views of the nature of physical law fail to uphold what we have established in the previous section: that the ideal laws of nature are preserved in the counterfactual event of their idealizations' truth. Let me begin by introducing the targets of discussion in this paper, as well as calling to attention the shared weakness which renders them susceptible to the argument I present.

To recall, my argument targets on the one hand contemporary Humean analyses of the laws of nature. Humeanism about the laws of nature in its contemporary form takes the shape of the "Best-Systems Account" of laws, according to which the laws of nature are those regularities whose statements in terms of universal generalizations are theorems of the best axiomatized descriptive system of the world's physical facts (see, among many others, Lewis 1973, 1983, Beebee 2000, and Cohen and Callender 2009). The idea is that the best "system" is the one that strikes the best balance between simplicity and descriptive or explanatory power. The canonical formulation, owing to David Lewis, goes like this:

> "…a given regularity might hold either as a law or accidentally, depending on whether other regularities obtain that can fit together with it in a suitable system. … [the system] must be as simple in axiomatisation as it can be without sacrificing too much information content; and it must have as much information content as it can have without sacrificing

> too much simplicity. … only the regularities of the system are to count as laws" (1983, p41)

The Humean account has two components. First, the laws of nature are in some deep sense *nothing more than* mere regularities in nature. For even considered as elements of a representational system, all that the laws of nature do is *describe* what happens (cf. Beebee 2000). They merely state—and neither govern nor in any deep way explain—the physical regularities. Second, the property of being a law is to play a particular role within a system. A proposition's lawhood is determined in part by the relations it stands in to other propositions within that "best" system, and also, as I shall argue, by the relations that the physical regularities responsible for its lawhood stand in to the other law-constitutive regularities.

My argument also targets two varieties of contemporary anti-Humeanism about the laws. The first anti-Humean view that falls in the scope of those that I will be targeting is Dretske (1977), Tooley (1977), and Armstrong's (1983) view that the laws are relations of necessitation between universals. The view, quite familiarly, is that the lawhood of all Fs being G is a matter of there being a relation N(F, G) that obtains between the universals F and G, and which necessitates the physical regularities that all Fs are Gs. The relation itself, N, is generally taken to be an ontological primitive (cf. Hildebrand 2013).

The second sort of anti-Humeanism I have in mind is dispositionalism, which has become increasingly popular in recent years. Dispositionalism takes a similar "ontological" approach to the laws of nature as the universals-view: on this view, the laws of nature are causal relations between event-types, necessitated by essentially dispositional properties (Bird 2005, 2007; Chakravartty 2007). For instance, *being charged* looks like it essentially confers on its bearers the disposition to repel or attract other charged objects in proportion with charge strength and in inverse proportion to distance (Shoemaker 1980). And this essentially dispositional character of charge, universally shared among all charged particles, looks like it can ground the lawhood of the fact that all charged particles engage in the behaviors described in Coulomb's Law.

3.2 On Lawhood and (Metaphysical) Analysis

These three views are well-entrenched. And their deep disagreements are well-documented. But I shall argue that they are vulnerable to a shared weakness. This weakness is due to the fact that these views of the laws of nature alike purport to be—or aim to provide—*analyses* of the notion of lawhood. That is, each of them in their own way purports to show that the property of *being a law* can be dispensed with in favor of some other or more primitive properties.

The Humeans' reductionist tendencies are well-known. David Lewis, in the canonical formulation of Humean supervenience, asserts that "*all there is* to the world is a vast mosaic of local matters of particular fact" (1986, ix, emphasis added). And focusing on the Humeans, John Carroll writes:

> "To take laws seriously, philosophers dread that they would have to recognize necessary connections or other similarly mysterious existing entities as really existing in nature…Whatever drives the suspicions, the preferred method of squelching them is clear. Convinced that there is a significant class of more basic concepts, philosophers seek a definition of 'law of nature'. They seek an analytic completion of '*P* is a law of nature if and only if…' (1994, p5)

The core of the Humean project is, of course, to show that the mysterious "glue" that they take other metaphysicians to require—in this case the glue of lawfulness—can be dispensed with. Humeans seek, in other words, to supply an analysis of what lawhood really *amounts to*.

But these anti-Humeans do, too. They, too, purport to have views that declare what lawhood really amounts to. Dispositionalists and other necessitarians about the laws, for instance, judge that the apparently distinct phenomenon of nomological necessity *just is* metaphysical necessity (Bird 2005, 2007, Mumford 2004). And although defenders of the universals-approach to lawhood can attest to the ontological primitiveness of their nomic necessitation relation between universals, the view itself still amounts to an analysis and dispensation with lawhood. Tyler Hildebrand observes:

> "the locution 'It is a law that'…can be translated into different statements containing different primitive terms. For example, one who thinks of laws as relations between universals can translate 'It is a law that all Fs are Gs' into 'Universal F stands in the relation of nomic necessitation to universal G'…In this analysis, the concept of lawhood is analyzed in terms of the supposedly more primitive concepts of universals and nomic necessitation" (2013, pp2-3)

So, while the dispositionalist view and the view of laws as relations between universals are distinctively anti-Humean in their "gluey" metaphysics, they too aim in their own way to dispense with the notion of lawhood. For by claiming that facts about lawhood can be entirely understood in terms of dispositional essences, or relations between universals, these views too declare that lawhood amounts to something else.

This shared feature of these views of the laws has, to my knowledge, not been discussed in the extant literature. And it is the feature that, as I shall argue, opens these views—despite their many deep differences—to a powerful objection. But there is an objection you might raise at this point. One might wonder whether there are multiple aims and methods to giving an account of the laws represented among these three views and their proponents. My argument turns on the fact that they are all analyses. But are these views all trying, as I have said, to give an analysis of what lawhood amounts to? You might think that there are two projects that might be going on here. First, philosophers might be aiming to give an account of what is going on, metaphysically speaking, in everything that is called a law. Second, philosophers might only aim to understand the metaphysics that makes our lawful world possible, that is, the metaphysics underlying whatever it is that *turns out* to be a law, but without an eye to any particular law (except perhaps fundamental laws and symmetries).[41] On this second approach, it is possible—though not required—to have a view that rules out certain law-ascriptions as veridical, whereas it might appear that the first approach instead takes all the apparently veridical law-ascriptions as its starting point and aims to accommodate them. We might put the thought this way. On one approach, the core question is, "What is going on (metaphysically) in these things here, the laws of nature?" On the second approach, the core question is, "What sort of metaphysics or structure would underlie the very possibility of our concept of law being physically realized?"

This objection deserves careful thought. Indeed it may appear that many philosophers who have engaged in this inquiry seem to be answering one of these questions but not the other. Some portions of the literature on laws allot more attention to, for instance, understanding how to

[41] Thanks to an anonymous referee for raising this insightful objection.

accommodate within the metaphysics the thorny issues that arise regarding ascriptions of 'law' within scientific practice (Roberts 2008) or to how nomological concepts appear in scientific modeling (Cartwright 1983). Only relatively recently have Humeans, for instance, begun to work out whether and how *ceteris paribus* laws and law-ascriptions can fit into the Humean picture of laws (Cohen and Callender 2009). Are all these philosophers engaged in the same project?

Yes. In response to the core of the objection, I deny that these two purportedly distinct "aims" of metaphysical inquiry into the laws actually differ. For in order for them to be genuinely distinct, they would have to have different target phenomena. The intension of 'law' specified in one question would have to be different than in the other. But I do not think that there are—or rather, that there should be—different contents to the target phenomenon 'law'. For one thing, we do not have much of an *a priori* concept of physical law that we might independently conduct metaphysical inquiry regarding. And even if it did, surely it *shouldn't*. It should be the case, methodologically speaking, that the content of our concept of natural law be guided by the apparently veridical law-ascriptions made by science. If so, to give a metaphysical analysis of the laws and to give a metaphysical picture of what might underlie any law are the same task.

To be sure, there is room for reasonable philosophical disagreement about which of these law-ascriptions we should hold in higher esteem; some philosophers may be more amenable to declaring certain portions of science mistaken about such-and-such being a law of nature (e.g., how physicalists might view the apparent laws of the special sciences). Still, it seems clear that, given that our concept of law *is* (and certainly ought to be) supplied by what the sciences discover, all of these metaphysicians are in fact engaged in inquiry regarding the same thing. That they may sometimes come to disagreement about which things are in fact laws, or which parts of scientific treatments of laws are worth attending to, is simply a result of the thorniness of philosophical discourse. The aim of our philosophical discourse in these cases still has been to deliver an analysis, within our metaphysics, that might tell us what the slippery phenomenon of "law" pervasive in our scientific theories actually amounts to.

3.3 The Laws as they Actually Are

Can these views accommodate the fact that the ideal laws would have been preserved, if their idealizations had been true? I shall argue that they cannot. The reason is that, if the idealizations made by (e.g.) the ideal gas law had been true, then what is responsible for making it a law of nature in the actual world (its "lawmakers") would not have been preserved. But remember that these views are meant to be *analyses* of the laws. They are views only infelicitously of what "make" the laws laws: they are really views of what the laws of nature are constituted by and, ultimately, amount to. So, if a law's actual analysans—what it ultimately amounts to—is different in the event of its idealizations' truth, then that just means that, by the lights of whatever analysis of lawhood one has taken on, the law itself is not preserved in that event.

This summarizes the argument. To see in more detail what I mean by all of this, recall first that statements of ideal law are vacuous. The physical world again contains no ideal objects, properties, or relations. What is it, then, that is responsible for making an ideal law into a law? That is, what is it in the world that the fact that (e.g.) $PV=nRT$ is a law answers to? It must be something other than what the ideal law actually represents the world as containing, since there are no such properties or objects. Regardless of whether you are a Humean or an anti-Humean, you must agree that if there is anything at all that an ideal law ultimately is constituted by, it must be *non*-ideal stuff.

At this point, you might wonder: "Really? Must it really be vacuous?" Yes. Consider the ideal gas law, for instance. Is there really no such physical state of affairs such that *PV=nRT* in gases? Indeed there is not. For consider the regularities and relations that actually obtain in a gas sample. The constituents of that sample will see some of their kinetic energy change into potential energy as they move around, due to friction and inelastic collision—contrary to the states of affairs that the ideal gas law describes. The constituents of the sample will see slight variations from their "ideal" trajectories due to electromagnetic interactions, which again, the ideal gas law represents as absent. Go looking for the physical relations and behaviors that obtain in a sample of gas, and you will only find these highly determinate causal relations, properties, and regularities, all of them non-ideal. We might say that the equation '*PV=nRT*' approximates some actual physical facts. But the fact that, literally, a gas's *PV=nRT* simply does not obtain.

So what does the fact that ideal gas law is a law actually amount to, if there is no such physical fact? Let us begin by calling the actual, highly-determinate physical fact that obtains '*@PV=nRT*', and consider the Humean and anti-Humean analyses in turn. By the lights of the Humean best-systems account, the laws of nature do nothing more than state which regularities there are (i.e. those that belong in the best system), the account implies that the laws in some sense reduce to the regularities they describe (Hall 2015). This is why Humeans sometimes speak in terms which suggest that those regularities *themselves* "count as laws" (Lewis 1983, p41) or "express laws" (Loewer 1996, p180). So, for Humeans, the fact that the ideal gas law is a law reduces to the fact that *@PV=nRT*. Or to put it slightly more long-windedly, the fact that '*PV=nRT*' is a law reduces to the fact that there is a particular regularity, i.e. *@PV=nRT*, that obtains with the appropriate level of generality and at an appropriate level of fundamentality that that (idealized) expression of it counts as a theorem of the best system.

The same general thought applies to the anti-Humeans we've chosen to target, *modulo* a few changes to account for the particularities of their views. Anti-Humeans will standardly note that

> "In common parlance we use the term 'law' to refer both to relations between causal properties and to linguistic devices employed to represent them" (Chakravartty 2003, p405)

And that

> "laws are *expressed* by singular statements of fact describing the relationships between properties and magnitudes" (Dretske 1977, p263)

That is, we can draw a distinction between laws – that is, the things in the world – and laws considered as "statements" of law, i.e., equations like *'PV=nRT'*. This might seem to help the problem, since as Chakravartty notes, "In-principle vacuity attaches, where it does, to expressions representing laws, not to laws themselves. Causal laws are never vacuous in principle, given that they are relations between causal properties" (ibid). So we have a picture of what it is that we ascribe to the equation '*PV=nRT*' when we say that it is a law. The fact that it is a law is analyzed in terms of its successfully "expressing" some metaphysical state of affairs that actually is "the" law.

But it is important to note that anti-Humeans thereby should not—indeed, cannot—say that the equation '*PV=nRT*' is a law because it successfully represents the physical fact or the ontological law *that PV=nRT*. For again, there is no such fact or law; the equation '*PV=nRT*' is strictly speaking vacuous. It must be a statement of a law because, despite representing the fact that *PV=nRT* as obtaining, it actually expresses some other state of affairs. And we have already identified which one: the state of affairs that @*PV=nRT*. In other words, it must be the fact that @*PV=nRT* obtains, and that '*PV=nRT*' manages to express that it obtains, that constitute the fact that that equation is a law in the actual world.[42]

There is an objection that anti-Humeans might raise here. Can uninstantiated properties – e.g, Platonic universals, or dispositional properties – serve as the analysantia of the fact that '*PV=nRT*' is a law? I think not. For while it is certainly open metaphysically whether properties can exist uninstantiated, it is hard to see how this suggestion might help here. The law '*PV=nRT*', despite being vacuous, is still applicable within a wide range of scientific endeavors and hence appears to earn its status as a law because of events and behaviors that are physically realized. If what it was that constituted the fact that '*PV=nRT*' is a law were uninstantiated universals or dispositional properties, this would help us avoid the vacuity of the law but would also imply that it is a total mystery why it manages to be useful in actual scientific practice. Surely, given that '*PV=nRT*' is a law of the sciences of our actual, non-ideal physical world, if there is anything that explains why, it is some actual physical goings-on, and not some uninstantiated business in

[42] Take note: the reasons that defenders of the Dretske-Tooley-Armstrong view and dispositionalists should agree that this non-ideal physical relation, @*PV=nRT*, is what really constitutes the law, are slightly different. Dispositionalists are pretty straightforwardly committed to this conclusion: the (dispositional) properties that are instantiated in our world are non-ideal ones, so they cannot be the sorts of properties that the ideal causal relation that *PV=nRT* would relate. By contrast, defenders of the Dretske-Tooley-Armstrong view can agree that the property of pressure, *P*, is instantiated, since they can identify that property as a universal that admits of many determinate magnitudes, some of which are in fact physically realized. But they should not say that the *relation* that *PV=nRT* is instantiated, since it is not the case in our non-ideal physical world, magnitudes of *P* are related with magnitudes of *V*, *R*, and *T* in the exact way the ideal gas law says. So, the reasons they have for doing so are slightly different, but proponents of both dispositionalism and the Dretske-Tooley-Armstrong view alike must say that '*PV=nRT*' is a law because of the non-ideal relation that @*PV=nRT*.

Plato's heaven. And if so, then it must still be the physically-realized state of affairs that *@PV=nRT* that accounts for the fact that it is a law.

3.4 A World of Pure Idealization

Now let us turn to the counterfactual supposition that the ideal gas law's idealizations are true. That is, let us turn to the counterfactual supposition that all gases are ideal, and suppose that there were—to borrow from a classic film—a world of pure idealization. Under this counterfactual supposition, all of our analyses of the laws declare that the ideal gas law is not in fact preserved.

If all gases were ideal, then the fact that *PV=nRT*, which cannot actually be found anywhere in our world, would be physically realized. And if all gases were ideal, then the fact that *@PV=nRT*, in all of its highly-determinate, non-ideal glory, would no longer be realized. For the sorts of objects ideal gases would be would be entirely distinct from those that actually exist and compose actual gases. Indeed, entirely different properties would have needed to be instantiated for the ideal gas law to be non-vacuous. The parts of the world that the fact that the ideal gas law amounts to would be "swapped out" for new, entirely different physical stuff.

This is where our analyses of lawhood encounter difficulties. Consider first the Humean. The Humean view, as we have seen, must take the fact that the ideal gas law is a law to reduce to the fact that *@PV=nRT* obtains. If gases had been ideal, then the regularity that *@PV=nRT* would not have obtained. For, to reiterate the point that was made in the previous section, the regularity that *@PV=nRT* is composed of events where gases in fact collide non-elastically, where they are subject to electromagnetic attraction and repulsion, where they are not perfectly hard point masses, and so on. If gases had been ideal, then none of these events would have obtained. So, if gases had been ideal, the regularity that *@PV=nRT* would not have obtained. But since *all there is* to the fact that '*PV=nRT*' is a law is the regularity that *@PV=nRT*, the Humean view judges that that law does not obtain in that counterfactual state of affairs.

(It will not help, by the way, for the Humean to insist that ascribing the property of *being a law* to a proposition really means ascribing to it the property of *being a theorem of the best deductive system.* For they might think to avoid the argument in the previous paragraph by insisting that the ideal gas law *is* preserved; after all, if gases had been ideal, '*PV=nRT*' would likely still show up in the best axiomatic system. However, a characterization of the World of Pure Idealization as one in which the actual ideal gas law is preserved is not only infelicitous, but contrary to the heart of how Humeanism about the laws wants to treat them. For even if some of the same propositions—including, say, '*PV=nRT*'—were still included in the best system, it is clear that what would *happen* in a state of affairs where gases are ideal is radically and fundamentally different from what actually happens in gases. And for the Humean, which laws there are is ultimately a matter of what happens. Since what happens in gases is radically different, it is not the case that the ideal gas law is preserved.)

The anti-Humean views under consideration fall prey to the same weakness. The fact that *PV=nRT* is a law is analyzed in terms of the fact that a particular state of affairs, namely *@PV=nRT*, obtains, and the fact that the *equation* '*PV=nRT*' expresses it. However, if gases had been ideal, neither of these analysantia of *PV=nRT*'s actual lawhood would have obtained. If gases had been ideal, it would not have been the case that *@PV=nRT* obtains or is an actual, metaphysical law. And *a fortiori* it would not have been the case that the equation *'PV=nRT'* would be a law by representing or expressing it. Of course, the equation *'PV=nRT'* would have expressed something else: it would have expressed the fact that *PV=nRT*, which it actually represents but merely vacuously. Nevertheless, the parts of the world in which the fact that '*PV=nRT*' is a law actually resides would have been different. And this means that by the lights of these anti-Humeans, the ideal gas law would not be preserved in the counterfactual event of its idealizations' truth.[43]

Humean and anti-Humean alike take the fact that the ideal gas law is a law to amount to some other fact or set of facts that, if gases really had been ideal, would not have obtained. So, the Humean and these anti-Humean analyses of lawhood alike deliver the judgment that the ideal laws are not preserved in the event of their idealizations' truth. But as I have argued, there is a powerful presumption that the ideal gas law *would* be preserved, if its idealizations had been true. This presumption stems from the very nature of our concept of lawhood, and the observation that the ideal laws – like any other laws of nature – are not accidents; they are preserved under those counterfactual assumptions consistent with them. So the fact that these popular analyses of the laws fail to uphold the counterfactual preservation of the ideal laws casts doubt on their attractiveness as accounts of the nature of physical law.

[43] You might wonder whether necessitarians about the laws will accept this argument since they, believing laws to be metaphysically necessary, will take them to hold under any counterfactual antecedent. However, I do not think this rule applies here. The antecedent 'If gases had been ideal…' will, by necessitarian lights, look to be metaphysically impossible, because it looks to violate many of the natural laws (including whatever the *actual* gas-properties-and-laws are). When faced with a counterpossible antecedent like this, that is, one that directly violates something you take to be metaphysically necessary, one ought to reasonably take that particular metaphysical necessity not to be preserved. For instance, even if one were a mathematical Platonist who took numbers to exist necessarily, one should still agree that under the counterfactual antecedent 'If there were no abstracta…', 'The number 7 exists' would not hold. This reply presumes that sometimes, counterpossible conditionals can be non-vacuously true, for which I think there is ample evidence. (See, e.g., Nolan (1997), and for scientific examples, Jenny (2018).) Thanks to an anonymous referee for raising this objection.

4 Towards A New Account of Lawhood

I have argued that contemporary Humeanism and two dominant varieties of anti-Humeanism about the laws fail to accommodate the datum that the ideal laws are preserved counterfactually in the event of their non-vacuity. The argument against both of these views can be generated by following a simple recipe. First, identify what it is in the world that some ideal or otherwise vacuous law actually consists in. Then, observe that these analysantia are physically non-ideal, and would not obtain if things had been physically ideal. It follows that that law is not preserved under that counterfactual supposition, since what in our metaphysics it actually is constituted by differs from what it would be constituted by, if its idealizations had been true.

Can this argument be avoided, by any account of lawhood? In this final section, I show that a variety of views of lawhood can. In particular, views of the nature of lawhood which do not attempt to analyze laws in terms of physically-realized facts can avoid the argument. The fact that these views can avoid my preceding arguments is a tentative reason to endorse them, although in the remaining space here I will not be able to adjudicate between these views as competitors to each other.

The first view that is worth considering is one defended by Marc Lange (2000, 2009), which takes the fact that a proposition is a law to be accounted for by its counterfactual stability. Lange's notion of counterfactual stability is a technical one, but it extends the intuitive idea of counterfactual preservation that has appeared throughout this paper. Roughly, a set of propositions S is stable iff the members of the set still would have been true given any counterfactual antecedent that is consistent with the members of the set taken jointly. To illustrate with an example, a set that includes the generalization, 'All the people in the theater are sitting' fails to be stable since at least one member of that set would be rendered false in the counterfactual event of there being a fire in the theater. Lange argues that the laws of nature are the members of the counterfactually stable sets, and takes counterfactual facts to be primitive and irreducible parts of our ontology. A fact or proposition's lawhood, then, is a matter of its enjoying the right degree of counterfactual stability relative to these primitive counterfactual facts. And on this account a proposition can even be "more of a law" (i.e., be more necessary) than another by being a member of a subset of the laws that is more counterfactually stable than the initial set of laws.

This view is compatible with the datum that the ideal laws are preserved in the counterfactual event of their idealizations' truth. For one can use it avoids my argument by avoiding any modal difference in what constitutes the fact that the ideal laws are laws. If you thought, like Lange does, that there are primitive counterfactual facts in our ontology, then you might easily also think that those counterfactual facts are not physically realized: maybe they are Platonic entities, or maybe they are just the occupants of a *sui generis* ontological category. If so, then the same counterfactual facts and relations of stability that actually constitute the fact that such-and-such an ideal law is a law can obtain even in ideal physical states of affairs. The conclusion that familiar versions of anti-Humeanism and Humeanism face – that the ideal laws' lawhood is not preserved in the event of their idealizations' truth – thus does not seem to apply to Lange's view.

Another view of the laws that avoids my arguments throughout this paper is Timothy Maudlin's (2007) view that the laws are sui generis entities, that occupy their own fundamental ontological category. Here is a pithy formulation and motivation of the view:

> "Nothing in scientific practice suggests that one ought to try to reduce fundamental laws to anything else. Physicists simply postulate fundamental laws, then try to figure out how to test their theories; they nowhere even attempt to analyze those laws in terms of patterns of instantiation of physical quantities." (Maudlin 2007, p105, original emphasis)

Maudlin's view, properly understood, should apply only to fundamental physical laws. But as I have already pointed out, idealization (and the concomitant vacuity) are present even in fundamental physical theories. This view that the laws are unanalyzable also avoids my arguments throughout this paper. If there is nothing that constitutes some law's lawhood, then of course there is no possible modal difference in what constitutes a law's lawhood.

One last view of the laws which is worth considering is Harjit Bhogal's (2017) "minimal anti-Humeanism". According to this view, the laws are the true universal generalizations which are not grounded in their instances. (For those who, like me, are suspicious of metaphysical grounding (e.g. Wilson 2015), we can simply say that the laws are the true universal generalizations which are not determined by, or which fail to non-trivially supervene on their instances.) Why should we think that this is what lawhood is? For one thing, Bhogal argues that this view of laws helps avoid the circularity problem that Humean reductionism about the laws faces. To put it briefly, Humeans say that the laws are determined by their instances, but science says that the laws also explain their instances, so the Humean picture of the laws places them in an explanatory circle. If we simply deny that the laws are determined by their instances, then we can avoid the explanatory circle.

The arguments I have provided throughout this paper provide an additional reason to support a view like Bhogal's minimal anti-Humeanism. For a view like this one can avoid my arguments.

If the laws of nature are not determined by their instances, then it is not the case that the "lawmakers" for the ideal laws—i.e., what constitutes the fact that they are laws—are non-ideal physical regularities. And if there had been ideal physical objects, it would not have been the case that the lawmakers for the ideal laws are *ideal* physical regularities. The laws simply aren't determined by their instances; so—just like for Maudlin's view—there is no possible modal difference in what constitutes the fact that there is this or that ideal law. On both Maudlin's and Bhogal's views, like Lange's, the ideal laws can be counterfactually preserved in the event of their idealizations' truth.

6 Conclusion

It is an important datum about the laws, which we should aim to uphold, that the ideal laws would be preserved if their idealizations were true. I have argued that if anything physically realized is responsible for the laws' being laws, then the ideal laws are not counterfactually preserved in the event of their idealizations' truth. Two popular varieties of anti-Humeanism about the laws are vulnerable to this argument, as is Humean reductionism about the laws. Their failure to counterfactually preserve the ideal laws casts doubt on their attractiveness as analyses of lawhood.

By contrast, it seems that a variety of views of laws which take lawhood not to be constituted by physically-realized properties or regularities avoid my argument. Because they take the lawmakers (insofar as there are lawmakers at all) not to need to be physically realized, they avoid saying that there is a modal difference in what constitutes the ideal laws' actual, vacuous lawhood and counterfactual, non-vacuous lawhood. So these views are capable of upholding the datum that the ideal laws' actual lawhood is preserved in the event of their idealizations' truth. This is a reason to consider these views of lawhood more seriously moving forward.

Chapter 5: The Metaphysical Implications of Model Transfer Between Sciences

1 Introduction

Model transfer is the scientific practice of taking a representational model that is built for one particular kind of target system, in some particular scientific domain, and applying it for representational work on another, disparate kind of target system, in another, disparate scientific domain. Contemporary science is replete with examples of model transfer. The Lotka-Volterra predator-prey model, which has its basis in population ecology, has a wide variety of transfer applications in social sciences (Donhauser and Shaw 2018). The quantum theory of atoms in molecules is a case of the virial theorem being transferred from astrophysics to physical chemistry (Price 2018). In "econophysics," models of the transfer of kinetic energy in a gas are transferred to model macroeconomic interactions (Chakrabarti et al. 2013).[44] The Ising model for total energy in a ferromagnetic-paramagnetic phase transition has been transferred to model the firing state of neuronal clusters (Schneidman et al. 2006). The Klein-Gordon equation, which describes the behavior of a relativistic electron, can also be used to describe the free motion of a string in a rubber sheet. The list goes on and on.

Despite how common model transfer is in contemporary scientific practice, very little philosophical attention has been paid to it. This is somewhat surprising, since model transfer immediately raises a number of philosophical issues. All of the very limited philosophical attention to model transfer thus far has been on broadly its epistemic implications. For example,

[44] See Gallegati et al. 2006 for criticism from traditional economics, and Thebault and Bradley 2018 for philosophical discussion.

how can it be possible to gain evidence about one sort of system by way of another very radically different sort of system (if it even is possible)?[45] Given the prevalence of model transfer in contemporary science, must we revise our conception of what our scientific theories nowadays are constituted by, and hence what scientific knowledge consists in?[46] Presumably there are both legitimate and illegitimate cases of transferring a model from one domain to another: what justifies the legitimate cases, and what does this justification tell us about the epistemic status of idealization and analogical reasoning in scientific modeling?

This paper examines the *metaphysical* implications of model transfer in the sciences. I wish to bracket the broadly epistemic issues I have just mentioned, and focus instead on this question: what does the success of model transfer methods tell us about the ontology of the natural world? This is not to say that the epistemic issues are unimportant. Clearly it is important to understand what legitimizes model transfers and how to judge their success. But still, there obviously *are* legitimate cases of model transfer, which meet every *prima facie* criterion of empirical success. And focusing on those legitimate cases, we face an important metaphysical question to ask which has heretofore been entirely neglected. What is it in the world that drives the empirical success of legitimate cases of scientific model transfer? It is a conviction of selective scientific realists—one that I share—that empirical success in our sciences does not simply happen without something in the world to account for it. This paper takes the selective realist's stance and argues that the success of model transfer provides evidence for a surprising ontological conclusion: that there are irreducibly higher-level, multiply-realizable causal process kinds.

[45] Donhauser and Shaw 2018, Castellane and Paternotte 2018
[46] Humphreys 2004, 2018, Zuchowski 2017

I proceed as follows. In the next section, I set the background by restricting our discussion to a few cases of clearly legitimate model transfer. Then, I introduce a puzzle about how to characterize these cases vis-à-vis received views on representation and ontology. This puzzle motivates a naturally plausible response in terms of an ontology of higher-level, irreducibly multiply realizable causal processes. This response not only enjoys a high degree of pretheoretical plausibility, but—as I argue—it also satisfies plausible scientific realism-driven constraints on the metaphysical implications of our scientific theories. We thus have both strong *prima facie* reasons and theoretical reasons to accept an ontology of higher-level, irreducibly multiply realizable, "generic causal processes". However, this is much more radical, ontologically speaking, than it initially appears; that we have good reason to accept it is surprising. I finish by discussing some of its broader metaphysical implications, which are equally surprising. These implications span into the metaphysics of the laws of nature, a new defense of ontic structural realism, and a new conception of the unity of the sciences.

2 Three Examples of Successful Model Transfer

2.1 Introductory Remarks

In this section, I begin by introducing some examples of successful model transfer. As we shall see, a striking feature of these cases is that the model's origin domain—the sorts of objects and properties it was *built* to describe—are radically different from the sorts of objects and properties in its transfer domain. The fact that these models can successfully describe and predict events in systems radically different from the sorts of systems they were built to describe is a surprising empirical fact that calls out for explanation.

I intend to do more than simply introduce a few cases of model transfer. I also wish to restrict the focus of our current discussion to a few cases of model transfer where the model in question is broadly dynamical or causal in character, and in both its origin domain and transfer domain provides causal-explanatory about the targeted physical system's behavior. The reasons for this restriction are simple. There is a long tradition in both the history of science and the metaphysics of science of taking observations of a causal or dynamical nature as strong evidence for new ontology. For example, on account of causal irregularities observed in the Stern-Gerlach experiment, physicists accepted that electrons have the fundamental property of spin. Because of causal observations in the photoelectric effect, Einstein proposed (and the scientific community accepted) that photons exist and that light is not continuous. The list, again, goes on and on. My arguments about the metaphysical implications of model transfer follow in these broadly scientific realist footsteps that have paved much of the road of the history of science.

2.2 The Ising Model

To introduce my first example of model transfer between sciences, consider the fact that when an iron-based magnet is heated up past a certain point, it loses its otherwise-permanent ferromagnetic properties. Other magnetic materials you stick to it will fall right off. This is an observable effect of a phase transition: just like liquids have transitions into solid and gaseous phases, ferromagnetic materials have phase transitions where they become paramagnetic or antiferromagnetic.

At the microphysical level, what explains a ferromagnetic material's phase transition is that the electrons that help constitute the material become unaligned in their spin direction. Spin generates magnetic fields; in normal conditions, a ferromagnet's constituent spins are all aligned, hence at macroscopic measurement scales it manifests a magnetic field. However, increasing the heat eventually causes its spins to flip in various directions; thus, in a phase transition, ferromagnets lose their overall magnetic field.

The Ising model provides a simple and effective representation of a ferromagnetic phase transition. Though my discussion will not involve technical details at all, it is still worth introducing the mathematical statement of the model to introduce what its content is. The Ising model in two dimensions represents a ferromagnetic system as a lattice of up-or-down spins.[47] The total energy E of the system to be given by the following function:

$$E = -J \sum_{<i,j>} s_i s_j - \mu \sum_j h_j s_j$$

Again, technical details are unimportant, but what is important to note for our discussions is that the model denotes electromagnetic properties. For example, $\mathcal{J}$ denotes the net magnetic interaction between adjacent spin sites such that the summation function describes the entire lattice's magnetic output (if J < 0, the interaction is antiferromagnetic as opposed to ferromagnetic if J > 0). The core insight of the Ising model is that in the model as well as in nature, macroscopic properties like being ferromagnetic or antiferromagnetic are realized by interactions between pairs and clusters of spins, rather than individual spins taken

[47] There also exist one-dimensional and multi-dimensional (>2) Ising models, but our discussion need not involve these in any detail.

independently. But it is clear what the Ising model's content in its origin domain is: it was built as a model of phase transition in ferromagnetism.

There is a surprising transfer of the Ising model to a domain radically different from that of condensed-matter physics. It has recently seen uptake in neuroscience. Suppose we want to understand how the firing state of a cluster of neurons considered as a whole relates to the firing state of the cluster's individual constituent neurons. You might think that the best way to represent this is to assume that when one neuron in the cluster fires there is no statistically significant correlation that its neighboring neurons also do. A recent experimental result considers this sort of method for predicting firing rates in a 40-ganglion-cell cluster in a salamander's retina when presented with a film recording of a natural environment (Schneidman et al. 2006). And this modeling method

> "fails disastrously…If the cells were independent, [the probability *P(K)* that *K* of these cells generate a spoke in the same small window of duration] would approximate the Poisson distribution, whereas the true distribution is nearly exponential. For example, the probability of $K = 10$ spiking together is ~10^5x larger than expected in the independent model…The independent model makes order-of-magnitude errors even for very common patterns of activity." (p1007)

It turns out not only that a model that countenances strong correlations between the firings of individual neuron neighbors makes up for these dramatic empirical failures, but that the model that does so best is the Ising model:

> "…we will see that the energy function relevant for our problem is an Ising model […] Here, the Ising model is forced upon us as the least-structured model that is consistent with measured spike rates and pairwise correlations; we emphasize that *this is not an*

> *analogy or a metaphor, but rather an exact mapping* [...] Looking in detail at the patterns of spiking and silence of one group of 10 cells, we see that the predicted rate for different [system-wide measurements] are tightly correlated with the observed rates over a very large dynamic range, *so that the dramatic failures of the independent model have been overcome*" (p1008-9, emphasis added)

This experimental result is surprising. We can distinguish this result from earlier, purely theoretical applications of Ising-model principles to model artificial neural networks; this was an attempt at transferring the model that relied on particular hypotheses about how neural networks evolve over time. This result, in virtue of its actual empirical confirmation of the Ising model's application to networks of real neurons, is different in kind. Indeed, this "transfer" is not so much an application of the Ising model to a different scientific domain as it is a *discovery* that the Ising model characterizes the behavior of a system quite radically different from its origin domain. This discovery is surprising enough that it seems to call out for a deeper explanation. What, in the world, could account for the remarkable success of a model very much built to characterize a particular kind of physical behavior (ferromagnetic phase transition), in characterizing a radically different kind of behavior in a radically different kind of physical system?

2.3 The Black-Scholes Model

Let me now introduce a second example of model transfer. If you heat up a metal bar at some point along its length, the hot section will gradually diffuse its temperature into the adjacent sections of the bar. The mathematical model that is most frequently used to represent this well-known phenomenon is known simply as the heat equation or diffusion equation:

$$\frac{\partial u}{\partial \tau} - \frac{1}{2}\sigma^2 \frac{\partial^2 u}{\partial x^2} = 0$$

The heat equation is due to Jean Baptiste Joseph Fourier, in 1807, when the study of thermodynamics was still nascent. And the heat equation is in every sense built specifically as a thermodynamical model. Fourier's derivation represents the metal bar in one dimension, so that heat conducts only along the length of the rod. The derivation also relies on Fourier's Law, which describes the relationship between the time derivative of heat transfer in a material and the temperature gradient. Additionally, the element σ denotes the thermal conductivity of the metal bar's material. From these observations it should be clear that the heat equation's content is thermal conductivity behavior.

However, the heat equation (or at least its form) has enjoyed a remarkable level of uptake in other areas of science. Here I present one particularly surprising transfer of the heat equation to model another phenomenon; what is remarkable about this case is, as in the previous one, the extremity of the differences in the sorts of phenomena that the model is used to characterize.

A stock option is a traded financial instrument that gives its holder the option, but not the obligation, to purchase shares of an underlying stock at an agreed-upon price (the "strike price").[48] Imagine that I hold an option for 100 shares of stock in company C at $15.00 per share, but before the option expires, the price of C rises to $25.00 per share. My holding of the option would allow me to purchase 100 shares of C for much less than they are currently worth. Anyone who is interested in enjoying this great deal on shares of C would have good reason to offer to

[48] Here for simplicity's sake I focus on call options only and not put options, and also for simplicity's sake speak as if the stock option allows exercise before expiration date (whereas the Black-Scholes model in fact only describes European stock options, which can only be exercised at expiration date).

buy the option from me. Hence, stock options can be profitably traded: the option itself is liable to increase or decrease in value with changes in the price of the underlying stock.

How does the price of a stock option over time relate to changes in the underlying stock's price? In the 1970s, Fischer Black and Myron Scholes developed what is now known as the Black-Scholes model for options pricing. The equation at the heart of the model describes the time differential of the option value V as it relates to the price of the underlying stock S:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

It is well-known that the Black-Scholes model is a transformation of the heat equation from thermodynamics. If one looks only at the first two elements on the left-hand side of the equation (i.e., ignoring '…+ *rS*…, which provide boundary conditions) one can see that it shares its form exactly with the heat equation. The only difference is that the coefficient on the second-order is positive in the Black-Scholes model, while it is negative in the heat equation. This is because one uses the Black-Scholes model to work backwards in time from a final options price, while the heat equation describes the thermal evolution of a system forward in time. But the transformation of the Black-Scholes model into the heat equation is a well-practiced exercise in mathematical finance (Wilmott et al. 1995, Ch 5.4), where the heat equation is used to identify the Black-Scholes model's solution. That one equation solves the other is further evidence that they indeed are the same mathematical model.

Myron Scholes and Robert Merton were awarded the Nobel Prize in economics for the Black-Scholes model (Fischer Black was deceased). So it clearly has the support of the scientific community as a successful theoretical model of options pricing. And it is also widely used by participants in options trading; so it enjoys uptake beyond just the scientific community. But the philosophical intrigue of the Black-Scholes model—especially its relation to the heat equation from classical thermodynamics—has gone undiscussed. And this is a remarkable lacuna, since it is striking that the same formula that describes the dynamics of heat in a metal bar also characterizes the dynamical relationship between a stock call option's value and the value of its underlying stock. Part of the reason why this is so remarkable is—as in the case of the Ising model—how radically different these domains are. I know not what the ontology of stock options is. But I do not need to know what the ontology of a stock option is to know that they are a radically different sort of thing, and their price behavior a radically different sort of phenomenon, from hot metal bars and thermal conductivity. What in our ontology could account for this striking convergence in the models that characterize these systems?

2.3 Electrodynamics and Maxwell's Method of Physical Analogy

The last example of model transfer I shall offer here is a historical one, to demonstrate that the practice enjoys historical roots. Consider two distinct physical phenomena. First, consider a vortex in a fluid, the sort that you can easily create if you spin a closed bottle of water around while it is upside down. Second, consider the flow of an electric current through a conductive material (e.g., a copper rod). As the current flows, it generates a magnetic field around it.

James Clerk Maxwell, one of the two fathers of classical electromagnetic theory, used a model transfer to construct and derive the laws that describe electromagnetic currents from models of fluid vortex motion. Maxwell's method of physical analogy—that is what he called it—is famously exemplified all throughout his seminal paper "On Physical Lines of Force" (1861-2). The use of the mathematical analogy to vortex motion suffuses the entire paper, and to describe all the ways in which Maxwell uses the mathematics of vortex motion to derive the laws that characterize electrodynamic behavior would be nearly to reconstruct his entire four-part paper. So in a representative expository passage, Maxwell writes:

> "My object in this paper is to clear the way…by investigating the mechanical results of certain states of tension and motion in a medium, and comparing these with the observed phenomena of magnetism and electricity […] By making use of the conception of currents in a fluid… I propose now to examine magnetic phenomena from a mechanical point of view" (Part I, p162)

In short, Maxwell's derivations in "On Physical Lines of Force" are a model transfer from fluid mechanics to electromagnetism.[49] In an in-depth reconstruction of Maxwell's reasoning, Nancy Nersessian agrees: "In constructing the mathematical representation of the electromagnetic field concept, Maxwell created several models of an imaginary fluid medium drawing from the source domains of continuum [fluid] mechanics" (2002a, p144). And Maxwell's use of this transfer from vortex motion to construct the laws of electrodynamics was highly empirically successful, producing results consistent with observations known from experiment. For instance, Maxwell's

---

[49] Maxwell was not the only early scientist of electromagnetism to transfer to its study the mathematical resources of fluid mechanics. Hermann von Helmholtz writes:

> "[There is] remarkable analogy between the vortex-motion of fluids and the electro-magnetic action of electric currents. . . . I shall therefore frequently avail myself of the analogy of the presence of magnetic masses or of electric currents, simply to give a briefer and more vivid representation" (1858, p43)

Maxwell was aware of von Helmholtz's use of this technique, writing,

> "I have seen in Crelle's Journal for 1859, a paper by Prof. Helmholtz on Fluid Motion, in which he has pointed out that the lines of fluid motion are arranged according to the same laws as the lines of magnetic force, the path of an electric current corresponding to a line of axes of those particles of the fluid which are in a state of rotation. This is an additional instance of a *physical analogy*, the investigation of which may illustrate both electro-magnetism and hydrodynamics" (1861-2, fn12)

There are some differences between Maxwell's and von Helmholtz's use of the method of physical analogy. I have chosen here to focus on Maxwell's contribution, because of its convergences with observations and because it is more influential in the history of electromagnetic theory. Alisa Bokulich (2015) provides a detailed discussion of the differences between Maxwell's and von Helmholtz's contributions.

derivations from the vortex-fluid model converged with the known experimental results that electric and magnetic actions occur perpendicular to each other, that magnetism is dipolar, and that polarized light is rotated by magnetic force.[50] It is for good reason that "On Physical Lines of Force" has joined the historical record as a seminal and groundbreaking work in the history of electromagnetic theory. Maxwell's use of physical analogy in this paper (or model transfer, as we now should recognize) catalyzed the further development of classical electrodynamics and its mathematics. It is a remarkable success story of model transfer in the history of science.

Just like the other two cases of model transfer I have presented, it is a striking feature of Maxwell's that the two scientific domains in question are radically physically different. Magnetic fields and fluids are very distinct sorts of physical systems. The exertion of a magnetic field in an electromagnet is a very different sort of phenomenon than the creation of a spinning vortex in a bottle of water. In Maxwell's derivations in "On Physical Lines of Force," he most frequently uses the mechanics of motion, tension, and pressure in a vortex to model electromagnetic behaviors. But none of these sorts of properties actually obtain in an electromagnetic field; they are simply the wrong sort of physical entity. That the same mathematical model characterizes the dynamical behavior of both of these kinds of systems, just like we saw with the Ising model's transfer and the heat equation's transformation, calls out for explanation. What in our ontology could account for the success of these remarkable convergences?

3 Towards an Ontological Account of Successful Model Transfer

[50] Cf. Nersessian 2002a, p147.

3.1 A Puzzle

These examples show that there are cases of model transfer uncontroversially successful enough to warrant philosophical inquiry, and inquiry from more than an epistemic or justificatory stance. These examples share a common pecularity in the radicalness of the difference between their origin domains and their transfer domains. That the transferred models can generate empirically successful results given this disparateness in the kinds of objects and behaviors they originally characterize and those that they are transferred to describe seems remarkable. How can this success be possible, if the ontology and content of the model is radically different from that of the transfer domain?

In this section I move us towards an ontological account of these striking convergences. And I want to begin by motivating why we need one. The initial philosophical intrigue of these successful model transfers becomes more vivid when we remember a feature common to all of them which I mentioned at the beginning. These models are all broadly causal-dynamical in their content, and as a result can be used to answer explanatory questions regarding the systems they target. For example, we can ask an explanatory question regarding the diffusion of heat throughout a metal bar: why did this or that spot on the bar become hotter? The heat equation furnishes a causal explanation that cites the temperature of another region of the bar as the explanans-event. Or we can ask what explains the occurrence of a phase transition in a particular ferromagnetic system. The Ising model allows us to trace the evolution of the system over time, so that the phase transition is accounted for in terms of causal facts like the fact that heating up the ferromagnet causes its spins to flip more and more. Ditto for vortex motion and the exertion of force in a particular direction on the edge of the vortex's center. All of these models, in virtue of their causal and dynamical content, furnish the sort of information that can be cited in causal explanations of various events in their target systems.

Now, consider the common idea that to successfully give an explanation of some explanandum, one must correctly describe the underlying causal, non-causal, or ontological facts that are explanatorily relevant to it. Wesley Salmon was one of the first to explicitly endorse this idea under the banner of the ontic conception of explanation, writing that

> "what explains an event is whatever produced it or brought it about [...] The linguistic entities that are often called *explanations* are statements reporting on the actual explanation" (1990, p133)

And while Salmon's endorsement of the ontic conception is often mischaracterized as being a view on which all explanations are causal, clearly he does not think so.[51] In another illustrative passage, he writes:

> "Explanations tell us how the world works. [Mechanistic] explanations are local in the sense that they show us how particular occurrences come about; they explain particular phenomena in terms of collections of particular causal processes and interactions—or, perhaps, in terms of non-causal mechanisms, if there are such things" (1990, p184)

Both of these widely-cited passages rely on the idea that for an argument, model, or other sort of representation to actually successfully qualify as giving an explanation for some explanandum, they must get the facts right. They must report on the explanatory dependence-structure of the physical system—the "actual explanation"—or the particular causal processes that led up to the explanandum, and so on. Thus Michael Strevens also writes that "an

[51] See also Jansson and Saatsi (2017), Povich (2018), Craver and Povich (2017), among others, on not conflating the ontic conception of explanation with the view that all explanations are causal.

explanation in the ontological sense is a collection of explanatory facts, whereas an explanation in the communicative sense is a representation of such facts" (2007, p319). And James Woodward also says:

> "[A] widely shared idea about explanation…is that successful explanation is at bottom a matter of getting fundamental ontology right; on this view, an explanation can't be satisfactory if it doesn't tell us what the basic objects or constituents are that make up the systems whose behavior we are trying to explain. According to this view, the statistical mechanics of gases is a satisfactory explanatory theory because it tells us what sorts of entities gases are made up of or composed of; by contrast, the generalizations of phenomenological thermodynamics have little or no explanatory import because they are noncommittal or unrevealing regarding the basic constituents of gases or compatible with fundamentally mistaken assumptions regarding those constituents." (Woodward 2003, p223)

This all is what I take to be the kernel of the ontic conception of explanation.[52] And it is clearly not a hard view to get behind. The idea is simple enough: to give an explanation of some physical event, you have to represent it accurately, or at least accurately represent what is ontologically relevant to its occurrence.[53]

[52] There has been some controversy lately on whether the ontic conception is, as I have stated it, the view that explanations must *represent* the ontological dependence-structure of the explanandum (see, e.g., Woodward 2003, Povich 2018, Craver and Povich 2017), or as the view that explanations are "full-bodied things in the world" (see Bokulich 2016, Illari 2013, Wright 2015 for criticism of *this* view). I accept the first formulation and also broadly sympathize with criticisms of the second formulation, but the details are unimportant here.

[53] There is of course a voluminous literature on reconciling the presence of idealizations in scientific models with models' abilities to provide factive explanations. See Bokulich (2007), Strevens (2007), Psillos (2011), Batterman and Rice (2014), among many others, for discussion.

Thus, recalling the explanatory content of the models we considered in the previous section, there is a *prima facie* puzzle when one considers the cases of their transfer that we have seen. It is a datum of our scientific practice that the transferred models are considered and used explanatorily in their transfer domains. And the fact that their transfer generates remarkable empirical successes seems to be evidence that they latch onto real causal structure in their transfer domains; after all, it is causal structure that generates observation and measurement! But again, these models' contents—the sorts of objects and properties they describe—are radically different from the objects and properties that constitute their transfer domains. How, then, if the ontic conception of explanation is correct, can these models be explanatory in their new domains at all? They get the ontology of their transfer domains radically wrong.

This point gains more force when we consider models not simply as representational devices that have descriptive and representational content, but as miniature worlds of their own. It is familiar within the philosophical literature on scientific modeling to conceive of the relationship between model and world as a relationship, really, between one system and another: model system and target system. Idealized scientific models, after all, come prepackaged with ontologies and physical laws all their own (Cartwright 1983, Sugden 2000, Suarez 2002, Contessa 2007). The representational relation from model system to target system is therefore conceived of variously as one of isomorphism (van Fraassen 1992, 1994), or qualitative similarity (Giere 1988, 1999), or inferential surrogacy (Suarez 2004, Contessa 2007, Kuorikoski and Ylikoski 2015); in all of these views, the model is conceived of not as having content in a merely representational way, but as a world unto itself. So the challenge that these cases of explanatory model transfer pose is not simply that the models' content represents their target systems in their transfer domains wrong. If the world of a model contains, e.g., idealized spins or an idealized metal bar, one can see how that might be the right sort of system to map onto non-ideal spins or non-ideal metal bars. But it is hard to see how there could be the same sort of mapping to non-ideal neurons or non-ideal stock options. In other words, considered as worlds unto themselves, the ontology of the models simply seem like they are of the wrong sort to facilitate any sort of representational connection to their transfer domains.

3.2 A Solution

So, we have two distinct philosophical questions regarding these successful model transfers. First, what explains how two radically different physical systems can be successfully characterized by the same model? Second, how can the same model provide explanatory information about both systems, if explanation is a matter of ontology?

You are likely thinking that there is a simple and obvious solution that answers both of these questions, that does not sacrifice the attractive and plausible judgment that they manage to do so—and manage to produce empirical successes—by latching onto real causal structure in both domains and by somehow being correct regarding ontology. Here's how the idea likely goes. When a model is transferred to a domain different from its origin domain, obviously we do not transfer wholesale all of the content of the model, ontology and all. Instead, what is transferred is a version of the model that is abstracted away from its target-specific ontological content, but which still preserves the sort of causal processes that the model characterized in its original domain. So, insofar as it is the same model that characterizes the two radically different physical systems, it is the model considered at a higher level of abstraction than specifics about concrete ontology. In other words, it is the model's higher-level, more abstract content that characterizes both systems.

This is probably the most intuitive solution to both of the puzzling features of model transfer. You might say: it simply seems obvious that if the same model successfully characterizes physical system A and physical system B, and those two systems contain very different kinds of objects and properties, then how the model does so must be by characterizing them at a level of description that does not represent the systems as containing particular kinds of objects and properties. Likewise, you might think it is simply obvious that cases of model transfer can still be provide veridical causal explanation. For even if they characterize the *things* in the system wrong, the transferred models are considered at a level of description where they still characterize the system's *processes* correctly. On this simple and intuitive view, transferred models still manage to tell us about what there is in their new systems, not by telling us which objects and properties there are, but by telling us *what happens*.

This solution is also corroborated by a plausible interpretation of what happens in scientific practice when models get transferred. It should be rather obvious that when we actually transfer a model from one domain to another, we abstract away from the particular denotations that the model originally had, and end up substituting them with new referents. Indeed, this is exactly how Maxwell's historical transfer case worked:

> "The modeling process Maxwell used throughout [his] analysis went as follows. First he constructed a model representing a specific mechanism. Then he treated the dynamical properties and relations generically by abstracting features common to the mechanical and the electromagnetic classes of phenomena. He proceeded to formulate the

> mathematical equations of the generic model and substituted in the electromagnetic variables" (Nersessian 2002a, p147)

And this is what we should expect: in order to get a model of one kind of system to work as a representation of another kind of system, you need to consider its higher-level content. But what is important is still the idea that there *is* some higher-level, more abstract content to the model, where the model ultimately has content where it represents "*types* of entities and processes that can be considered as constituting either domain" (ibid). On the natural solution being offered here, this high-level, abstract content is not only what facilitates the model's transfer from one domain to another, it is what explains the model's success in characterizing both systems. In other words, the model ends up characterizing both systems *because* they share in these "generic" causal processes.

I endorse this answer, and I shall defend it further in a few moments. And I agree that it is intuitively plausible and even seems obvious. But it should not be taken lightly on the grounds that it seems simple or obvious. For suppose you think that the empirical successes of our legitimate scientific practices occur because our sciences latch onto ontology, and you are presented with cases where one causal-dynamical model, applied to two apparently ontologically disparate systems, manages to be successful in describing both of them. Then *of course* you should think that there is in fact something ontologically shared between the otherwise disparate systems that underlies this success. In other words, if you are a scientific realist, and believe that in some broad sense there are no miraculous successes in legitimate science, then the idea that generic causal processes explain successful model transfers ought to be one of the very first options on the table. Indeed, as far as I can see, it is the *only* scientific realist response on the table. In model transfers, the success to be accounted for is simply that a model built to represent and detect causal and dynamical structure appears to detect the same structure in two different systems. Any adequate ontological account of this success—that is, any adequate realist account—must therefore countenance something sufficiently similar between the two systems to explain this sameness in what represents them. And generic causal processes are the precise sort of thing apt to do so.

To say that an ontology of generic causal processes is the best and only realist option on the table is not to say that the case for them cannot be strengthened beyond appeal to underlying intuitions about the successes of science. The case for them can be strengthened, and I shall do so. However, more important right now than any strengthened defense is the question of exactly what it is that generic causal processes would be: we must clarify what we are asked to believe in before evaluating whether belief is indeed warranted. It turns out that the idea that there are "generic causal processes" is quite radical ontologically speaking, and it turns out to have far-reaching metaphysical consequences. So, before strengthening and clarifying the case for generic causal processes, I shall first articulate exacty why this view is ontologically novel.

4 The Ontological Novelty of Generic Causal Processes

The novelty of an ontology that includes these generic causal processes is that it requires a reworking of how causal processes are typically conceived. In every extant discussion of casual processes, they are individuated both at the token level and at the type level by particular material realizations. To begin with, I take it that there is a naturally plausible view of what causal processes are where they are ordered collections of causal events. Thus Jonathan Schaffer says that "processes are lawful [causal] sequences" (2001, p77-8). Likewise Christian Loew writes that "causal processes are sequences of events that lead toward an outcome" (2019, p93), and on his view, it is simply a part of how processes are individuated that they have the same material constituents:

> "Two causal processes are (cross-world) identical, just in case they have: (i) the same form, and (ii) the same constitutive material parts. [...] Processes have a main actor (or set of actors) that carry out or undergo the process. These objects are constitutive for the process such that identical processes need to have the same constitutive parts" (ibid, p93-95)

It is clear that in these formulations of the idea of a causal process, they are constituted at the material, things-bumping-into-each-other level of the world where there is real, material, causal *oomph*. And this is what we should expect, if causal processes are just part of the physical order of the world. Where it is that physical laws and relations hold—i.e., where the *oomph* is—there too shall we find the processes.

This underlying part of the natural view of causal processes is preserved in sophisticated philosophical treatments of what a causal process is, too. For instance, the *locus classicus* of causal process views has for some time been Wesley Salmon's, in which a causal process (as opposed to a pseudoprocess) is a chain of causal interactions in which a "mark" is transmitted from object to object (1984, p147). A mark, in Salmon's understanding, is an alteration to a characteristic of an object that is introduced by a single local interaction between concrete objects. On this view, the material constituents of types of causal processes must be the same, since what it is that individuates a causal process is that it transmits a mark and how it does so; marks are tied to the objects that bear them. Or consider Philip Dowe's conserved-quantity characterization of a causal process: a causal process is a "world line" of some concrete object—a four-dimensional history of the object—which manifests some conserved quantity (1992, p210).[54] Just like on Salmon's view, Dowe's "world line" characterization ties the realization of a process to some particular object (or, we can easily suppose, a type of object). Douglas Ehring's (1998) view of processes holds that processes are those causal interactions in which the selfsame property instance, i.e. trope, persists; at the type-level, causal processes are individuated by the particular property-types they pass along. Lastly, on Michael Strevens's criterion of process individuation, two causal processes are of the same type only if their microphysical realizers are such that "you can get from any realizer to any other realizer by a process of minimal tweaks" (2012, p498). In all of these views, it is clear that what individuates causal processes are its particular constituent material object- and property-types. A difference in material constituents or types of properties involved means a different causal process. This underlying individuation principle is not hard to get behind; again, it reflects a naturally plausible view of what causal processes are.

But if the physically disparate systems in these model transfer cases exemplify the same causal process, then this widely-held and naturally plausible view of the individuation of causal processes is false. For if two radically physically disparate systems in nature can exemplify the same causal process, then there are some kinds of causal processes that are radically multiply realizable. That is simply what the suggested solution to understand these model transfer cases entails: that the same dynamical process happens in both, e.g., a ferromagnetic phase transition and the firing of some neurons in response to a stimulus. The physical constituents of what is alleged to be the same process are very distinct, both in terms of which types of objects there are and which property-types there are.

[54] Dowe's view draws from David Fair's (1979) "transfer of energy" view of processes, which I do not mention here for this reason.

What individuates causal processes, if not their material constituents? That is, in virtue of what is it true that ferromagnetic phase transitions and firing ganglion-cell clusters exemplify the same causal process? It seems to me that the appropriate answer to this question must just be that they just *do*; there is nothing in virtue of which it happens. There simply is a kind of causal process that is the one that both of those systems exemplify. For it is not as if there are laws of nature that govern both neurobiological behaviors considered as such and ferromagnetic behaviors as such. So it is not as if we can appeal to the similarity of the pairwise causal relations at any successive stage in the material constituents to explain why they are the same process. Maybe you think we can appeal, as Loew does, to the idea that causal "processes [have] a form such that they can lose or gain some of their parts as long as their form stays the same. The form of a process…consists in the way in which it unfolds toward the outcome" (2019, p94). But this suggestion still pushes us towards the idea that the fact that the two model transfer systems share the same causal process is fundamental. After all, to make sense of generic causal processes in model transfer, the idea of the "form" of a process must be divorced from material realizers or properties like energy and the causal role individual events play, contra Loew's account (ibid, cf. Steward 2013). To explain the criteria under which the origin and transfer domains in a model transfer exemplify the one kind of causal process's form, we would have to say that their processes have the same form despite their radical physical disparities. But this implies that there are facts in our ontology about which causal-process-forms are which, that are independent of material realization; i.e., there are facts about what the bare, unrealized form of this or that kind of causal process is. It must be these "bare" forms that the generic causal processes exemplified in successful cases of model transfer correspond to.

Therefore, the puzzles posed by model transfer push us towards the notion that there are irreducibly higher-level, explanatorily fundamental multiply-realizable kinds. In these successful and legitimate model transfers, we discover that there is a metaphysical unity in the kind of systems involved: they share in their causal process-kind. Or maybe, following the previous thought, we can say that their causal processes share their high-level "form" (a phrase which also recalls ontic structural realism; more on this later). However we choose to describe the result, it is deeply surprising: some ontologically fundamental scientific kinds are *by nature* "generic," high-level, and radically multiply-realizable.

Additionally, this shows how the multiple realizability here is different in kind from the multiple realizability present in other areas of the sciences. In every canonical case of multiple realization in the sciences, it is possible to identify some mid-level physical property or process predicated of the lowest-level physical realizers that explains why the property is instantiated in some case. This is because every canonical case of multiple realization is one where the property in question is functionally described. To take an example from Fodor's (1974) classic paper on multiple realizations, the property of *being money* seems to be radically multiply realizable by, e.g., paper, coins, wampum shells, etc. But it seems clear that these instantiations are all explained in terms of the functional role of money and the underlying mental-cum-social disposition to treat coins, wampum, etc., as being able to be exchanged for goods and services. Hence, in ordinary cases of multiple realizability, the multiply-realizable property is—despite being autonomous in many ways from its realization-properties—still not explanatorily or ontologically fundamental.

By contrast, if there are generic causal processes that the origin and transfer domains in model transfer cases both exemplify, that they both do so is not apt to be explained in these ways. To begin with, and most importantly, the causal process is something that occurs in each of these whole systems, considered in isolation. So it is not the case that there is a functional or causal role that the whole system plays that could be the underlying explanans for their containing some process or another. Similarly, the radical differences between systems make it impossible to identify the sort of single physical or causal property that explain why ordinary multiply-realizable properties are instantiated. Again, the sort of ordinary causal behaviors a hot metal bar engages in are about as far as I can imagine from the sort of behaviors stock option engage in. Perhaps there is a very high level of abstraction at which they are the same: indeed, perhaps successful model transfers are evidence that physically disparate systems have the same phase-space structure.[55] But even this sameness would not provide an explanation for *why* the disparate systems contain the same causal process (the phase dimensionality of a system is not, after all, an underlying causal or functional role), it would only be a further implication of the discovery that they do in fact contain the same process.

So the sorts of generic causal processes that, as has been suggested, are exemplified in and explain the successes of legitimate model transfers are different in kind from other kinds of multiple realizability across the sciences. As I have argued in this section, if there are generic causal processes, there are facts about which are which, and what they are like, that are totally independent of material realization. In other words, if there are generic causal processes, they are ontologically and explanatorily fundamental.

5 Strengthening the Case for Generic Causal Processes

[55] See Andersen (2017) for a characterization of causation as transfer of information in phase space; her view is highly compatible with what I say here about the ontology of generic causal processes.

Thus far I have argued for generic causal processes on the simple and intuitive basis that they appear to be the best and perhaps only sort of thing that can give an ontological account of our cases of successful model transfer. And now we have seen exactly what an ontological commitment to them would entail. Now I wish to show that the case for generic causal processes can be strengthened beyond appeal to underlying scientific realist intuitions about the empirical successes of our legitimate sciences.

A closer look at what underlies the process of model transfer, and what sort of reasoning and physical principles are relied on in a successful transfer, furnishes two additional reasons to believe in generic causal processes. Let us recall the idea that in order to transfer, e.g., the Ising model to another domain, it clearly must be the case that what is transferred is not the whole model, kitchen sink and all. Instead, the process must involve some sort of abstraction that preserves the core dynamical content of the model while still removing it from the particular denotations and referents the model had in its origin domain. This is part of what the impetus was, in the first place, for thinking that generic causal processes provide a plausible and attractive ontological account of how model transfer works.

I wish to call attention to the way in which the abstraction process serves a justificatory role in the transfer of the model, and the way in which this abstraction is responsible for generating the model's predictive content in its new domain. For consider what Nancy Nersessian says in two passages regarding Maxwell's model transfer. First, she says:

> "Maxwell was able to formulate the laws of the electromagnetic field by abstracting from specific mechanical models the dynamical properties and relations that continuum-

> mechanical systems, certain machine mechanisms, and electromagnetic systems have in common. In their mathematical treatment these common dynamical properties and relations were separated from the specific instantiations…through which they had been rendered concrete. […] [This] mathematical structure…represented causal processes in the electromagnetic medium without requiring knowledge of specific causal mechanisms" (2002a, p159)

And again writes:

> "Maxwell took a fluid-dynamical system as his analogical source and made certain idealizing assumptions…then treated processes in the fluid generically, e.g., as "flow", rather than "fluid flow", and abstracted mathematical relationships into which the electromagnetic variables could be substituted" (2002b, p131)

Nersessian's exposition readily calls to mind the idea of generic causal processes we have been working with. For now, I want to focus on a particular aspect of the model transfer process here, though: what role does the analogical source and the idea of, e.g., "flow" actually play in transferring the model?

The role that the abstraction process plays here is that it in fact generates the predictive content of the model as it is taken from its origin domain and applied to its transfer domain. This can be seen not only in Maxwell's case, but in the other two cases as well. The content of Maxwell's derivations, in the electromagnetic context, are justified and fueled by the idea that there is a kind of behavior that we might simply call "flow". There is a counterfactual supposition to the effect of: *if flow were present in electromagnetic fields*, we would expect to see thus-and-so. Likewise, in the case of the Black-Scholes model, the derivation of the model is motivated by the idea that the underlying stock price follows the sort of statistical behavior known as a Brownian motion; Brownian motions satisfy the heat equation. And so, too, in the case of the Ising model's transfer: the driving assumption is that meso-scale correlations of a micro-scale property determine its macro-scale manifestations, not individual instantiations. (We simply substitute 'neuron firing' for 'spin flipping'.) In each of these cases, once the model has been interpreted in terms of the particular objects and properties in its transfer domain, it generates new predictions about how those objects and properties will behave. But it only generates these because of underlying justificatory principles regarding what this-or-that kind of behavior looks like in general. And where these principles come from are the previously, empirically-observed behaviors of the origin domain.

What this means is that when a model is transferred to its new domain and its predictions actually come to be observed, there is a confirmation of the generic behaviors' presence in that domain. Or here is a better way to put it: observation of a transferred model's predictions is evidence that the model's justificatory principles in that domain have been causally *detected*. After all, what is it that fueled those predictions in the first place? Again, the answer is: generalizations from behaviors that were observed in the origin domain.

This powerfully augments the case for generic causal processes. To begin with, notice that in order to draw or justify a particular prediction by appeal to the observed behaviors of, e.g., fluid flow, but in a system that contains no fluids, one must do what Nersessian suggests in the passage quoted above: treat the observed behavior in question as not essentially fluid-based. But what this amounts to is simply agreeing that there simply is a broader type of behavior or process, *flow*, that admits of many determinations. And to agree that there simply is this broader type of behavior, that is the way things flow, without reference to what it is that is flowing, is to endorse genuine differences in kinds of behavior that are divorced from material realizations. In other words, it means endorsing essentially what I have called generic causal processes.

A second reason this augments the case for generic causal processes is that, as we have seen, there is good reason to think that it is possible to detect when something genuinely does or does not instantiate this broader generic type. This follows from the simple fact that these assumptions about how *flow* (for example) looks, regardless of material realization, can generate predictions in this, that, or any other domain. While it is serendipitous that Maxwell was right about the analogy from fluid flow to vortex flow, I take it that it is sufficiently obvious that if we were to use those same "flow" principles to generate predictions about how a bushel of apples would behave, none of our predictions would be satisfied.

Thus the case for generic causal processes is just another case of selective scientific realism about this or that kind of entity. Realism can and should be selective in the sense that scientific representations contain a great deal of information which whose reality is explicitly worth disavowing, e.g., those of idealized frictionless planes, perfectly continuous fluids, and so on. To identify what it is that we actually should be realists about in our theories and practices, it has variously been suggested that we distinguish between what is essential and what is idle in generating predictions in our theoretical reasoning (Kitcher 1993, Psillos 1999), or that we identify what causally underlies what can ends up being detected and measured (Cartwright 1983, Chakravartty 1998, 2007). The case for generic causal processes satisfies this sort of strategy, broadly construed. For the justificatory reasoning that is involved in transferring a model essentially involves appeal to what are generic causal processes, viz., facts about the generic structure or form of flow, or of Brownian motion, etc. And it is the appeal to these generic facts that allows for substitution of variables in a model's new domain, e.g., from electromagnetic variables to neurobiological ones, and which accordingly drives the predictions. That the predictions driven by appeal to generic causal processes is the same sort of detection of underlying ontology or causal structure that fuels selective realism about many other sorts of entities. The upshot of all of this is that there is a powerful scientific realist case for generic causal processes which goes beyond simple appeals to intuition. There is good reason to think that model transfers are cases in which generic causal processes, if they really exist, would be detected; and that there are successful model transfers is good evidence that we have in fact detected generic causal structure. I conclude on the basis of all of this that there are generic causal processes.

6 Broader Implications

6.1 Valediction

I have argued for the reality of generic causal processes because they are the best and only sort of thing that can make sense of the datum of successful model transfers between sciences. And I have shown how the case for generic causal processes can be supplemented by careful attention to the reasoning processes that are involved in actually transferring a model (or justifying its transfer). We thus have both strong *prima facie* reasons to be realists about generic causal processes, and strong theoretical precedent as well. This ontological commitment, I have also shown, is surprisingly radical. Although it comes off the tongue easily as an account of why model transfers are successful, endorsing it in fact requires agreeing that some ontologically fundamental kinds are by nature multiply realizable. This by itself is a surprising and novel conclusion. But I am not finished yet. Here, in this final section, I show that realism about generic causal processes yields further metaphysical implications.

6.3 A New Face for Ontic Structural Realism

One of the most surprising implications of the reality of generic causal processes is that it implies a sort of ontic structural realism. Or to put it more precisely, realism about generic causal processes implies realism about the same sort of things that ontic structural realists urge us to be realists about.

Let me rehearse the idea of structural realism. Many years ago, Larry Laudan brought a blight upon the peaceful Land of the Realists in the form of the pessimistic metainduction: the empirical successes of science do not warrant belief in the entities or properties present in our theories, since past theories were empirically successful despite lacking successful reference. The structural realist's strategy is to deny that there is nothing that successfully refers in these past theories: past theories were successful because, even if they were incorrect about which objects and properties nature contains, they were correct about the sort of *structure* in nature. Indeed, one of the core motivations of structural realism is that the mathematical structures of the successful parts of superseded theories are preserved in their contemporary descendants. This now-famous passage of John Worrall's puts this point exactly:

> There was an important element of continuity in the shift from Fresnel to Maxwell—and this was much more than a simple question of carrying over the successful empirical content into the new theory. At the same time it was rather less than a carrying over of the full theoretical content or full theoretical mechanisms (even in "approximate" form)

> … There was continuity or accumulation in the shift, but the continuity is one of form or structure, not of content. (1989, p117)

Because the structure of our theories is preserved, structural realists say, we can explain the success of both past and present science in terms of their structure. And if so, then we should be either epistemically or ontologically committed to the structure of nature that our theories ascribe to it. Epistemic structural realism, which I will not discuss here, is the view that the beliefs in nature that our theories warrant amount to the abstract, structural information they contain. Ontic structural realism, by contrast, is the view that we are warranted in believing in the *reality* of structures, whatever they turn out to be.

This is the kernel of structural realism. There has been voluminous discussion over exactly how to characterize structural realism, over whether it is well-motivated to begin with, and what its ontology looks like, exactly (see Ladyman 2014 for a survey). I have no contribution to this particular voluminous discussion, but what I wish to discuss is that the reality of generic causal processes amounts to ontic structural realism of a certain kind. For notice how Worrall, in the famous quote, characterizes what the structure of a theory is: it is the "form" of the characteristic equations. That is what is alleged to be preserved in historical episodes of theory-change.

We have seen that the idea of a generic causal process amounts to a realist interpretation of the "embodied form" of a model's characteristic dynamical equations. Indeed, the idea of a causal process directly implies that there is a reality to the form of a process that is preserved between differences in material realization. A belief in the reality generic causal processes is very much a belief in the same sort of thing that ontic structural realists allege is preserved during theory change. There is, so they say, the form or structure of nature that one theory told us about, and it remains present even in another theory that contains a radically different ontology. Their conclusion is that we ought to believe in the reality of these forms. And I have argued for the reality of the forms of dynamical equations on entirely different grounds.

Hence, realism about generic causal processes constitutes realism about the same sort of thing that ontic structural realists ask us to be realists about. To be clear, generic causal processes do not offer us an *argument* for ontic structural realism, in the way that defending structural realism as a theory requires careful attention to the history of science. Rather, realism about generic causal processes amounts to an ontological commitment to (a version of) the ontology of structuralism, and it is surprising that such an ontological commitment can be generated without direct appeal to typical structuralist reasoning.

It is still open exactly as to what the ontic structural realists, defending realism in the historical sense, might urge us to believe. There are an incredible variety of options. Some ontic structuralists have defended *eliminative* structural realism, a view on which there is nothing *but* relational structure (French and Ladyman 2003, 2011). Some defend the view that individuals have no intrinsic properties (French 1999). In general, structuralists have encountered a great deal of difficulty in trying to explain exactly what they mean by "structure." By contrast, realism about generic causal processes gives us a sort of realism about structure that carries its own clear ontological implications: there are facts about types of causal processes and their behaviors that are independent of any material realization. There still are individuals and their properties; after all, they are what are apt to be the material instantiators of these generic types. So insofar as generic causal processes commit us to the same sort of thing that ontic structuralists endorse, we can see that they commit us to a *non-eliminative* version of structuralism. In any case, the upshot is twofold. First, realism about generic causal processes implies a realism about the same sort of ontology that is defended by ontic structuralists, and it also manages to do so while spelling out exactly what the ontology of generic causal processes might look like.

6.4 A New Conception of Laws of Nature

A second implication of endorsing an ontology of generic causal processes is that it implies a sort of primitivism about the laws of nature. For the characteristic dynamical equations of our cases of model transfer are the expressions of the laws of nature that govern (at least) the origin systems. The heat equation, for example, seems clearly to be a law of how heat diffuses, and when its variables are substituted in content to denote (e.g.) particulate diffusion, then it manages to state the nomic facts about that process, too.

As we have seen, however, if there are generic causal processes, then the mid-level interpretations of these characteristic dynamical equations, which are separated from any particular physical interpretation, also manage to describe dynamical patterns in nature. Indeed, insofar as the particularized versions of these characteristic equations express the laws that govern those particular systems, we should think too that the generic form of them also express laws of nature. And if so, then that is just to say that generic causal processes enjoy the status of being causal laws of nature.

Generic causal processes are ontologically fundamental and the facts about which ones are which obtain independently of material realization. Similarly, the facts about which systems they can and cannot be found in are ontologically fundamental and obtain independently of material realization. So, if generic causal processes are laws of nature, then there are laws of nature the lawhood of which and the existence of which do not supervene on physical facts. For example, if generic causal processes are individuated independently of material realization, then a difference in what the "Humean mosaic" (Lewis 1973, 1986) looks like would not make a difference to the facts about which kinds of generic causal processes there are. So, there would be some laws of nature—the generic causal processes—that are not analyzable in Humean terms. Likewise, a difference in which dispositional properties are concretely instantiated (Bird 2007, Chakravartty 2007) would not necessitate a difference in which kinds of generic causal processes there are, nor would a difference in the relations between particular kinds of physically-instantiated properties (Dretske 1979, Armstrong 1983). So generic causal processes would not be analyzable in the mainstream anti-Humean terms, either. This amounts to a kind of primitivism about the laws; of course, it need not be the case that *all* laws of nature fail to non-trivially supervene on physical facts. But if there are generic causal processes, it seems clear that *some* laws are ontologically primitive.

Primitivism about the laws of nature has some recent philosophical precedent, e.g., in Timothy Maudlin's work:

> "Nothing in scientific practice suggests that one ought to try to reduce fundamental laws to anything else. Physicists simply postulate fundamental laws, then try to figure out how to test their theories; they nowhere even attempt to analyze those laws in terms of patterns of instantiation of physical quantities." (Maudlin 2007, p105, original emphasis)

Still, there has been very little uptake of primitivist views of the laws. If there are generic causal processes, however, there are powerful reasons to believe in the ontological primitiveness of at least some causal and dynamical laws. It is surprising that an ontological account of how a particular kind of modeling practice manages to be successful carries this implication.

6.5 A New Conception of the Metaphysical Unity of the Sciences

I shall close with one last implication of my arguments in this paper. There is an old story according to which the sciences are unified as follows: science is arranged hierarchically, with fundamental physics at the explanatory base, chemistry above it, biology next up, followed by the psychological, behavioral, and social sciences. The degree to which science is epistemically unified is a matter of there being truth-preserving representational or inferential connections between these levels. And the degree to which the subject matters of these sciences are *metaphysically* unified is a matter of the extent to which events of one scientific kind can be explained in various ways by events in another, typically lower, scientific domain.[56] On one typical way of thinking about the metaphysical unity of the sciences, the sciences are metaphysically unified to the extent that physical-science properties and events are fundamental and explanatory of all the rest.

If there are generic causal processes, ontological reductionism to the physical sciences is not possible. For if there are such processes, then—as we have seen—they are scientific kinds which are ontologically fundamental, and the facts about which ones there are do not depend on material realization. As it turns out, the arguments against reductive physicalism that generic causal processes provide are distinct from previous multiple-realizability arguments against

[56] See Cat (2017) for a survey of this literature.

reductive physicalism, even though generic causal processes would by nature be multiply realizable. The idea that multiply realizable properties refute reductive physicalism is well-rehearsed: if some property-type is multiply realizable across a variety of physical properties, then there is no one physical property-type that it could be identical to, and if so, then reductive physicalism is false.[57] Physicalists respond to these arguments in a variety of ways. For example, insofar as the allegedly multiply-realizable properties have functional instantation criteria, one can always identify underlying physical properties or causal roles that both undermine the appearance of genuine multiple realizability and support a type-type identity claim (Shapiro 2000). Or physicalists can simply deny that there is genuine unity in the alleged higher-level properties (Churchland 1986, Bechtel and Mundale 1999). So multiple-realizability arguments' ability to rebut ontologically reductionist views is, despite their wide uptake, still quite uncertain.

However, as we have seen, it is not anything about the *multiple realizability* of generic causal processes that motivates the idea that they are ontologically fundamental and not reducible to physical properties It was issues regarding the individuation and occurrence-conditions of generic causal processes that pushed us towards the idea that they are ontologically fundamental. I have discussed how, because generic causal processes are not functional properties and are not functionally defined, their exemplifications in physical systems are not apt for explanation in terms of underlying physical properties. Moreover, it is not an open option to deny that there is genuine unity in which causal models apply to which systems: it is empirically confirmed, remember, that there is an "exact mapping" of the Ising model to ganglion-cell clusters. Thus the reality of generic causal processes provides us with an argument

[57] Fodor (1974), Antony (2003), Funkhouser (2007), among many others.

against reductive physicalism that is immune to established objections against multiple-realizability arguments; indeed, their irreducibility is of a fundamentally different kind than the kind that functional properties are alleged to enjoy.

The reality of generic causal processes refutes ontological reductionism. So it undermines the prospects for the sciences to be metaphysically unified in the traditionally-conceived way. But if there are generic causal processes, there is an entirely new kind of metaphysical unity to the sciences. After all, if there are generic causal processes, then radically physically disparate systems can contain one and the same kind of causal process. This is not interlevel, reductive unity *per se*: rather, this tells us that there are patterns in nature that span across a variety of scientific domains. These unifying patterns are what underlie the ability of certain sorts of models to latch onto causal processes across domains. This sort of metaphysical unity is, perhaps, not what most people sign up for when they go searching for order and unity in nature. But it is striking and remarkable all the same. And it is even more striking and remarkable that we can, on occasion, detect and discover these metaphysical unities through model transfer.

Works Cited

Adams, Ernest W. (1993). On the rightness of certain counterfactuals. *Pacific Philosophical Quarterly* 74 (1):1-10.

Aldus, R., Fennell, T., Deen, P.P., Ressouche, E., Lau, G.C., Cava, R.J., Bramwell, S.T. (2013). Ice rule correlations in stuffed spin ice. *New Journal of Physics*, 15(1):013022.

Andersen, Holly (2017). Patterns, Information, and Causation. *Journal of Philosophy* 114 (11):592-622.

Antony, Louise (2003). Who's afraid of disjunctive properties? *Philosophical Issues* 13 (1):1-21.

Armstrong, D. M. (1983). *What is a Law of Nature?* Cambridge University Press.

Armstrong, D. M. (1996). *A World of States of Affairs.* Cambridge University Press.

Baron, Sam; Colyvan, Mark & Ripley, David (2017). How Mathematics Can Make a Difference. *Philosophers' Imprint* 17 (3).

Batterman, Robert W. & Rice, Collin C. (2014). Minimal Model Explanations. *Philosophy of Science* 81 (3):349-376.

Bechtel, William and Jennifer Mundale, 1999. "Multiple Realizability Revisited: Linking Cognitive and Neural States," *Philosophy of Science*, 66: 175–207.

Beebee, Helen (2000). The non-governing conception of laws of nature. *Philosophy and Phenomenological Research* 61 (3):571-594.

Bhogal, Harjit (2017). Minimal Anti-Humeanism. *Australasian Journal of Philosophy* 95 (3):447-460.

Bird, Alexander (2005). The dispositionalist conception of laws. *Foundations of Science* 10 (4):353-70.

Bird, Alexander (2007). *Nature's Metaphysics: Laws and Properties.* Oxford University Press.

Bokulich, Alisa (2008). Can classical structures explain quantum phenomena? *British Journal for the Philosophy of Science* 59 (2):217-235.

Bokulich, Alisa (2011). How scientific models can explain. *Synthese* 180 (1):33 - 45.

Bokulich, Alisa. (2015). Maxwell, Helmholtz, and the unreasonable effectiveness of the method of physical analogy. *Studies in History and Philosophy of Science Part A*, *50*: 28–37.

Borghini, Andrea & Williams, Neil E. (2008). A dispositional theory of possibility. *Dialectica* 62 (1):21–41.

Bradley, Seamus, & Thébault, Karim P. Y. (2018). Models on the move: Migration and imperialism. *Studies in History and Philosophy of Science Part A*.

Bramwell, S.T., Giblin, S.R., Calder, S., Aldus, R., Prabhakaran, D., Fennell, T. (2009). Measurement of the charge and current of magnetic monopoles in spin ice. *Nature*, 461(7266):956–U211.

Brogaard, Berit and Salerno, Joe (2007). "Why Counterpossibles Are Non-Trivial." *The Reasoner* 1:5-6.

Brogaard, Berit and Salerno, Joe (2013). "Remarks on Counterpossibles." Synthese 190:639–660.

Carroll, John W. (1994). *Laws of Nature*. Cambridge University Press.

Carroll, Sean (1994). *Spacetime and Geometry: An Introduction to General Relativity*. Addison-Wesley.

Cartwright, Nancy (1983). *How the Laws of Physics Lie*. Oxford University Press.

Castellane, Ariane, & Paternotte, Cédric. (2018). Knowledge transfer without knowledge? The case of agentive metaphors in biology. *Studies in History and Philosophy of Science Part A*.

Castelnovo, C., Moessner, R., Sondhi, S.L. (2008). Magnetic monopoles in spin ice. *Nature*, 451:42–45.

Cat, Jordi (2017). The Unity of Science. In *The Stanford Encyclopedia of Philosophy*.

Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., & Chatterjee, A. (2013). *Econophysics of income and wealth distributions*. Cambridge University Press.

Chakravartty, Anjan (1998). Semirealism. *Studies in History and Philosophy of Science Part A* 29 (3):391-408.

Chakravartty, Anjan (2003). The dispositional essentialist view of properties and laws. *International Journal of Philosophical Studies* 11 (4):393 – 413.

Chakravartty, Anjan (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press.

Chaplin, M. (2007). Water's hydrogen bond strength. *Condensed Matter* 0706.1355.

Churchland, Patricia S. (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.

Cohen, Jonathan & Callender, Craig (2009). A better best system account of lawhood. *Philosophical Studies* 145 (1):1 - 34.

Contessa, Gabriele (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science* 74 (1):48-68.

Contessa, Gabriele (2010). Modal truthmakers and two varieties of actualism. *Synthese* 174 (3):341 - 353.

Craver, Carl F. & Povich, Mark (2017). The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A* 63:31-38.

Donhauser, Justin, & Shaw, Jamie. (2018). Knowledge Transfer in Theoretical Ecology: Implications for Incommensurability, Voluntarism, and Pluralism. *Studies in History and Philosophy of Science Part A*.

Dowe, Phil (1992). Wesley salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science* 59 (2):195-216.

Dretske, Fred (1989). Reasons and causes. *Philosophical Perspectives* 3:1-15.

Dretske, Fred I. (1977). Laws of nature. *Philosophy of Science* 44 (2):248-268.

Ehring, Douglas (1997). *Causation and Persistence: A Theory of Causation*. Oxford University Press.

Ellis, Brian (2001). *Scientific Essentialism*. Cambridge University Press.

Fair, David (1979). Causation and the flow of energy. *Erkenntnis* 14 (3):219 - 250.

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

French, Steven & Ladyman, James (2003). Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese* 136 (1):31-56.

French, Steven & Ladyman, James (2011). In defence of ontic structural realism. In Alisa Bokulich & Peter Bokulich (eds.), *Scientific Structuralism*. Springer Science+Business Media. pp. 25-42.

French, Steven (1999). Models and mathematics in physics: The role of group theory. In Jeremy Butterfield & Constantine Pagonis (eds.), *From Physics to Philosophy*. Cambridge University Press. pp. 187--207.

Funkhouser, Eric (2007). A liberal conception of multiple realizability. *Philosophical Studies* 132 (3):467-494.

Gallegati, M., Keen, S., Lux, T., & Ormerod, P. (2006). Worrying trends in econo- physics. *Physica A: Statistical Mechanics and Its Applications*, 370(1), 1-6.

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

Giere, Ronald. (1999) *Science without Laws*. University of Chicago Press.

Godfrey-Smith, Peter (2009). Models and fictions in science. *Philosophical Studies* 143 (1):101 -116.

Hall, Ned (2015). Humean reductionism about laws of nature. In Barry Loewer, Jonathan Schaffer (eds.), *A Companion to David Lewis.* Wiley. 262-277.

Handfield, Toby (2004). Counterlegals and Necessary Laws. *Philosophical Quarterly* 54 (216):402-419.

Hellie, Benj, Murray, Adam, & Wilson, Jessica M. (forthcoming). Relativized metaphysical modality. In Otávio Bueno & Scott Shalkowski (Eds.), *Routledge Handbook of Modality*. Routledge.

Hendry, Robin Findlay & Rowbottom, Darrell Patrick (2009). Dispositional essentialism and the necessity of laws. *Analysis* 69 (4):668-677.

Hildebrand, Tyler (2013). Can Primitive Laws Explain? *Philosophers' Imprint* 13:1-15.

Hildebrand, Tyler (2014). Can bare dispositions explain categorical regularities? *Philosophical Studies*167 (3):569-584.

Humphreys, Paul (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method.* Oxford University Press.

Humphreys, Paul. (2018). Knowledge transfer across scientific disciplines. *Studies in History and Philosophy of Science Part A.*

Illari, Phyllis (2013). Mechanistic Explanation: Integrating the Ontic and Epistemic. *Erkenntnis* 78 (2):237-255.

Jaag, Siegfried (2014). Dispositional essentialism and the grounding of natural modality. *Philosophers' Imprint* 14.

Jackson, John David (1975). *Classical Electrodynamics* (2nd ed.). Wiley.

Jacobs, Jonathan D. (2010). A powers theory of modality: or, how I learned to stop worrying and reject possible worlds. *Philosophical Studies* 151 (2):227-248.

Jansson, Lina, & Saatsi, Juha. (2017). Explanatory Abstractions. *British Journal For The Philosophy Of Science.*

Jenkins, C. S. & Nolan, Daniel (2012). Disposition Impossible. *Noûs* 46 (4):732-753.

Jenny, Matthias (2016). Counterpossibles in Science: The Case of Relative Computability. *Noûs.*

Justin, Clarke-Doane (forthcoming). Modal Objectivity. *Noûs.*

Kistler, Max. (2013). The Interventionist Account of Causation and Non-causal Association Laws. *Erkenntnis, 78* (S1): 65–84.

Kitcher, Philip (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.

Kment, Boris (2014). *Modality and Explanatory Reasoning*. Oxford University Press.

Kripke, Saul A. (1980). *Naming and Necessity*. Harvard University Press.

Kuorikoski, Jaakko & Ylikoski, Petri (2015). External representations and scientific understanding. *Synthese* 192 (12):3817-3837.

Lange, Marc (2000). *Natural Laws in Scientific Practice*. Oxford University Press.

Lange, Marc (2004). A note on scientific essentialism, laws of nature, and counterfactual conditionals. *Australasian Journal of Philosophy* 82 (2):227–241.

Lange, Marc (2009). *Laws and Lawmakers*. Oxford University Press.

Lange, Marc (2016). *Because Without Cause*. Oxford University Press.

Lewis (1973). *Counterfactuals*. Blackwell.

Lewis (1986). *On the Plurality of Worlds*. Blackwell.

Lewis, David (1970). How to define theoretical terms. *Journal of Philosophy* 67 (13):427-446.

Lewis, David (1973). *Counterfactuals*. Blackwell.

Lewis, David (1983). New work for a theory of universals. *Australasian Journal of Philosophy* 61 (4):343-377.

Loew, Christian. (2019). Causes As Difference-Makers For Processes. *Philosophy and Phenomenological Research*, *98*(1): 89–106.

Loewer, Barry (1996). Humean Supervenience. *Philosophical Topics* 24 (1):101-127.

Martin, C. B. (2007). *The Mind in Nature*. Oxford University Press.

Maudlin, Tim (2007). *The Metaphysics Within Physics*. Oxford University Press.

Maxwell, James Clerk. (1861/62). On physical lines of force.

Merricks, Trenton (2001). *Objects and Persons*. Oxford University Press.

Morrison, Margaret. (1999). Models as Autonomous Agents. In M. S. Morgan, & M. Morrison (Eds.), *Models as mediators* (pp. 38–65). Cambridge: Cambridge University Press.

Mumford, Stephen (2004). *Laws in Nature*. Routledge.

Mumford, Stephen (2004). *Laws in Nature*. Routledge.

Murray, Adam, & Wilson, Jessica M. (2012). Relativized metaphysical modality. *Oxford Studies in Metaphysics*, 189–226.

Nersessian, Nancy J. (2002a). Maxwell and the method of physical analogy: Model-based reasoning, generic abstraction, and conceptual change. In D. Malament (Ed.), *Reading natural philosophy: Essays in the history and philosophy of science and mathematics*. Chicago: Open Court.

Nersessian, Nancy J. (2002b). Abstraction via Generic Modeling in Concept Formation in Science. *Mind & Society* (Vol. 5).

Nolan, Daniel (2013). Impossible Worlds. *Philosophy Compass* 8 (4):360-372.

Nolan, Daniel (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic* 38 (4):535-572.

Nolan, Daniel (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic* 38 (4):535-572.

Norton, John (1987). The logical inconsistency of the old quantum theory of Black body radiation. *Philosophy of Science* 54 (3):327-350.

Paul, L. A. & Hall, Ned (2013). *Causation: A User's Guide*. Oxford University Press UK.

Povich, Mark (2018). Minimal Models and the Generalized Ontic Conception of Scientific Explanation. *British Journal for the Philosophy of Science* 69 (1):117-137.

Price, Justin. (2018). The landing zone – Ground for model transfer in chemistry. *Studies in History and Philosophy of Science Part A*.

Psillos, Stathis (2011). Living with the abstract: realism and models. *Synthese* 180 (1):3-17.

Psillos, Stathis (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.

Reutlinger, Alexander (forthcoming). Is There A Monist Theory of Causal and Non- Casual Explanations? The Counterfactual Theory of Scientific Explanation. *Philosophy of Science*.

Roberts, Bryan W. (manuscript). *Philosophico-Scientific Adventures.*

Roberts, John T. (2008). *The Law-Governed Universe.* Oxford University Press.

Saatsi, Juha & Pexton, Mark (2013). Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations. *Philosophy of Science* 80 (5):613-624.

Saatsi, Juha & Vickers, Peter (2011). Miraculous Success? Inconsistency and Untruth in Kirchhoff's Diffraction Theory. *British Journal for the Philosophy of Science* 62 (1):29-46.

Salmon, Wesley (1990). *Four Decades of Scientific Explanation*. Pittsburgh University Press.

Salmon, Wesley. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Schaffer, Jonathan. (2001). *Causes as Probability Raisers of Processes. Journal of Philosophy* (Vol. 98).

Schneidman, Elad, Berry, Michael J., Segev, Ronen, & Bialek, William. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*(7087): 1007–1012.

Schwinger, Julian, et al. (1998). *Classical Electrodynamics.* Avalon Publishing.

Shaffer, Michael J. (2012). *Counterfactuals and Scientific Realism*. Palgrave MacMillan.

Shapiro, Lawrence A. (2000). Multiple realizations. *Journal of Philosophy* 97 (12):635-654.

Shoemaker, Sydney. (1980). Causality and properties. In Peter van Inwagen (Ed.), *Time and Cause.* D. Reidel.

Sider, Theodore (1999). Presentism and ontological commitment. *Journal of Philosophy* 96 (7):325-347.

Stalnaker, Robert (1968). A Theory of Conditionals. *Studies in Logical Theory, American Philosophical Quarterly*, Monograph: 2, 98-112.

Steinel, T., Asbury, J.B., Zheng, J., Fayer, M.D. (2004). Watching Hydrogen Bonds Break: A Transient Absorption Study of Water. *J. Phys. Chem. A* 108, 10957.

Steward, Helen (2013). Processes, Continuants, and Individuals. *Mind* 122 (487):fzt080.

Strevens, Michael (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.

Strevens, Michael (2012). Replies to Weatherson, Hall, and Lange. *Philosophy and Phenomenological Research* 84 (2):492-505.

Suárez, Mauricio (2002). An inferential conception of scientific representation. *Philosophy of Science* 71 (5):767-779.

Sugden, Robert (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology* 7 (1):1-31.

Swoyer, Chris (1982). » The Nature of Natural Laws «. *Australasian Journal of Philosophy* 60 (3):1982.

Tan, Peter (2017). Interventions and Counternomic Reasoning. *Philosophy of Science* 84 (5): 956-969.

Tan, Peter (2019). Counterpossible Non-vacuity in Scientific Practice. *Journal of Philosophy* 116 (1):32-60.

Thébault, K. P. Y., Bradley, S., & Reutlinger, A. (2016). Modelling inequality. *British Journal for the Philosophy of Science.*

Tooley, Michael (1977). The nature of laws. *Canadian Journal of Philosophy* 7 (4):667-98.

Tugby, Matthew (2013). Platonic Dispositionalism. *Mind* 122 (486):fzt071.

Tugby, Matthew (2014). Categoricalism, dispositionalism, and the epistemology of properties. *Synthese* 191 (6):1-16.

Tugby, Matthew (2015). The alien paradox. *Analysis* 75 (1):28-37.

van Fraassen, Bas C. (1992). From Vicious Circle to Infinite Regress, and Back Again. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992:6-29.

van Fraassen, Bas C. (1994). Interpretation of science: science as interpretation, in: J. Hilgevoord (Ed.) *Physics and Our View of the World*. Cambridge University Press.

Vander Laan, David (2004). Counterpossibles and Similarities. In Frank Jackson & Graham Priest (eds.), *Lewisian Themes: The Philosophy of David K. Lewis*. Oxford, UK: Clarendon Press. pp. 258-275.

Vetter, Barbara (2015). *Potentiality: From Dispositions to Modality*. Oxford University Press.

von Helmholtz, Hermann. (1858). On integrals of the hydrodynamic equations which express vortex-motions. In *Crelle's Journal für die reine und angewandte Mathematik* (Vol. 55, pp. 25-55). English Translation by P. Tait (1867) in Philosophical Magazine, Series 4, 33(226):485e512.

Vickers, Peter (2013). *Understanding Inconsistent Science*. Oxford University Press.

Weisberg, Michael (2007). Three Kinds of Idealization. *Journal of Philosophy* 104 (12):639-659.

Whittle, Ann (2008). A functionalist theory of properties. *Philosophy and Phenomenological Research* 77 (1):59-82.

Williamson, Timothy (2007). *The Philosophy of Philosophy*. Blackwell.

Williamson, Timothy (2018). Counterpossibles. *Topoi* 37 (3):357-368.

Williamson, Timothy (forthcoming b). Counterpossibles in Metaphysics. In Kroon, Frederick (ed.), *Philosophical Fictionalism.*

Wilson, Alastair (2016). Grounding Entails Counterpossible Non-Triviality. *Philosophy and Phenomenological Research* 92 (3).

Wilson, Jessica M. (2014). Hume's Dictum and the asymmetry of counterfactual dependence. In Alastair Wilson (Ed.), Chance and Temporal Asymmetry. Oxford University Press.

Woodward, James & Hitchcock, Christopher (2003). Explanatory generalizations, part I: A counterfactual account. *Noûs* 37 (1):1–24.

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

Worrall, John (1989). Structural realism: The best of both worlds? *Dialectica* 43 (1-2):99-124.

Wright, Cory (2015). The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A* 54:20-30.

Zuchowski, Lena. (2017). Modelling and knowledge transfer in complexity science. *Studies in History and Philosophy of Science Part A.*