

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Laws and causation

Citation

Hall, Ned. 2011. Laws and causation (working title: Causation and the Sciences). In The Continuum companion to the philosophy of science, ed. Steven French and Juha Saatsi, 96-119. New York: Continuum International Publishing Group.

Published version

https://www.bloomsbury.com/uk/the-continuum-companion-to-the-philosophy-of-science-9781441187611/

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:33979279

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story

Causation and the sciences

Ned Hall

§0 Introduction

The last several decades have seen a profusion of philosophical work on causation – most of it well outside the confines of philosophy of science. That's odd. You would have thought that a central aim of any scientific discipline is to map some aspect of our world's causal structure. So you would have thought that, concerned to illuminate the structure of scientific inquiry, philosophy of science would automatically take up the obvious epistemological and metaphysical questions about causation: How do we find out about causal structure? What is the nature of causal structure? And so, finally, you would have thought no argument required for the view that philosophy of causation is simply a branch of philosophy of science – any more than one needs to argue, say, that philosophy of space and time is a branch of philosophy of science.

So much for what you would have thought. Since at least the early 20th century, a strong undercurrent of suspicion has flowed through the philosophical world concerning the scientific legitimacy, or relevance, of *any* notion of causation. Sometimes this undercurrent bubbles to the surface, as in this famous old quote from Bertrand Russell (1953):

> The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.

Likewise in this more recent polemic from John Norton (2003; p. 2):

> Mature sciences, I maintain, are adequate to account for their realms without need of supplement by causal notions and principles. The latter belong to earlier efforts to understand our natural world, or to simplified reformulations of our mature theories, intended to trade precision for intelligibility. In this sense I will characterize causal notions as belonging to a kind of folk science, a crude and poorly grounded imitation of more developed sciences.

Take such suspicions seriously, and you will naturally consider philosophical study of causation pointless – at least, for those who want to understand the structure and aims of scientific inquiry.[1]

Happily, a countercurrent has pushed in the other direction. Wesley Salmon's important work on causation and explanation provides an obvious example (see for

[1] Maybe it has a point, if you want to sort out issues of legal or moral responsibility, etc.

example his 1984). More recently, philosophers *and* scientists have taken an explicit interest in the epistemology of causation, making use of "structural equations" to contribute important advances in this area. (See for example Spirtes et. al. 2001, Pearl 2009; more on structural equations in §4.3.3, below). A bit more tentatively, philosophers of science have delved into the metaphysics of causation, with results that have helped to clarify why, and in exactly what sense, the causal structure of the world ought to be a target of scientific inquiry. (See for example Hitchcock 2003, Woodward 2005, Strevens 2009.)

This essay contributes to the anti-skeptical effort. But I differ from some of my allies in holding that an adequate account of causation needs guidance from a background metaphysical position about laws of nature. To do causation right, you need a Grand Metaphysical View. Better: an awareness of the range of options for such a View, and of the ways in which choices among these options will affect the character of a theory of causation.

Here is the View I favor, quickly sketched:

Two distinct fundamental features characterize reality as a whole. First, there is a total history of complete physical states. Second, there are fundamental laws that dictate exactly how earlier states generate later states. It is the job of fundamental physics, and fundamental physics alone, to map this basic structure. All other natural facts are ultimately explicable in terms of it.[2]

Looking at the world *just* through the lens of fundamental physics, we won't see the need for any interesting, richly structured concept of causation. True, we could say that each complete physical state "causes" each later complete physical state – but why bother, once we have the fundamental laws in hand?

But the value of a concept of causation derives from details of our actual human predicament. First, we need *control* over our world (e.g., so it doesn't kill us). Second, we need to *understand* it (because we're curious). Third, while grasping a complete, correct physics would obviously facilitate understanding and control, we can only build up to such a grasp by way of *piecemeal approximations*.[3] The scientific value of causal concepts is precisely that they facilitate control, understanding, and piecemeal approximation. For the basic function of any causal concept is to record – with

[2] The qualifier "natural" merely signals a distinction between facts about the physical structure of the world, and facts – if such there be – about mathematics, or logic, etc. In addition, this picture needs adjustment to accommodate relativistic spacetimes, e.g. as follows: Pick an arbitrary point in spacetime. Consider its past light cone. Then the laws dictate how the complete physical state of earlier slices of this past light cone generate complete physical states of successive slices. In what follows, I mostly won't bother with this refinement.

[3] And even with a complete, correct physics in hand, computational limitations would continue to make the use of approximations essential.

widely varying degrees of precision and informativeness – facts about how localized aspects of the world at one time nomologically depend on localized aspects at earlier times. Knowledge of such dependencies yields control: think of the case where localized facts about the future depend on what you do right now. It yields understanding: for this requires, among other things, an organized knowledge of the localized patterns of dependency that the world exhibits. (See especially Woodward 2005 and Strevens 2009.) And it facilitates the project of constructing better and better approximations to a complete and correct physics: for to know a causal fact is to have a nugget of information about how the state of the world at an earlier time leads to the state of the world at a later time.

Finally, it is (*pace* Norton) precisely one of the central aims of the special sciences – by which, we can now see, we should mean *any science other than fundamental physics* – to articulate the kinds of causal facts that facilitate control, understanding, and piecemeal approximation.

That's the view, highly compressed. Now I'll try to unpack it rather more systematically, starting with some

§1 Remarks on methodology

The contemporary causation literature has two off-putting features, especially for those whose interest in causation stems from an interest in science. In both cases philosophers of science have legitimate cause for complaint, but shouldn't get carried away: the methodological corruption in this literature is not nearly so severe as to deprive it of real importance.

§1.1. The role of intuitions

The first off-putting feature concerns the way in which the literature deploys intuitions, in the form both of judgments about specific cases and alleged "platitudes" concerning the nature of the causal relation. A philosopher advances an analysis of the schema "event C is a cause of event E". Someone devises a fiendish counterexample: a case (typically hypothetical) about whose causal structure we all have a firm intuition not captured by the analysis. Then it's back to the drawing board, either to add further bells and whistles (perhaps appealing to "platitudes" for guidance), or to scrap the analysis, and try something new.

A quick example: Start with the simple idea that for C to cause E is for it to be the case not only that C and E both occur, but that if C had not occurred, then E would not have occurred. But what if C not only brings about E, but also cuts off some backup process that would have brought about E, had C not occurred? Trouble, surely, since if C had not occurred, E would have occurred anyway, thanks to the

backup. The canonical example makes use of the "neuron diagrams" that populate much of the literature (thanks to the influence of David Lewis; see in particular his 1986b):

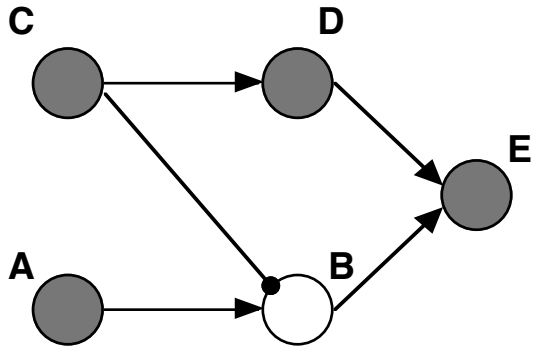


Figure 1

Circles represent neurons; arrows represent stimulatory connections between neurons; lines with blobs on the end represent inhibitory connections. Shading a circle indicates that the neuron fires. The order of events is left to right. In Figure 1, **A** and **C** fire simultaneously; **C** sends a stimulatory signal to **D**, causing it to fire, while **A** sends a stimulatory signal to **B**. But, since **C** also sends an inhibitory signal to **B**, **B** does not fire. Finally, **D** sends a stimulatory signal to **E**, causing it to fire. Figure 2 shows what would have happened if **C** had not fired:

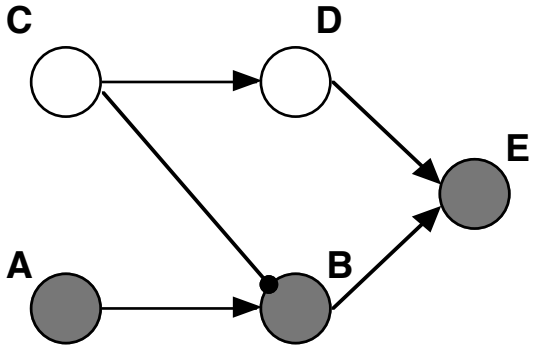


Figure 2

So we tinker, declaring that what it is to for C (the firing of **C**) to be a cause of E is for there to be a *chain* of dependence linking C to E (essentially the analysis proposed in Lewis 1973a). In the diagram, E depends on D (for if **D** had not fired, **E**

would not have fired[4]), and D in turn depends on C. And we make the tinkering look less *ad hoc* by appealing to the supposedly *a priori* insight that causation is transitive.[5]

But wait! Examples of *late preemption* show that this strategy won't work. Here is a favorite, which for future reference I'll call "Suzy First": Suzy and Billy, two vandals, throw rocks at a window. Each has perfect aim. But Suzy throws her rock a bit quicker, hence it gets there first. Had she not thrown, the window would have broken all the same. Still, only *her* throw counts as a cause of the breaking. Yet (exercise) no chain of dependence links her throw to the breaking.

You can no doubt see how things are going to go from here. Ever more clever analyses will confront ever more clever counterexamples. And you can easily come to wonder *why we are bothering* – what of value we would learn, even if we succeeded in constructing an analysis of causation that ran the gauntlet of every possible example. As the convolutions pile up, the hope of finding insights into causation *that will ground insights into science* is fading fast.

This is an overreaction. To be sure, we should repudiate the attitude that treats intuitions as non-negotiable data – an attitude on display in the following quote from Lewis:

> When common sense delivers a firm and uncontroversial answer about a not-too-far-fetched case, theory had better agree. If an analysis of causation does not deliver the common-sense answer, that is bad trouble. (Lewis 1986b, p. 194)

It makes no sense to think that "common sense" is such an infallible guide to the discovery of an account of causation *that will illuminate the structure of scientific inquiry*. All the same, common sense *is* a guide of sorts – namely, to where a potentially useful causal concept or concepts might be found. (Compare mathematics, where ordinary intuitions have oft served as guides to the construction of useful concepts.) The appropriate rule is really this one: If an analysis of causation does not deliver the common-sense answer, that is some evidence – defeasible, of course – that something of importance has been overlooked.

I think that hypothetical cases have important lessons to teach us philosophers of science – some of them, anyway, and in a fashion that may require some subtlety to discern. No reason to keep you in suspense: I think that cases with the structure of Suzy First highlight the importance of mediation by a *causal mechanism*, and that once we get the right account of what causal mechanisms are, it will naturally fall out

---

[4] Note the now-standard reading of the counterfactual as "non-backtracking": thus, we do not reason that if **D** had not fired, that could only have been because **C** did not fire, whence **E** would have fired all the same. More, in §REF below, on how to ground this reading in an account of the truth-conditions for counterfactuals.

[5] For an example of a regrettably naïve appeal to transitivity along these lines, see Hall 2000.

that a central task of the special sciences is to discover them and accurately describe their structure. We'll return to this topic in §4.4.2, below. For now, the crucial point is just this: We should reject the too-uncritical Lewis stance towards intuitions, and adopt instead the moderate attitude that sees intuitions as clues to where potentially fruitful concepts of causation might be found.

§1.2. The proper target of analysis

The second off-putting feature of the literature is its almost exclusive focus on expressions of the form "event C is a cause of event E". Do scientists – at least, *qua* scientists – really pay much attention to statements of this form? Probably not.[6] So an account of causation that *limited* itself to analyzing sentences like this might seem to deprive itself of interest, at least to philosophy of science.

That's hasty. Observe, first, that *nearby* claims clearly matter to science. Take a case where Suzy, by herself this time, throws a rock at a window, breaking it. We should all agree – because it's true – that Suzy's throw is a cause of the window breaking. Yawn. But our interest might perk up if we are told, in detail, what it was about her throw *in virtue of which it was a cause of the breaking*. Or again, if we are told, in detail, what small variations in the manner or circumstances of her throw would or would not have led to a breaking (of exactly what kind). *This* sort of information can conceivably form the ingredients for a causal generalization of the kind that the special sciences traffic in. It is also the sort of information possession of which would give you a properly fleshed out *explanation* for why the window broke – yielding an understanding deeper than the shallow understanding you possess when you know merely that the window broke because Suzy threw a rock at it. So clarity about singular causal claims can set the stage for clarity about *other* kinds of causal claims, ones that are the life-blood of the special sciences.

§2 Two conceptions of causation

I indicated above that an account causation in the sciences should lean on a grand metaphysical view for guidance. I won't try to provide a survey. But one distinction really must be put on the table, between two profoundly different ways of conceiving the subject matter of the theory of causation. On one conception – which traces back to Aristotle's notion of "efficient causation" – objects (or "substances") in the world possess, in virtue of their essential natures, *sui generis* causal powers or

[6] With some obvious and important exceptions – e.g., think of Feynman's study of the Challenger disaster. More generally, it would be a foolish philosopher of science who insisted that investigating singular causal claims could *never* be of interest to science. It's just that such investigation is not often the primary focus.

capacities, and it is by reference to the operation of these powers or capacities that we explain change. We can, if we wish, say that one *event* is a cause of some later *event*, but this style of description misses the metaphysically important features of the case: what is metaphysically *fundamental* is always the way in which one or more objects in the world act on others in virtue of their respective causal powers or capacities. If you want to understand what causation is, then you first need to understand what it is for an object to possess a causal power or capacity.

The second, rival conception grants that we can speak of objects in the world as possessing powers or capacities. But this is loose talk, explicable in terms of something more metaphysically fundamental: patterns of dependency in the way that the events that constitute our world's history unfold. So a philosophical theory of causation should focus on these dependency relations.

Which conception one endorses will profoundly affect how one thinks about scientific inquiry.[7] Each deserves careful development. (And squabbling over them really ought to stop until the best versions of each are on the table. Fat chance.) But for the sake of tractability, I'm going to plant my flag with the second conception, and assume, from here on, that *causal structure* is, at bottom, *dependency structure*.

§3 Laws and counterfactual structure

And this dependency structure, in turn, I will take to be *counterfactual* dependency structure – a structure reflected in the pattern of truth-values for counterfactuals relating the state of some localized part of the world at one or more places and times to the state of some localized part of the world at one or more *later* places and times. This proposal raises three obvious questions:

Why understand dependency as *counterfactual* dependency?

What is the right account of the counterfactuals in terms of which dependency is to be understood?

---

[7] For example, the first conception lends itself to skepticism over whether our world is governed by simple physical laws at all: after all, why expect the operations of objects' causal capacities to be so tidy? So it should come as no surprise that Nancy Cartwright, one of the most vigorous and creative contemporary advocates of the first conception, vehemently rejects a conception of physics as aiming to provide us with such laws. (See especially her 1999, and, for an earlier and forceful statement of a closely allied view, Anscombe 1971.) The second conception of causation, by contrast, is the one that seems almost inevitable, if you adopt the sort of "fundamentalism" about physics that Cartwright opposes: the view that fundamental physics should be in the business of looking for mathematically elegant laws, that describe with perfect precision and accuracy the evolution through time of the world's complete physical state. (Which is not to say that these laws must be deterministic: precision and accuracy can take a probabilistic form, as well. For example, a "fundamentalist" will think it perfectly appropriate to assume that decay phenomena, even if they are irreducibly stochastic, are characterized by *perfectly numerically precise* objective probabilities.)

Why the temporal asymmetry? Why not also focus on future-to-past counterfactual dependence?

And one more question, less obvious but quite important:

In giving an account of causal structure in terms of counterfactual dependency structure, does it matter what conception one has of the fundamental laws themselves?

Let's take these questions in turn, starting with a very brief answer to the first: We should understand dependency as *counterfactual* dependency in part because, as this essay tries to show, doing so is fruitful. And in part for the reason broached in the introduction: The fundamental laws of our world *endow* it with a counterfactual structure; it serves some of our most basic epistemic and practical needs to investigate that structure; causal concepts can be very naturally understood as the essential conceptual tools needed for such investigation. More on these points, as we progress. Now let's turn to the remaining questions.

§3.1. A simple account of (simple) counterfactuals

Take the canonical form for a counterfactual to be the following: "If it were the case that A, then it would be the case that B" where A and B can be filled in by any declarative sentence. In typical cases, A will be false; hence the label "counterfactual". Since the work of Lewis and Stalnaker in the '60's and early '70's (see for example Lewis 1973b, Stalnaker 1968), it has become standard to assume that truth conditions for counterfactuals are to be given by some kind of *similarity semantics*, e.g. the following:

"If it were the case that A, then it would be the case that B" is true iff: among the possible worlds in which A is true, the one most similar to the actual world is one in which B is also true.

The search is then on for an analysis of the relevant notion of "similarity".

To my mind, this has been a mistake, leading at times to pointlessly byzantine constructions (e.g. Lewis 1979). In fact, I think it is not so useful to look for a one-size-fits-all semantics, especially given the incredible variety counterfactuals can exhibit. Instead, I will simply give a plausible recipe (adapted from Maudlin 2007b) for evaluating a limited but important range of counterfactuals.

Require that the antecedent A concern events at (or around) a specific time t, and the consequent B matters after t. Consider an example. Suzy is standing near a window, not throwing anything. If Suzy *had* thrown a rock at the window, then the window would have broken. In outline, the recipe for evaluating this counterfactual proceeds in two stages. First stage: we construct an alternative state – a *nomologically possible* state – of the world at the given time t in which, instead of just standing around,

Suzy is throwing a rock at the window. Second stage: we extrapolate this state forward, and check to see if the window breaks. If so, the counterfactual is true. If not, false.

How do we construct this alternative state? By making *localized* changes to the actual state of the world, sufficient to make it the case that Suzy throws – but involving no extra, gratuitous alterations. In part, the prohibition on gratuitous alterations gets taken care of, by insisting that the changes be localized to the portion of the world described in the antecedent (i.e., Suzy and her immediate environment, at time t). Thus, we are not to imagine a possible world in which Suzy throws, but the window has been moved, or has had a brick wall erected in front of it, etc. But the prohibition demands more: e.g., we are not to imagine Suzy throwing a rock with a tiny bomb embedded in it, which will explode well before the rock reaches the window. Don't expect an exact account of what sort of alteration qualifies as "gratuitous". It is fair to say, though, that the more sharply we can pin this issue down in any given case, the more rigorous use we can expect to get out of counterfactuals – and that will matter a great deal, in many scientific contexts.

Both indeterminacy and context-dependence arise naturally at this first stage. Many rocks lie scattered around Suzy, all ready to hand. So there is surely no fact of the matter about *which* one she would have thrown. Or suppose that she really hates breaking windows. In a context that makes this disposition salient, we can truly say that if she had thrown a rock at the window, she would have missed (intentionally); in a context that makes her skill salient, we can truly say the opposite.[8]

But it does not follow that either indeterminacy or context-dependence creeps in at the *second* stage, where we extrapolate from the counterfactual state of the world in order to see whether the consequent comes out true. And indeed there won't be if we adopt – as I now propose we should – the obvious story to tell about how such extrapolation proceeds: the counterfactual state of the world constructed at the first stage is to be evolved forward *in accordance with the actual fundamental laws of nature*.[9]

§3.2. Temporal asymmetry

The Recipe stipulates that the antecedent of a counterfactual concerns a time or times earlier than the time or times covered by the consequent. Why? Not because it's a Revealed Metaphysical Truth that causation must be asymmetric (and we have an eye towards analyzing causation in terms of counterfactuals). A more plausible

[8] There are several *other* ways in which context-dependence can creep in; it's a useful exercise to try to come up with them.

[9] If these fundamental laws are stochastic, then indeterminacy of a special sort will show up. Notice, though, how different it is from the "which rock?" variety.

and, frankly, much more interesting reason is the following: while we can certainly employ the Recipe in the future-to-past direction, the kind of dependence we thereby uncover is unfit for sustained scientific investigation.

Consider a simple example (here I draw on Elga 2000). Suzy throws a rock at a window, breaking it (by herself, say). After the breaking, shards of glass lay scattered on the ground, and the rock lies a bit further off. Imagine this process running in reverse – a physical possibility, given the time reversal symmetry our fundamental laws appear to display. Shards of glass and a rock are lying on the ground; then they fly up into the air in such a way that the shards of seal themselves into an unbroken pane, which flexes in such a way as to give the rock (which has traveled in just the right way to be at the right location) a little extra push; it flies through the air; Suzy's hand closes over it. At time 0, the rock and shards lie on the ground; at later time 1, the window contains an unbroken pane of glass. How does this time-1 fact counterfactually depend on the state of the world at time 0?

Extremely sensitively. Consider what it takes for the shards to fly up and form an unbroken pane. Vibrations in the earth converge beneath each individual shard so as to propel it upwards. Subtle air currents adjust the trajectory of each shard so that it reaches its appointed spot within the windowpane at just the right time. Processes internal to each shard prepare their surface molecules so that bonds will form as the shards come together. So if we examine the localized state of the world within any little patch of ground, or indeed any little volume of air in the surroundings of the window, we reach the following verdict: if the state of the world in that region had been *even a little different* at time 0, there would have been no unbroken pane of glass at time 1.[10]

But that means that, while the state of things at time 1 counterfactually depends on the state of things at time 0, it does so in an *utterly non-discriminating fashion*. Compare the structure of the dependency relations when we run the example in the normal direction: the ultimate state of the window – broken – depends on facts about what Suzy is doing at an earlier time, but does so in a usefully insensitive fashion (i.e., if she had thrown just a tiny bit differently, the window still would have broken, etc.). And the broken state of the window does *not* depend in any such systematic way on, for example, the state of the little patch of ground a few feet away from Suzy.[11]

[10] A bit more carefully: for any such little region at time 0, the overwhelmingly vast majority of variations on its time-0 state are such that, if it had that variant, then there would have been no window at time 1.

[11] The only obvious dependence is this: if the state of that patch had somehow been such as to *interfere* with Suzy, then the window would not have broken.

In sum, when we ask, "How does the state of the window depend on localized states of the world at a slightly earlier time?", we can give, in the *normal* case, a fairly interesting and richly detailed answer. But in the temporally bizarre world, we cannot: all we can say is that it is true of pretty much every localized region near the window that, had its state been different at time 0, there would have been no unbroken window at time 1. And what goes for breaking windows goes for any process that exhibits such thermodynamic irreversibility – which is to say, the vast majority of processes studied by the special sciences. So, while a world that was the time reverse of our own would of course have a quite interesting past-to-future *fundamental physics*, it would not have an interesting past-to-future *localized dependence structure*. And that is just to say that our *own* world does not have an interesting future-to-past localized dependence structure. So the reason for the stipulation in the Recipe is something more physically deep and metaphysically deep: given the overall thermodynamic structure of our world, there is no point to considering localized, future-to-past dependence.

The Recipe shows how a world's fundamental laws can endow that world with a rich, past-to-future counterfactual structure. There is no reason to suppose that science cannot fruitfully investigate that structure. But I suspect that how one thinks such investigation ought to go will depend on the view one takes of the nature of fundamental laws. The next section looks into this issue.

§3.3. Rival accounts of law

Philosophers who hold that there *are* fundamental laws of nature divide into profoundly opposed camps over their metaphysical nature. The most basic split concerns the following questions: Are laws mere patterns in the phenomena (patterns that are in some way salient, to be sure – but still, nothing more than patterns)? Or are they something more, something that somehow *governs* or *constrains* those phenomena? I'll call those who take the first, laws-as-mere-patterns view "Humeans", their opponents "anti-Humeans".

Illustration. Suppose that physicists had determined Newtonian particle mechanics to be the complete and correct physics for our world. Then we would have learned quite a lot about the world's fundamental structure: say, that (1) there are (nothing but) finitely many point particles, (2) each possessing an unchanging value for mass and charge, and (3) each moving, at all times, in accordance with Newton's second law of motion together with the appropriate force laws. But questions about the metaphysical status of these very claims would remain.

For anti-Humeans, (1) – (3) capture basic facts about the metaphysical structure of nomological possibility and necessity. Now, views about the exact *way* in which

the capturing gets done will vary: Some will think that magnitudes such as mass and charge are fundamentally *modal*, endowing particles that possess them with the power to affect and be affected by other particles in the way described by the Newtonian equations; others will think of these magnitudes as non-modal, but hold that the world contains some extra metaphysical ingredient that *constrains* or *governs* particle behavior. Either way, an anti-Humean will view our Newtonian particle world as serving up a *substantive* answer to the question, "Why do all particles move, at all times, in conformity to the Newtonian equations?" (One sample answer: Because, in virtue of their masses and charges, they interact with each other in such a way that they so move. Another: Because the fundamental laws constrain them to so move.)

The Humean thinks that there is no such substantive answer, for she holds that the fundamental physical magnitudes are non-modal, and that what the laws are supervenes on *de facto* particle behavior in a way that undercuts any talk of "governing". The most that can be said is that the fact that the particles so move is a particularly useful or important fact to take note of. On the best going version of Humeanism (for which see Loewer 1996, Beebee 2000, and my 2010), what makes this fact special is roughly that it captures, in an easy-to-state form, quite a lot of information about the world.

Anti-Humean views have a common feature that is very important to recognize. According to all of them, there are facts about what does in fact happen in our world – about what its total history of states is. But in addition, according to each of them the world has a *determinate and objective modal structure*. The views differ as to where exactly this structure comes from. But they all recognize it. And so they all can (and should) agree that the Recipe provides a valuable conceptual and linguistic tool for articulating this structure.

Matters are otherwise for the Humean. She can, of course, insist that what *she* calls "fundamental laws" are to be used in giving truth conditions for counterfactuals in exactly the way described by the Recipe. Then she will mostly agree with her rivals about which counterfactuals are true or false.[12] But by her lights, that looks like nothing more than an arbitrary semantic decision to use the "if it were the case that …, then it would be the case that…" construction in a certain way.

[12] Mostly, but not completely. For example, an anti-Humean will be committed to counterfactuals like this one: if there had been just one particle, then it would have been the case that, had there been a second, the two would have interacted in precisely such-and-such a way. The Humean should, by contrast, probably say that if there had been just one particle, then the total history of the world would have been too simple to render it determinate what the laws were – and so there would have been no fact of the matter about what would have happened, had there been a second particle.

Now, this point is tricky. After all, in some good sense, the way we use our linguistic constructions *must* involve some semantic decision. Still, if you have a view about the sorts of structures the world objectively contains, then that will motivate judgments as to whether certain ways of representing the world are more or less *apt*. In the case of our Newtonian particle world, anti-Humeans all think that this world has some sort of objective structure that goes beyond the facts about how particles happen to move. And so each of them will see the counterfactual construction as specially suited to articulating this extra structure. But for the Humean, there is no such extra structure. There is just a decision that, when evaluating counterfactuals, we must always hold fixed those de facto claims about the world that she counts as "fundamental physical laws".

One wonders at the basis for this decision. Why not allow much greater flexibility? You are wondering what would have happened, had you performed a certain experiment. Well, perhaps the answer will vary, depending on whether you hold fixed physical regularities, or rather chemical regularities, or (yet again) biological regularities … etc. An anti-Humean would insist that there just is *some one thing* that would have happened if you had performed the experiment (modulo imprecision in the antecedent of the counterfactual, and any indeterminism in the fundamental laws). That is just to say that the world has, here as everywhere else, a determinate counterfactual structure – in which case we are obliged to hold fixed, in our counterfactual reasoning, those aspects of the world that most directly reflect this structure. But our Humean can make no sense of such insistence. So what mistake does *she* think we are making, if, in pluralist fashion, we hold that a question concerning what would have happened under certain counterfactual circumstances gets different (but equally correct) answers, depending on which patterns in the phenomena the context in which the question arises instructs us to hold fixed? Not a *factual* mistake, surely. What, then?

At any rate, having raised this issue I will set it aside for the remainder of the essay. In what follows, I'm going to assume, with the anti-Humean, that our world has a determinate counterfactual structure. And it is exactly this structure that provides the materials for *causal* structure.

§4 Causation as counterfactual covariation

Return to the guiding idea that was presented in the introduction: causal concepts matter to the sciences, for they provide tools by which those sciences can articulate – typically in an approximate and piecemeal manner – structures of dependency in the world. By means of such articulations we enhance our ability to control our world, and to increase our understanding of it.

§4.1. The central idea

What account of causation should we favor, if we see causal structure as constituted by patterns of dependency? By this point, the question would seem to answer itself. We should favor a *counterfactual* account, with the counterfactuals understood by means of the Recipe. And the guiding idea behind the account ought, surely, to be the following: a causal connection between goings-on in one localized region of space-time and goings-on in some later region of space-time is just constituted by a pattern of counterfactual covariation between these regions. Imagine you're God. You can reach in and tweak the state of the world in one localized spot, in any of various ways, and then watch how these alterations propagate forward, in accordance with the fundamental laws. If the alterations You introduce *here* lead to corresponding alterations *there*, then here and there are causally connected.

That, in essence, is the account of causation Lewis presents in "Causation As Influence" (2004), intended to replace his earlier counterfactual account of causation. But there are importat differences of emphasis. For Lewis, the name of the game was to give truth conditions for "C is a cause of E", roughly these ones: both events occur, and a suitably wide range of alterations of C would have led to an appropriately corresponding range of alterations of E. Lewis was not particularly interested in exploring in detail what "suitably wide" and "appropriately corresponding" should mean. He was happy to leave these matters vague, since he thought that the ordinary concept of causation he was targeting for analysis was itself vague.

But from a scientific perspective, the interest in the account resides in the nature of this correspondence. Who – qua scientist, anyway – cares whether event C is a cause of event E? What we really want to know about is the detailed structure of the counterfactual covariation between the region in which C occurs, and the region in which E occurs. Better still, if this detailed structure, or structures near enough like it, gets *repeated* in other regions of space and time. But this difference of emphasis should not lead us to lose sight of the fact that Lewis has drawn attention to a kind of objective structure the world possesses that clearly deserves to be called "causal structure", and is of clear scientific interest.

§4.2. Getting more sophisticated

So we now have the following slogan: Causal structure is localized dependency structure, of the sort articulated by the Recipe; investigating such structure is one of the main tasks of the sciences (the special ones, anyway). Getting more sophisticated requires looking a bit more closely how we might go about *representing* localized dependency facts.

Suppose we have one region R – as Lewis might put it, the region in which C occurs – and some later region S (say, the region in which E occurs). How should we capture facts about how the state of things in S depends on the state of things in R? We will evidently need some *taxonomy*, in terms of which we can specify the actual states of R and S as being selected from a number of possibilities. I suggest that a complete theory of the structure of the special sciences must in large part be a theory of such taxonomies. Here I'll rest content with some general remarks.

First, in very many cases, more than one taxonomy will be available. Something is going on in Suzy's body. We could describe it in physiological terms, or biochemical terms, or chemical terms, or… etc. For short, we should be *pluralists* about scientific taxonomies.

Second, while we should expect the possibilities articulated by any taxonomy to be exclusive – when some part of the world is in a given state, that is incompatible with it's being in any other state – we should not expect them to be exhaustive. Something is going on inside this beaker. As it happens, it can be accurately described in chemical terms, as can many alternatives to it. But suppose we are wondering what would have happened if the beaker had contained a *plasma*. Then our chemical taxonomy gives out. That doesn't mean we can't *describe* this alternative – just that we need a different taxonomy, borrowed from a different branch of science. For short, typical scientific taxonomies will be *limited in scope*. Plausibly, the only place where we could find a taxonomy *not* limited in scope is in fundamental physics.

Third, there is at least a strong presumption in favor of using *intrinsic* taxonomies – ones that distinguish possible states of a region only on the basis of features that characterize how that region is in itself, independent of how any other region is. Within physics and chemistry, this requirement looks non-negotiable, though the softer sciences may relax it.[13]

Fourth, a mark of maturation in a scientific discipline is that its taxonomies, and its methods for deploying them in stating generalizations about localized dependence structure, work to substantially rein in the kind of context dependence and indeterminacy in counterfactuals that we touched upon in §3.1, above. Examples bear this point out: Consider how firm our grasp is of counterfactual reasoning in, say, chemical or biochemical contexts, as compared to economic contexts. Notice that we have, here, an important way in which a science can fail to be "exact" that has nothing to do with whether that science deploys sophisticated mathematics: Does that science use its taxonomies in a way that adequately constrains the context dependency and

[13] The obvious example (to philosophers, anyway) is a psychology that tries to discern patterns of dependence between an agent's actions and the *contents* of her mental states – assuming, as has become commonplace, that such content is partly determined by how she is connected up to her environment.

indeterminacy in counterfactuals? You can festoon your economic theories with all the differential equations you like; but if you don't solve *this* problem, then you don't really have a proper science.

Finally, a sophisticated scientific taxonomy – and even many unsophisticated ones, including those that we more or less tacitly appeal to in ordinary life – will come equipped with standards for *comparative similarity*, that allow one to locate a given possible state within a landscape of surrounding or "nearby" alternatives. To see how ubiquitous this feature is – and to see, at the same time, its importance – consider our prosaic example of Suzy throwing a rock at a window. At one time and place, one thing happens: Suzy throws a rock. At a later time and place, something else happens: the window breaks. How do the goings-on in the second region depend on the goings-on in the first region? A crude way to map the patterns of dependency might deploy the following taxonomies. As for the window, it can either remain unchanged, or crack, or shatter (not exhaustive, obviously!). As for Suzy and her throw, we can map the alternatives (some of them, anyway) by allowing the *mass* and *color* of her rock to vary counterfactually. Then – even with these utterly simple-minded taxonomies in play – we will be able to say something *explanatorily much more interesting* than merely that the window broke because Suzy threw a rock at it. For we will can begin to articulate the explanatory *relevant* features of the throw, by noting, for example, that the *mass* of the rock mattered – in that had that mass been lower than a certain threshold, the window would merely have cracked, and had it been lower still, the window would have remained unaffected. And we can note that the color of the rock is *irrelevant*, in that, had the rock been any different color, the window would have broken all the same.

In short, we highlight what is explanatorily important by contrasting what actually happens with alternatives that are similar and different in certain specified respects. So we can advance the following tentative generalization: A taxonomy contributes to the explanatory power of explanations proffered by means of it to the extent that it comes equipped with similarity metrics that enable such contrasts. Equipped with such a taxonomy, it may even be possible to write down an interesting equation relating the state of R to the state of S.[14] Notice that the sign that you are working with a taxonomy highly sophisticated in this respect is precisely that the space of alternatives it depicts has a natural mathematical representation.

[14] It will always be possible to write down *some* equation: all you have to do is index the possible states of R by some numbers, and do the same for S; then you can gerrymander an equation to fit whatever the pattern of dependence between S and R happens to be. It goes without saying that the use of mathematical equations in such cases will hardly be illuminating.

§4.3. *Advantages*

Several advantages accrue to thinking of scientific inquiry as organized around the discovery and description of causal structure, understood as localized dependency structure. Here I will sketch some of the more prominent ones.

§4.3.1. The scientific topic-neutrality of causation

The first point is one that John Campbell has emphasized (in conversation): On the conception of causal structure as localized dependency structure, causal structure turns out to be a highly *generic* feature of the world. Indeed, the variety of ways in which we can usefully represent such structure seems limited only by our creativity and imagination in coming up with clear and effective taxonomies. And it is evident that these taxonomies can overlap in their coverage: dependency patterns can be discerned in the same stretch of reality at a chemical level, a biological level, a psychological level, etc. We can thus do justice to the attractive idea that the sciences are in the business of investigating the causal structure of the world, while leaving ample room for significant differences in their methods of investigation. They are involved in *one and the same task* – but depending on the scale at which they are looking, they might bring radically different descriptive resources to bear on carrying it out.

§4.3.2. The causal character of special science "laws"

Second, we're now in a position to correct a tempting mistake concerning the content of generalizations in the special sciences. Let me explain, by means of a pair of contrasting examples.

Start with Newtonian particle mechanics. Presenting that theory, we make some de facto claims about the behavior of all particles at all times – say, that this behavior conforms to such-and-such equations. But then (depending on our views on the metaphysics of fundamental laws), we go on to comment that this claim about particle behavior has a special sort of status: it is nomologically necessary, or holds as a matter of law, etc. We therefore respect a sharp distinction between the *content* of a law – a description of universal but purely de facto behavior of stuff in the world – and the *modal status* of that law.

Now switch to, say, this example of a simple generalization from metallurgy: All metal bars that are heated expand in direct proportion to their change in temperature. Of course, this generalization needs to be qualified in order to be true: What if you heat a metal bar so much that it melts? What if you set off a bomb next to it as you heat it? And so on. Set these issues aside (for now – we will return to one of them in §4.4.2). Assuming we have an appropriately qualified version of the generalization, we might think that it likewise just says something about the de facto behavior of stuff in the world. Wrong, I think. Rather, what the metallurgical claim *really* aims to report is not merely that, in all of space and time, there are no examples of

metal bars that are heated at one time, yet fail to expand by the appropriate amount shortly thereafter. I think it is intended to report, in addition, that such expansions *counterfactually depend* on the heatings.

To fail to notice that special science generalizations have this kind of content causes epistemological mischief (as Ward 2010 very nicely argues). Presented with a claim loosely stated in the form "all A's are B's", you might wonder why it can't be tested by looking for examples of non-B's that are non-A's: why can't we test the metallurgical law by looking for examples of things that don't expand, and checking to see that they are not previously heated metal bars? Of course you can point out statistical asymmetries (it is antecedently enormously probable that a randomly chosen non-B will be a non-A, etc.), but there is a better point to make: The idea that you could test a claim to the effect that the behavior of a metal bar counterfactually depends in a systematic way on whether and how it is heated, *without* doing direct experimentation on such bars, is ridiculous on its face. In sum, I suggest, with Ward, that the famous "Raven's Paradox" has earned its credentials *as* a paradox only thanks to a mistaken view about the contents of the generalizations to which it has been applied.

§4.3.3. Foundations of structural equations

Among philosophers and scientists interested in causation, one idea has gained currency in recent years that would seem to offer an alternative to our approach. On this rival view, causal structure is to be analyzed by means of *causal models*: roughly, systems of appropriately chosen *variables*, together with *structural equations* that relate them. Such models allegedly provide tools by which to analyze, in a controlled and rigorous fashion, certain specialized counterfactuals in terms of which causation is to be defined.

Alas, there are two painfully obvious foundational questions that, to my knowledge, the literature in this area has yet to adequately answer: What are variables? And what are the truth conditions for structural equations?[15]

Interestingly, our foregoing discussion suggests an obvious prescription for choosing variables and values, for an arbitrary system for which we might wish to construct a causal model: First, find a way to "carve up" the system into discrete, well-defined sub-systems. Second, for each relevant sub-system, and each relevant

[15] For example, scan Pearl's (2000) *book-length treatment* of causal modeling, for example, and you find nothing more substantive than scattered remarks such as the following: "The world consists of a huge number of autonomous and invariant linkages or mechanisms, each corresponding to a physical process that constrains the behavior of a relatively small group of variables." (p. 223) We are told nothing – *anywhere in the book* – that might yield an adequate understanding of such concepts as "autonomous", "mechanisms", or – most crucially – "constrains".

time or time-interval, introduce a variable to characterize the intrinsic physical state of that sub-system at that time, or during that time-interval. That is to say: impose some scientific taxonomy.

Furthermore, there is a simple, attractive – I would almost like to say: inevitable – story to tell about structural equations. We should take them to describe patterns of localized dependence of just the sort articulated by the Recipe – or rather, by the Recipe, when it is extended in a natural way to accommodate counterfactuals whose antecedents specify alterations to the state of the world at one or more localized times and places. But this is easily done (see my 2007 for details). In short, the scandalous state of the literature on structural equations is easily corrected. But once the correction is in place, we can also see that that literature has overinflated the significance of structural equations and the causal models that make use of them. Far from providing a foundation for notions of counterfactual dependence and causation, these models turn out to be nothing more than useful tools for representing antecedently understood patterns of counterfactual dependence. – Which is fine, given that their main value is for the *epistemology* of causation: after all, when it comes to testing some complicated causal hypothesis, half the battle consists in representing that hypothesis perspicuously enough that the way evidence bears on it can be sharply discerned. If anything, answering the foundational questions about causal models should remove any basis for what would otherwise be reasonable suspicion regarding their use in science.

§4.3.4. Putting metaphysics in its place

Adopt our conception of causation, and you are well positioned to guard against certain errors that have infected philosophical discussions in which theses about causation figure prominently. I will consider just one famous example, drawn from philosophy of mind.

Suppose you are a physicalist. You think that all mental phenomena are somehow grounded in physical phenomena. But not because you think that mental events are literally *identical* to physical events; rather, the latter "realize" the former. And realization is not identity: when a given mental event M is realized by, say, some neurophysiological event P, it *could have* been realized by any of many quite different physical events.

Along comes an argument that your view has untenable consequences (see for example Kim 1993). You would surely like to say, for example, that a certain mental event M that an agent undergoes causes some later physical behavior on her part. But then you are guilty of positing rampant causal overdetermination, for this physical behavior is – given the "causal closure of the physical" – *already* caused by M's realizer P. And that is unacceptable. It seems, then, that on pain of positing rampant

overdetermination, physicalists about the mind face a dilemma: either hold that the mind is causally inefficacious, or hold that mental events, properties, etc. are strictly identical to their physical realizers.

The right response is to ask, "What's so bad about rampant causal overdetermination?" If you thought of causation as some kind of metaphysical *juice*, such that for any thing to happen requires a certain amount of it, then you might have an answer: "Rampant overdetermination is implausible, because we should assume, in Razorish fashion, that the world contains no more juice than is necessary to get things to happen." But that answer relies on a confused metaphysical picture. Causation is, at bottom, localize counterfactual dependence of a certain sort. An agent's body moves at a certain time. The way that it moves can depend in all sorts of subtle and interesting ways on aspects of her prior neurophysiological state, and likewise on aspects of her prior mental state, even if the latter is not strictly identical to the former. These patterns of dependence are underwritten in a straightforward way by the fundamental physical laws – and, crucially, underwritten in a way that carries no suggestion whatsoever that one pattern of dependence will somehow metaphysically "crowd out" all others. There simply is no "exclusion problem".

§5 Problems and open questions

The account sketched here will only earn real plausibility once it is developed in detail. I want to close by mentioning two clear and serious problems that will need addressing, and drawing attention to some philosophically interesting issues about the sciences that these problems bring into focus.

§5.1. Causation in the social sciences

Consider counterfactuals used to describe features of social interactions. Billy and Suzy are lunching together. Billy is wearing a blue t-shirt. What would Suzy's reaction be, if he were wearing a green t-shirt? Sensible answer: Nothing; she doesn't really pay much attention to how he dresses. Crazy answer: "Oh my God! How did your t-shirt suddenly change color??" But notice that if we apply the Recipe to this example flat-footedly, we get exactly the crazy answer: make a localized change to the state of the world at the relevant time (i.e., switch the color of Billy's shirt from blue to green); leave the rest of the state of the world unchanged (which state includes Suzy's memory of seeing him wearing a blue t-shirt just moments before). Update in accordance with the laws.

Obviously, we need to add to the Recipe that, somehow, appropriate adjustments are to be made to the surrounding state of the world in cases like this. It is a wide open question how to effect this modification of the Recipe in a clear and rig-

orous fashion. But it is also an urgent question, if we wish to understand counterfactual and therefore causal reasoning in any of the social sciences.

§5.2. Late pre-emption and "no interference" clauses

Next, there are two problems that superficially appear unrelated, but that have, I think, deep interconnections. A proper working out of the approach to causation on offer here needs, I think, to solve these problems.

First, the idea that causal structure is localized dependence structure does nothing to capture the asymmetry that is so vividly on display in cases like Suzy First (where, remember, Suzy and Billy both throw rocks at a window, but hers gets there first). One option is to try to discern this asymmetry in subtle features of the localized dependence structure: for example, we might observe that the exact *time* at which the window breaks is a bit more counterfactually sensitive to the timing of Suzy's throw then it is to the timing of Billy's throw. For various reasons, I think that is a dead end (see Hall & Paul 2003 for details). A second option is simply to ignore the intuitive difference, reminding any doubters that our aim is to find a concept of causal structure of use in illuminating the structure of scientific inquiry, not a causal concept that does justice to ordinary intuitions. I hold out hope for a third option, which is to construct a more refined account of causal structure that does justice, in cases like Suzy First, to the very natural thought that what distinguishes Suzy's throw from Billy's, causally speaking, is precisely that a process of the right type connects her throw to the breaking, whereas no such process connects his throw to the breaking. Rather than try to go into details (for which see my 2004 and 2005), I will simply sketch what I think is the right approach.

We should understand "processes" simply to be sequences of events (not necessarily spatiotemporally contiguous). These can be compared with one another, and classified into types, on the basis of their intrinsic characteristics. Furthermore, some processes can be distinguished as ones that unfold from certain initial conditions, in highly sanitized circumstances in which *nothing else is happening* at the time that those initial conditions obtain.[16]

So, we can ask what a process looks like that begins with Suzy throwing a rock at the window, in circumstances in which Billy (and, for that matter, anything else in the environment whose presence might muck up the counterfactual relationships

[16] Making sense of this notion of "nothing else happening" will, I think, involve introducing a distinction between a *default* state of the world, and *deviations* from it. A situation in which the only things happening at a certain time consist in the obtaining of such-and-such conditions will just be a situation in which every part of the world save those involved in the instantiation of the conditions is in its default state. See Hitchcock 2007 as well as my 2007 for more discussion.

typically diagnostic of causation) is *absent*. Such a process will have certain distinctive features. Finally, it is because one can find, in Suzy First, a process sharing those features that connects Suzy's throw to the breaking – but no process sharing those features that connects Billy's throw to the breaking – that Suzy's throw counts as a cause of the breaking and Billy's does not.

Suppose this approach works. Surprisingly, the tools it deploys can be brought to bear on the seemingly very different problem of making sense of the "no interference" clauses that unavoidably adorn most generalizations in the special sciences. Metal bars, when heated, expand in proportion to their change in temperature. That is, *provided that nothing interferes*: If you set off a bomb next to a metal bar while you heat it, it will *not* expand in proportion to its change in temperature. Think about such examples for a few moments, and you will see that, outside of fundamental physics, it will be essential to tack on a "no interference" rider to any generalization if it is to have a hope of being true and informative. And that has raised a philosophical problem: How do we make sense of this rider? (See for example Earman et. al. 2002.) It's no help simply to observe that special science generalizations are generalizations about localized dependence structure. Yes; but as such, they still need the rider to stand a chance of being true.

Myself, I think we make sense of it as follows: If we have a generalization of the schematic form "if, at time t, conditions of type C obtain, then conditions of type E will follow at a later time t*", we should understand this generalization as making a perfectly definite claim about what would happen, if C-conditions obtained at t, and *nothing else were happening* at t. For if nothing else were happening, then, *a fortiori*, nothing else would be happening that could possibly interfere. And so the generalization is true just in case, in all such stripped-down conditions, E-conditions obtain at the appropriate time t*. Finally, I think we should add that in normal conditions – where, of course, *lots* of other stuff is going on – we have a genuine instance of such a law not merely if the C-conditions obtain at some time, and the E-conditions obtain at the appropriate later time, but if there is a process connecting the C-conditions to the E-conditions that shares the relevant features displayed by the process that would have unfolded in stripped down conditions – where, roughly, these "relevant features" are those in virtue of which this process has the counterfactual structure it does. Finally, call a type of process, individuated by such features, a *mechanism*. Then we can say not merely that the special sciences investigate localized dependence structure, but more specifically that they aim to discover and articulate the structure of such mechanisms.

§6 Conclusion

Let me close with a minor methodological moral. Philosophers who pursue an analysis of causation – especially, one that stresses the importance of ordinary intuitions – face a challenge: Why should anyone (except perhaps those interested in the semantics of the English word "cause") care whether they succeed? The challenge can be met, if the analysis yields a concept that has important and interesting connections to other things we care about. The connections between causation, counterfactuals, laws of nature, and the taxonomic schemes and generalizations of the special sciences – if they are as I have suggested! – provide a case in point.

Philosophers of science should take very seriously the possibility that philosophical investigation into the nature of causation – of a sort that straightforwardly deserves the name "metaphysics" – can provide them with valuable tools with which to probe and clarify the nature of scientific inquiry. Yes: such investigation should pay attention to what actually goes on in the sciences. But given that metaphysics, quite generally, aims to uncover basic facts about what reality contains, and what that stuff is like, how else should it proceed?

References

Anscombe, G.E.M., 1971. *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press.

Beebee, Helen 2000. "The Non-Governing Conception of Laws of Nature", Philosophy and Phenomenological Research 61: 571-593.

Cartwright, Nancy 1999. *The Dappled World*, Oxford: Oxford University Press.

Collins, John; Hall, Ned; and Paul, L. A. eds. 2004: *Causation and Counterfactuals*, Cambridge, MA: MIT Press.

Earman, J., J. Roberts, S. Smith 2002. "Ceteris Paribus Lost", *Erkenntnis* 57: 281-301.

Elga, Adam 2000. "Statistical Mechanics and the Asymmetry of Counterfactual Dependence." Philosophy of Science (suppl. vol. 68, PSA 2000): 313-324.

Hall, Ned 2000. "Causation and the Price of Transitivity," *Journal of Philosophy* 97: 198-222.

Hall, Ned and Paul, L. A. 2003: "Causation and Preemption," in Clark and Hawley (eds.) *Philosophy of Science Today*, Oxford: Oxford University Press.

Hall, Ned 2004: "The Intrinsic Character of Causation," in Dean Zimmerman (ed.), *Oxford Studies in Metaphysics, Volume 1*:255-300.

Hall, Ned 2005. "Causation and Ceteris Paribus Laws", Harvard Review of Philosophy 13: 80-99.

Hall, Ned 2007. "Structural Equations and Causation", Philosophical Studies 132: 109-136.

Hall, Ned 2010. "Humean Reductionism about laws of nature," ms.

Hitchcock, Christopher 2003: "Of Humean Bondage," British Journal for the Philosophy of Science 54: 1-25.

Hitchcock, Christopher 2007. "Prevention, Preemption, and the Principle of Sufficient Reason", Philosophical Review 116: 495-532.

Kim, Jaegwon 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.

Lewis, David 1973a. "Causation," *Journal of Philosophy* 70:556-567.

Lewis, David 1973b. *Counterfactuals*, Cambridge: Harvard University Press.

Lewis, David 1986a. *Philosophical Papers, Volume II*. New York: Oxford University Press.

Lewis, David 1986b. Postscripts to "Causation," in Lewis 1986a: 172-213.

Lewis, David 2004: "Causation As Influence," in Collins, Hall, and Paul 2004, *Causation and Counterfactuals*, chapter 3.

Loewer, Barry 1996. "Humean Supervenience", Philosophical Topics 24: 101-127.

Maudlin, Tim 2007a. *The Metaphysics Within Physics*, Oxford: Oxford Univ. Press.

Maudlin, Tim 2007b. "A Modest Proposal Concerning Laws, Counterfactuals, and Explanation", in Maudlin 2007a.

Norton, John 2003. "Causation as Folk Science", *Philosopher's Imprint* 3.

Pearl, Judea 2009. *Causality: models, reasoning, and inference (2nd ed.)*, Cambridge: Cambridge University Press.

Russell, Bertrand 1953. "On the Notion of Cause", in Russell, *Mysticism and Logic*, London: Penguin.

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

Spirtes, Peter, C. Glymour, R. Scheines 2001. *Causation, Prediction, and Search* (2nd ed.), Cambridge MA: MIT press.

Stalnaker, Robert 1968. "A Theory of Conditionals", in Nicholas Rescher (ed) *Studies in Logical Theory*, Oxford: Blackwell.

Strevens, Michael 2009. *Depth: An Account of Scientific Explanation*, Cambridge: Harvard Univ. Press.

Ward, Barry 2010. "Taking the Raven Paradox with a Grain of Salt", ms.

Woodward, J. 2005. *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford Univ. Press.