Causality and Determinism: Tension, or Outright Conflict?

Carl Hoefer
ICREA and *Universidad Autònoma de Barcelona*

Draft October 2004

Abstract: In the philosophical tradition, the notions of determinism and causality are strongly linked: it is assumed that in a world of deterministic laws, causality may be said to reign supreme; and in any world where the causality is strong enough, determinism must hold. I will show that these alleged linkages are based on mistakes, and in fact get things almost completely wrong. In a deterministic world that is anything like ours, there is no room for *genuine* causation. Though there may be stable enough macro-level regularities to serve the purposes of human agents, the sense of "causality" that can be maintained is one that will at best satisfy Humeans and pragmatists, not causal fundamentalists.

Introduction. There has been a strong tendency in the philosophical literature to conflate determinism and causality, or at the very least, to see the former as a particularly strong form of the latter. The tendency persists even today. When the editors of the *Stanford Encyclopedia of Philosophy* asked me to write the entry on determinism, I found that the title was to be "Causal determinism".[1]

I therefore felt obliged to point out in the opening paragraph that determinism actually has little or nothing to do with causation; for the philosophical tradition has it all wrong. What I hope to show in this paper is that, in fact, in a complex world such as the one we inhabit, determinism and *genuine* causality are probably incompatible with each other. After we see why this is so, we can appreciate better the different metaphysical options available to philosophers hoping to understand the complex issues concerning laws of nature, causality, and physical theory.

At first blush it should seem strange to be told that determinism and genuine causality are incompatible.[2] After all, let's consider a simple possible world governed by Newtonian mechanics, consisting of nothing but a box with rigid walls, and a few billiard balls inside colliding elastically with the walls and with each other. This possible world is surely

---

[1]Hoefer (2003), http://plato.stanford.edu/entries/determinism-causal/

[2]I will explain below what is meant by the modifier "genuine" here.

deterministic (none of the pathological elements of Newtonian physics are permitted, here)[3], and what could be a more standard example of cause-effect relations than that old cliché, billiard-ball collisions? Accepting these claims for a moment, what this shows is just that in very simple, idealized worlds, determinism may be able to coexist with genuine causality. My target is to consider a very different kind of possible world: namely, a world as rich and complex and apparently-irregular as ours is. We will see what problems are brought by this richness and complexity, below.

But surely there is more to say in favor of the compatibility of determinism and causality than just our toy billiard-ball-world example? In fact, isn't one precise way of presenting determinism this: If the world is governed by deterministic laws, then an[y complete] earlier state of the whole world may be considered a sufficient cause of a[ny] later state of the world, in all details. If this isn't genuine causality, then what is?

But this isn't genuine causality, in anything remotely like the senses of "X caused Y" or "P's cause Q's" that we employ in everyday life or in science. First, consider that in all reasonable examples of deterministic theories we have in hand, the determinism is bi-directional: *future* states of the world entail past events completely, as well as vice-versa. Yet it strains the notion of "cause" past the breaking point, to say that we should consider the state of the world 1 billion years from now as the *cause* of all our current events and actions. (Here, by the way, we see one reason why our billiard-ball world should be accepted as causation-bearing only with some hesitation: the physical laws give us no reason to think that "ball A struck ball B, giving it a new direction and momentum" is more correct than the time-inverse of the same process, "ball B struck ball A, causing it to change its direction and momentum".[4] If we, as human observers, were watching the billiard balls, we would surely single out one of these as the "correct" description (i.e., the one matching our internal sense of time-order). But then, with humans in it, this would no longer be a simple world, nor could it be one with Newtonian laws.)

Getting back to complete states of the world, we need not insist on bi-directional determinism in order to see why the relations between them are not relations of causation. For as we will see shortly, the relationships are ones of *entailment*, whereas causation is traditionally seen as a metaphysical but non-logical relationship. The temporal distance between states of the world can also be chosen to be quite large: the complete world-state of 1 billion years ago entails the events of today's world-state just as well as yesterday's does, yet it contains none of the facts or events that we might intuitively take to be causes of today's events. Finally, the relationships between complete states of the world lack a certain *robustness* that we normally attribute to cause-effect relations. If my phone call at 4:00 a.m. woke my sister (caused her to wake up), we normally suppose that my calling at 4:00:01 would have done the same, or that my calling in the

---

[3]We will help ourselves to global energy conservation (and any other extra laws we may need) to rule out determinism-wreckers such as space invaders. See Earman (1986) for details.

[4]Huw Price (1996) reminds us forcefully that physics does not seem to tell us that one of these descriptions is the "right" way of looking at the events; either one is as good as the other. I will largely avoid the issues of philosophy of time that arise if we explore these remarks further, however.

presence of a housefly on the wall near me (actually there was none) would have done the same, too. But if we add a housefly to the state of the world a billion years ago, or move some event by 1 second, the "butterfly effect" tells us that arbitrarily large changes in the world-state now might be the result. For all these reasons (and others I will not mention), it is not right to see the entailment relations between states of the world as cause-effect relations.

So determinism by itself is no form of, nor guarantee of, genuine causation in the sense that interests us. But are they not at least *compatible?* Over the next few sections we will see why they are not. Before that, we need to clarify both determinism and genuine causation.

> *Determinism:* The world is deterministic if and only if, given a specified way things are at a time t, the way things go thereafter is fixed as a matter of natural law.

"Fixed as a matter of natural law" means simply that the specification of how things are (everywhere) at time t, together with the laws of nature, jointly logico-mathematically determine a single possible future for the world. Notice that I have only defined the determination relation to be past –> future, so as to make the doctrine as weak as possible while still deserving the name. But we should bear in mind that the existing theories in physics that are deterministic, or close to it, are all theories in which the logico-mathematical determination works as well from later –> earlier as it does in the customary direction.

Determinism is only a conceptual possibility for our world if we take seriously that the world may be "governed" by fundamental, exceptionless laws of nature;[5] no-laws metaphysical views such as are held by Cartwright (1999) and other post-fundamentalists rule out determinism without further ado. Such philosophers will agree with me that determinism can't hold in a world like ours, and *a fortiori* cannot coexist with genuine causation in our world. But for the rest of us, both fundamental laws, and the idea that they may be fully deterministic, remain live metaphysical possibilities.[6] For us, then, what would these laws have to be like? Two features that the true laws seem likely to have will be important for what follows.

First, these laws will be mainly, if not totally, local and microscopic in character, stating constraints on the behaviors of the smallest bits of stuff that exist, and building up conclusions concerning macro-scale happenings out of those constraints. Field theories such as GTR and QFT have this character; so does statistical mechanics, in both classical and quantum forms. True, there may be some global constraints supplementing the local-interaction laws, e.g., perhaps a law that states that the total mass-energy, or the total angular momentum, in all of

---

[5]The laws must be taken seriously as *truths* concerning the world; whether they deserve to be said to "govern" is a separate issue, and need not be decided for our purposes here.

[6]I should mention that it is simply a popular misconception to think that the rise of quantum theories in physics has shown the falsity of determinism once and for all. On one important interpretation of non-relativistic QM, Bohmian mechanics, quantum mechanics turns out to be a fully deterministic theory of point-sized particles.

spacetime sums up to a certain value.[7] But the action in the laws, so to speak, will mainly be found at a local and microscopic level. By "local" here I do not mean to exclude such things as Newtonian gravity, taken as an action-at-a-distance, but rather laws that mention, or directly govern (so to speak) macro-scale objects, independent of their constituents. So, for example, the fundamental laws of nature are not likely to include things such as: "All ravens are black" or "No gold sphere is greater than 1 km in diameter." Such constraints as there are (most likely there are few or none) concerning the sizes of gold spheres and the colors of raven plumage will arise out of laws governing basic micro-kinds of things, together with historical and contingent conditions built into the meaning of expressions like "gold" and "raven".

A more serious example of a non-micro, non-local law would be the second law of thermodynamics. This brings us to the second feature of possible deterministic laws for our world that needs to be mentioned. The 2nd law and other "phenomenological" laws have tended to be seen, since the definitive victory of atomism (in its various quantum guises), as merely expressing overwhelmingly likely behaviors in certain circumstances, rather than behaviors mandated by fundamental law. We believe that entropy-reducing evolutions do occur sometimes for small, isolated systems, and we believe larger-scale anti-entropic behavior to be possible, if unlikely to be observed.[8]

In general, our two features to keep in mind say: macro-scale phenomena emerge from micro-scale initial conditions and interactions. And while certain dependable regularities seem to emerge at the macro-level, in fact what will occur in any given case depends on the *precise* initial conditions and external (boundary) influences that occur; and if the initial and boundary conditions are *just so*, many rather weird and uncanny outcomes may be determined and may occur, without any breakdown of determinism's grip.

So much for determinism and deterministic laws. What do I mean by *"genuine* causation"? The full answer to this is to be found in the next three sections, where I look at what I take to be the three most prominent accounts of causality that deserve to be called "genuine". At a minimum, genuine causation should include the paradigmatic cases of causal talk in everyday life, not merely events at the level of particles or spacetime points. And the point of talking about *genuine* causation is, of course, to contrast it with some lesser sorts of causation, metaphysically innocuous and, correspondingly, obviously applicable to events in our world. For example, imagine that a deeply empiricist philosopher says "Of course there is causation in our world, for all that is required for causation is a bit of semi-dependable regularity and the interest of some human onlookers. Basically, 'C causes E' is true if C's are usually followed by E's, and humans can sometimes bring C about as a tool for bringing about E." If this is all that one means by 'causation' or 'causality', then these notions clearly can exist in harmony with determinism – at least, if determinism is in fact a possibility for our world, as we are assuming.

---

[7]Such laws have been toyed with for variants of General Relativity, most commonly setting the total-stuff quantity to zero.

[8]The justification for saying that such behaviors are "unlikely" is, of course, a key unresolved problem for the philosophy of statistical mechanics.

Genuine causation is something stronger than roughly reliable regularity; it is meant to be metaphysically substantive. In particular, it must involve a non-trivial degree of *modal involvement* to the meaning of causal claims. But exactly what sort of modal involvement? Below we will look at three sorts of answer to this question, and check whether causation, so understood, can coexist with determinism. Before beginning, there remains just one other view of causation to mention in passing: *non-reductive causal fundamentalism*. This is the view that there is genuine causation, that it is a fundamental part of the correct metaphysical world view, and that it cannot be reduced to or analyzed in terms of any other notions. In particular, this sort of fundamentalist will resist even the minimal conceptual commitments we will see below: for example, a commitment to the claim that whenever a genuine cause C occurs, at least the probability of E must be thereby increased; or that, given that *c* caused *e*, then it is true that if *c* had not occurred, *e* would not have occurred. What philosophers subscribe to such a view of causation? I will not name names. But I fear that the ranks of such philosophers are growing, in part because of the failures of all philosophical analyses of causation proposed to date. Against these philosophers, my argument will not be able to gain a foothold. This is not surprising, for in the view of such philosophers, the causal relation is a kind of magical linkage, something we allegedly recognize all over the place, but which resists translation into any other conditions, terms or notions. Rather than indulging in mystery-mongering, I would urge these philosophers to adopt an appropriate anti-realist attitude towards causation. But that is a battle for another day.

Genuine Causation: Mackie's analysis. There is a centuries-old tradition of trying to understand causation in terms of either necessity, or sufficiency, or both. This tradition reached its maximum development in John Mackie's *INUS-condition* analysis of causation.[9] The analysis is meant to capture a non-probabilistic, traditional notion of causation: given that the cause occurs, the effect *has to* follow; or, as we often also think, *had the cause not occurred,* then the effect would not have ensued (other things being equal; the cause was *necessary* in the circumstances). Now Mackie himself was an empiricist, and at one stage at least he wanted to translate these modalities into empiricist-friendly facts about what happens *always* or *never*, in the actual world. I would argue that the attempt to make the account empiricist-friendly failed, but we can set aside that issue for our purposes. As I noted above, we are only interested in views of causation with genuine modal involvement. Since we are assuming the existence of laws (deterministic laws, in fact), we may as well say that the necessity here is *physical necessity*: i.e., for the effect not to have followed would have involved a violation of one or more law of nature.

An *INUS* condition is an *insufficient, but necessary* part of an *unnecessary, but sufficient* condition for a given event-type to occur. Let's take a simple traditional example like: striking the match caused it to light. This is not the only way to get a match to light; one can also insert it into an existing flame, for example. So while striking the match (plus other factors) may be

---

[9]Mackie (1965), (1974).

sufficient for the lighting to ensue, it is not necessary; it is one of several cause-complexes (i.e. conjunctions of events and/or circumstances) that are each *unnecessary but sufficient* to bring about the lighting of the match. Focussing on the striking itself, a moment's thought reveals that rubbing the match along the strike-board is not sufficient all by itself. If we do this in a vacuum chamber, no lighting will occur. So there must be oxygen around. And the match had better be dry. And so on: the true, sufficient cause is a long string of jointly-occuring circumstances or events, with the striking being just one member of the list – albeit, a necessary one. Let's not overwork our imaginations, and let's suppose that a proper sufficient condition for the match lighting is this: <strike, dry, oxygen present>. Each of the three elements here is an INUS condition for the match lighting – and indeed, in the right context at least, it is possible to imagine each of them as being singled out, for explanatory purposes, as *the* cause of the match's lighting.

So far, so good; but philosophers do have very good imaginations. The oxygen had better not be present in the form of a high wind; the air pressure had better not be too low or too high; the match had better not be *about to be doused by water, as soon as the striking ends*, and so on. There would seem to be an unlimited number of strange ways in which the three elements of our putatively-sufficient condition might be present, and yet the striking fail to produce a lit match. So we face an ugly choice. Either we accept that the true US conditions are arbitrarily long and not ever fully specifiable by us – which seems to render belief in genuine INUS conditions almost a matter of faith; or we lump together all the weird absences-of-preventers that we need, and say they are excluded by the "causal field", which is a catch-all condition that ensures that nothing abnormal is present, about to occur, etc. The catch-all is a problem for the analysis: it cannot be taken literally as meaning "and nothing that is a preventer of E is present", since 'preventer' is a fully causal notion, and its introduction here makes the account circular. But if instead we take the catch-all to be simply an element saying that "circumstances are normal", then one worries whether something this vague can be admitted as part of the analysis. (One also worries, of course, that "normal" here is just a sneaky way of excluding preventers without saying so outright.)

This is all fairly familiar material so far – but how does determinism affect the picture? The answer can be seen after we divide reasons why a putative US condition may fail into two types of enemy: the enemy from without, and the enemy from within. The kinds of problems we have dealt with so far, and which Mackie's causal field are meant to exclude, all count as enemies from without. But if microscopic-level determinism reigns, then in many cases we need to exclude the enemy from within: microscopic initial- and boundary-conditions that are just perverted and "atypical" enough to entail the non-production of the usual effect.

Philosophers of statistical mechanics talk about these sorts of things quite often. Given just the right initial and boundary conditions, that ice cube in your gin and tonic may grow bigger, and the liquid hotter. Or the end result of stirring the drink may be to leave the gin and the tonic in separate layers, as a matter, we might say, of "accident" or "bad luck" – but in fact, as a simple consequence of the initial conditions and boundary conditions that held when you began stirring. These sorts of things are fantastically unlikely, and we probably don't see any in our lifetimes. But the conceptual point still stands: because of micro-determinism of the type

likely to hold in our complex world, in most cases *there are no US conditions for a given macro-level effect, and hence no INUS conditions.* If the macro-level US condition proposed supervenes on the "wrong" initial conditions, and equally "wrong" boundary conditions obtain in the relevant environment, then all sorts of weird things may take place, including the failure of the customary effect E to ensue.

A few points are worth noting here. First, the anti-thermodynamic behavior examples mentioned above are just examples of what the "wrong" initial conditions can produce; the undesired outcome need not be a violation of the 2nd law, and indeed, may involve domains where the categories of thermodynamics are not applicable to begin with. The anti-thermodynamic examples are just one salient type of problem that bad initial conditions may produce, one already widely discussed in the philosophy of physics. Second, it is also not clear that the mischief that bad initial conditions can cause is always highly improbable, unlikely to be observed in our lifetimes. Going back to the match-striking: how often have you struck a match, watched sparks fly, and watched as the match fails to light? Often, if you buy the same brand of matches as me! This can happen when the match is dry, when there is no wind, when the force of the strike is sufficient (judging by other occasions when it was), etc. Why does the match sometimes not light? Perhaps the only real explanation, for a given occasion, will have to advert to what happened at the microscopic level. The initial and boundary conditions just weren't such as to entail the match lighting up; and moreover, that might be true of about 30% of the strikings *that share an identical macro-level description.* Why didn't the US condition lead to the effect? The only answer may be "the enemy from within".[10]

If this is right, no band-aid can be brought in to fix the account, as we did when we invoked the causal field to exclude the enemies from without. Suppose we add a secondary causal field assumption: ". . . and the initial conditions and boundary conditions underlying all these factors are not such as to lead to E's non-occurrence." The circle is now complete! In fact, we can throw away C and its mates in our original proposed US condition, and just keep the final clause: assuming deterministic laws, if the local initial and boundary conditions are such as to *not entail the non-occurrence of E*, then these conditions actually *entail the occurrence of E*. But then our original macro-level elements like <strike, dry, oxygen present, . . .> are no longer INUS conditions, because they are not *necessary* after all. In short, the INUS condition analysis has broken down completely. If genuine causation exists in our world, its nature must be explicated some other way, for genuine INUS conditions and micro-deterministic laws are incompatible.

Returning to the match-striking example, one thing that might occur to us immediately is that the problem I have highlighted for Mackie's analysis is precisely the sort of problem that drove many philosophers, in the early 1970s, to develop accounts of *probabilistic* causation.

---

[10]I realize that a defender of the INUS condition analysis might not give in so easily, and might insist that there are (in principle) specifiable macro-conditions that fail to obtain when my match doesn't light. But it is hard to see any plausible argument to back up this claim, and hard to imagine what the description of the lacking macro-condition would be like.

Striking the match doesn't guarantee its lighting, no matter how you specify the macro-conditions; but surely it at least raises the probability of lighting occuring? We will come back to this below as our third and final type of account of genuine causation. Before going probabilistic, we must explore a different sort of account, meant to be compatible with determinism: Lewis' counterfactual analysis.

Lewis' analysis. The original Lewis (1973) account of causation can be thought of as a schema for analyses based on counterfactuals; the details of what counterfactuals are, what their truth conditions are, how to analyze direct versus indirect causation, singular vs generic causation, and so on, are not fully fleshed out. The fully fleshed out Lewisian story about causation arguably does not give us genuine causation, because it is too Humean-empiricist. Lewis' official ontological picture insists that the panoply of events that occur in world history are all fully "distinct existences", *a la* Hume, and that the facts about laws and about causation strongly supervene on this panoply of *intrinsically* disconnected facts. But we do not need to enter into arguments here about whether Lewis' ontology ultimately supports genuine laws, and genuine causation, or not. As with Mackie above, since we are already assuming the existence of genuine, deterministic laws of nature, we can also help ourselves to the modalities and counterfactuals that rest on these laws.

Here is a greatly oversimplified statement of Lewis' account of causation. If *c* and *e* are both events that actually occur, and it is true that if *c* had not occurred, then *e* would not have occurred either, then *c* is a cause of *e*. If *c* and *e* are not actually occurring events, but it is a true counterfactual claim that *if c* had occurred, then *e* would have occurred, then again we can say that *c* is a cause (would have been a cause) of *e*.

Lewis' account has great surface plausibility. Very often, the reason we say that some thing *c* caused another thing *e* is precisely because we judge that, had *c* not occurred, then *e* would not have ensued. How do we judge this? Typically, we try to imagine what the situation would have been like without *c*, but with everything else held as near to the way it actually is as possible, and then consider whether in those circumstances, the unfolding of events in accordance with the laws of nature would have led to *e*'s not occurring.

But in making these judgments, we do not grind through calculations using a putative set of deterministic differential equations plus initial and boundary conditions. Instead, we use common sense and physical intuition, hoping that our conclusions do indeed track what the true, deterministic laws entail. The problem, then, is this: the deterministic laws, in most cases, probably do not entail a single possible outcome! The enemy from within will strike us here too. If the initial and boundary conditions of our hypothesized *not-c* situation are *just so*, the laws may entail that *e* comes about anyway – at least, for a wide range of cases we typically classify as cases of cause/effect. And turning to the case where *c* and *e* do not actually occur, it will be even more obvious that in certain cases, if the microscopic initial and boundary conditions are "wrong ones", then in the counterfactual situation in which *c* does occur, *e* will fail to occur.

In other words, if the facts about which counterfactuals are true and which false are based on what the deterministic microphysical laws entail, then they do not give us unequivocal

results very often. There will be a set of possible micro-states compatible with our *not-c* posit that lead to *not-e*, and a different set of micro-states, equally compatible with *not-c* and presumably macroscopically indistinguishable from the other set, that lead to *e*. Intuitively, we want to say that the latter set will be a set of "less likely" conditions, or that it will have "measure zero" compared to the normal-results set. Below we will come back to whether this response can save the day; first, let's consider a couple of examples.

Let's return to our example of the match that lit. Suppose it had not been struck – surely it is clear that it would not have lit? At the level of intuitive common sense, yes; at the level of deterministic microphysics, the answer is not so clear. Spontaneous combustion is known to happen, and although it never happens to ordinary matches in our experience, this "never" is presumably much like the "never" of "ice never grows larger in a drink". Moreover, we may consider a context in which a different causal claim is at issue. After many failed attempts to strike a match, my daughter finally does it firmly enough and gets a match lit. I say: "See: you have to strike with enough force! If you had struck the match a bit less forcefully, it wouldn't have lit." I thereby signal that I think the rough strike-strength as having caused the match to light, in contrast with weaker strike-strengths that have been employed on previous occasions. But in this case, our enemy from within doesn't need to make the match light spontaneously, but rather only make the match light from a less-than-actual-strength strike. The set of microscopic initial conditions and boundary conditions that would lead to *that* happening may be a very big set indeed. Or to change examples, suppose I see a feather delicately balanced on the edge of the table, swaying like a see-saw. A sudden small gust of wind blows it off the table. I say to my friend that the gust of wind caused the feather to fall. But here too, it is not so clear that the enemy from within might not have done exactly the same job, without the gust being present. Perhaps only very normal, tiny micro-gusts, such as happen all the time in a room of "still" air, would have sufficed, and lots of possible "no-sudden-gust" initial and boundary conditions would have led to precisely that outcome.

My claim is that these enemy-from-within examples show us that a world governed by micro-deterministic laws is a world where all sorts of outcomes are compossible with a specified *macroscopic* set of fixed conditions. And as for Mackie's, this fact causes big problems for Lewis' analysis of causation. The intuitively correct counterfactual consequence is only guaranteed if we restrict the counterfactual micro-conditions allowable to those guaranteed to deliver the right result; but that is cheating. The only general assertion we can make about what *would* have happened, had certain macro-conditions been different, is this: lots of different things *might* have occurred, and which would occur depends completely on the detailed micro-conditions with which we replace the actual situation.

Let's now return to the objection that while these enemy-from-within cases are indeed nomologically possible, they are so rare, improbable, unlikely etc., that we can set them aside as irrelevant to the truth conditions of our counterfactuals. There are two problems with this reply. First, I tried to give cases where it is not at all clear that the "enemy" initial conditions are greatly less likely than the "normal" ones (the softly-struck match), or where in fact it may be that the "enemy" conditions arguably are more numerous than the "correct-result" conditions (the feather case). The latter is perhaps a borderline case of overdetermination, and a defender

of Lewis' analysis might be prepared to agree that there is no clear causal fact in that case.

More to the point, though, I don't think that taking refuge in the "unlikeliness" of certain sorts of initial and boundary conditions is automatically meaningful. It is true that for many models of deterministic physical systems (notably, a gas of Newtonian particles in a box), there seems to be a natural measure to impose (or a range of suitable, natural measures) on volumes of phase space, and if we then equate volume with probability, we get the desired result that weirdness-producing initial conditions are "unlikely". But great care must be taken in using the everyday notion of probability in such abstruse, many-worlds scenarios. Are we to imagine God throwing some quadrillion-sided dice to decide which initial and boundary conditions to use, when setting up our *not-c* world? Many philosophers, myself included, object to the exportation of the probability concept to extravagant metaphysical domains outside of the actual world.

Furthermore, Lewis' own definition of the truth conditions for counterfactuals does not square easily with this proposed fix of the problem. Lewis says:

> $A \mathrel{\square\!\!\rightarrow} C$: the proposition that if A were true, the C would also be true. The operation $\mathrel{\square\!\!\rightarrow}$ is defined by a rule of truth, as follows. $A \mathrel{\square\!\!\rightarrow} C$ is true (at a world w) iff either (1) there are no possible A-worlds (in which case $A \mathrel{\square\!\!\rightarrow} C$ is *vacuous*), or (2) some A-world where C holds is closer (to w) than is any A-world where C does not hold. In other words, a counterfactual is nonvacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent. (1973, p. 560).

Evidently, to get the right counterfactuals coming out true in the face of the enemy from within, we will have to maintain that some "normal" world is closer to actuality than is *any* "enemy" world. But closer in what way, and for what reason? By hypothesis, they are macroscopically identical to each other, up to the point of the non-occurrence of *c*. We can't say that the enemy world is further away from reality because it involves a very improbable transition taking place; the actual world, and pretty much any possible world like it in complexity, is a world where very improbable transitions take place.

I don't want to explore the argumentative options further here, and will rest my case hoping to have shown that there is at least a tension between the counterfactual account of genuine causation, and the way events may unfold under deterministic laws. Stepping back to view the problem at a distance, we may put it like this: in a world where micro-determinism reigns, specifying the presence or absence of a macroscopic condition doesn't entail much, if anything, about what subsequent events may look like.[11]

---

[11]It might appear that a natural move at this stage, to save a counterfactual analysis, would be to back away from $-c \mathrel{\square\!\!\rightarrow} -e$ as the crucial counterfactual, and instead say something like: $-c \mathrel{\square\!\!\rightarrow} P(-e) > P_{@}(-e)$. See for example Schaffer (2001). I will briefly touch on the kind of problem such an account would face, in the next section.

Probabilistic causation. Just as philosophers turned to probabilistic causation in the 1970s after the failures of traditional accounts, so do we here, to see whether a less rigid notion of causation fares better in terms of compatibility with determinism.

At first blush it may strike some as strange to even spend time discussing a form of causation that is obviously in contradiction with determinism from scratch. And it is certainly true that *some* kinds of probabilistic causation that might be contemplated are incompatible with determinism. A view that takes probabilistic causation to exist if and only if there are genuinely probabilistic, i.e. stochastic fundamental laws of nature, would be one such. But that is not the only way to understand probabilistic causation,[12] and we should not assume that determinism entails that there are no non-trivial objective probabilities in the world.[13]

Strongly empiricist views can be ruled out straight away, however, as not offering genuine causation. If one takes probabilities to be simply reasonably-stable actual frequencies, and takes Prob(Contract Malaria|Sprayed Swamps) > P(Contract Malaria|Not-Sprayed Swamps) – perhaps along with some other conditions designed to rule out spurious correlations – to be the kind of thing we mean when we say that spraying swamps causes a reduction in malaria, then one has a view that is perfectly at home in a deterministic world. What such a theory is, of course, is the probabilistic extension of a naïve Humean regularity account of causation, and like its brother, it is not what most philosophers would call a robust, genuine form of causation.[14]

So if we are to have *genuine* probabilistic causation that is compatible with determinism, we want our probabilities such as P(E|C) to be appropriately modal and non-accidental. In other words, we need to postulate something like stable *propensities* or *causal capacities* that are possessed by certain objects or systems, propensities or capacities that supervene on the micro-physics (and contingent-but-stable features of the world), without necessarily being explicitly reducible to those features. I will break the discussion of this idea into two parts: first, assuming that these causal capacities have a fixed, intrinsic *strength*, at least if the causal conditions are adequately spelled out; and second, assuming that the capacities do not have fixed and stable

---

[12] Nor even an obviously interesting way: how do we build up from stochastic fundamental laws to probabilistic causes and effects at the macro-level? If something like ordinary quantum mechanics is true, do its probabilities have any obvious relation to the probability of getting headache relief if one takes an aspirin?

[13] I argue this point at length in Hoefer (unpublished).

[14] I choose this example because it is one used by Nancy Cartwright in her famous paper "Causal Laws and Effective Strategies". In that paper Cartwright does advocate a simple frequentist account of probability, but the point of her argument is to show that the "other conditions" needed to rule out spurious correlations cannot be captured within the constraints of Humean empiricism. Extending the view and arguments in subsequent works (esp. (1989)), Cartwright later appears to advocate a form of irreducible singular probabilistic causation (one of the *irreducible fundamentalist* views mentioned earlier). For some problems with her 1983 arguments, see Hoefer (2004).

strengths.

The former sort of account claim will claim that it is true in general that P(E|C) > P(E|-C), at least in a range of contexts where other causal factors for E are held fixed; and for *at least one* (perhaps more, perhaps for all) clearly expressible type of causal set-up, there is a determinate probability *x* for the production of the effect.[15] We can use another cliché'd example here, such as: P(lung cancer within 10 years|smoke 2 packs a day) = 0.232.[16] Or, to return to our matches: P(lights|brand X, dry, no wind, strike-strength L) = 0.5. In general: P(E|Full-C) = *x*.

Unfortunately, in a world of micro-determinism, the sensitivity of outcomes to particularities of the micro-level initial and boundary conditions is in conflict with the hoped-for stable strength of the causal capacities. Both the enemy from within and the enemy from without are relevant here. Take the latter sort of problem first. If there are enough microscopic black holes whizzing through space, and if they happen to frequently collide with match-heads as they are being struck, the rate (actual frequency) at which striking will produce a lit match may be different than what it would be with few or no such collisions.[17] But we don't need exotic things like black holes; micro-gusts of wind will do just as well. The moral here is that the actual frequencies of E-outcomes after C-conditions may depend on a lot of factors and influences extrinsic to what we normally take to be the relevant parts of the C-conditions. As we saw in discussing the problems for Mackie's analysis, it may be unrealistic to suppose that there is, in principle, a completable list of factors that should be excluded. In other words, "Full-C" probably does not exist, in most cases.

But let's suppose that we are allowed to stipulate a proper-shielding condition that takes care of such unwanted influences; the actual frequencies may indeed fluctuate, and fail to be "right", when such influences are operative. Then the account of probabilistic causation we are looking at claims that, for systems properly shielded, when C obtains there is a definite, objective probability of E being produced – the "true" probability, as opposed to the actual frequencies that may obtain in our improperly shielded world. Here the enemy from within strikes, undermining the notion that any such "true" probability P(E|Full-C, Shielding) really exists. We know that in general, it is the actual micro initial conditions and boundary conditions

---

[15]What we have in mind here is something like one of Mackie's US conditions, but now instead of being sufficient for the ensuing of the effect, the set of conditions is sufficient to *produce E, with objective probability x*.

[16]The number here is made up, so smokers need not panic. On second thought, they should go ahead and panic.

[17]The idea here is that these collisions occur, in fact, more frequently than do collisions with other similar-sized objects, and for no particular reason (i.e., as a contingent accidental by-product of the actual initial and boundary conditions). If you don't like micro black holes, substitute alpha particles or other invisible death rays.

that determine, on a given Full-C-occasion, whether E results or not. So for some range of IC/BCs an E-outcome occurs, and for (most of) the rest of the possible IC/BCs, E does not occur, and in our world members of the former set occur, stably and reliably, in about *100(x)%* of Full-C cases.

It is clear, though, that this proportion could have been different, with exactly the same laws and exactly the same "causal facts" (other than the alleged strength of Full-C) had the contingent facts about IC's and BC's been different. Even more easily, the frequency distribution of E given Full-C might have failed to be stable and reliable. And yet, evidently, nothing about the nature of C, or E, or the real relationship that sometimes connects them, would thereby be any different. In other words, the alleged genuine causality we set out to explore, has turned out not to be genuine after all: it has no modal robustness, and in the end turns out to be just another empiricist-acceptable variant of the rough-regularity account of causation.[18]

We can deal with the second sort of account, that does not even postulate a given *intrinsic* strength to the probabilistic causation relation, quickly: it begins where the previous account was forced to end up. Without fixed, *true* probabilities of E given Full-C's, probabilities that somehow follow from the nature of C and E (and perhaps other factors, such as the laws), what this account of causation says is basically this: C's sometimes help give rise to E's, though there is no further clarification of the relationship that we can give, except perhaps on a case-by-case basis. Such a position, plausible in the extreme, is not the sort of *genuine* causation that we set out to explore.[19]

---

[18]It may be supposed that one can appeal to a "natural" distribution of IC's and BC's on something like thermodynamic & Stat Mech grounds, to shore up the idea that a certain probability *x* emerges naturally from the physical situation of shielded Full-C, whereas other values could only result and be stable in virtue of "weird coincidences" or "conspiratorial IC's" and suchlike. For simple problem set-ups like shielded coin-flippers or gases in boxes and so forth – i.e., for problems that have a uniform *micro*-description that we can handle with physical theory – this claim will sometimes be plausible. But unlike "rigid rectangular box of volume V with a Newtonian gas of identical particles in a M-B distribution at temperature T", "Smokes 2 packs a day, . . ." has no canonical distribution of micro-descriptions; and the boundary conditions (i.e. external influences, at the micro-level) are even less plausibly regimentable. There is just no reason to suppose that one set of micro-states of *males-smoking-2-packs-daily+environment* counts as "normal", "probable" or what have you, while another one does not.

[19]We arrive here at the doors of a *mechanism*-based account of causation (see Glennan (1996)). Since mechanisms work because of the dependable relations of their parts or sub-processes, and these are either (a) further mechanisms, (b) brute causal relationships, or (c) reducible/derivable from the laws of nature, the logical terrain for our purposes does not change at all. Mechanisms only salvage genuine causation if their smallest sub-parts end up involving brute causal relationships, and if those relationships are to count as genuine causation, presumably they must be understood in something like one of the ways discussed in the text

Conclusion. My goal was to show that *genuine* causation – causation of a robust metaphysical type, with some genuine modal component, and applicable at the level of ordinary events – is either in tension with, or in outright conflict with, the supposition of underlying determinism. Perhaps the tension can be overcome with an elaboration of a Lewis-style counterfactual account, perhaps using counterfactual probabilities rather than certainties; but the considerations of the last section go against such a hope. In a world of micro-determinism, there may be objective probabilities, but they are hostage to the facts about distributions of initial- and boundary-conditions, and this undercuts their ability to serve as the foundation of *genuine* causal relations.

We saw that standard analyses of causation face problems of the enemy from without sometimes, and of the enemy from within always. In a rich and complex world such as ours, with few or no macro-regularities of a robust and exceptionless kind, the relations between macroscopic things or circumstances supervene on the actual micro-facts, and so those relations can be (fairly) reliable, useful, controllable and so forth – but they can't be metaphysically fundamental.

Throughout, I have been looking at views of causation that locate it, at least in part, at the macro-level: gusts knocking down feathers, smoke causing cancers, and so on. But the defender of causation could accept the fragility of (what we call) causal relations at the macro-level, and yet insist that there is indeed genuine causation in our world – in the fundamental laws governing the micro-things, themselves. This is a view strongly supported by Nancy Cartwright's work on laws in physics: the Schrödinger equation tells us, among other things, how the capacities of a charge-bearing object work (say, a proton) *cause* a distinct charged object (say, a captive electron) to behave. In general, looking at the fundamental physical theories we have in hand, Cartwright and others argue that they are through-and-through stories about the causal powers of things. Other philosophers, such as Huw Price and myself, are uneasy with such a construal.

Fortunately or not, this debate cannot be settled using current knowledge. Are there really any *true, exceptionless* fundamental physical laws? If so, are they mere regularities, or necessary truths, or statements of causal powers? These questions can only be answered if or when progress in physics clarifies whether a true fundamental physics is possible for our world. In the meantime, we have good reason to suspect that *if* such laws do exist, and are deterministic, then genuine causation at the level of ordinary events is at best a fragile, pragmatically useful epiphenomenon.

---

above.

Bibliography

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

Cartwright (1989) *Nature's Capacities and Their Measurement* (Oxford University Press).

Cartwright (1999) *The Dappled World* (Cambridge University Press)

Earman (1986) *A Primer on Determinism* (Dordrecht: Reidel publishers).

Glennan (1996) "Mechanisms and The Nature of Causation," *Erkenntnis* **44**: 49-71.

Hoefer (unpublished) "The Third Way on Objective Chance", available on request.

Hoefer (2003) "Causal Determinism", *Stanford Encyclopedia of Philosophy*, plato.stanford.edu

Hoefer (2004) "Humean Effective Strategies", *Proceedings of the 12th International Congress on Logic, Methodology and Philosophy of Science* (ed. Petr Hajek, Luis Valdes-Villanueva and Dag Westerstahl). Elsevier North-Holland Press. Manuscript available at philsci-archive.pitt.edu

Lewis (1973) "Causation", *The Journal of Philosophy* **70** (1973): 556 - 567.

Lewis (1994) "Humean Supervenience Debugged", *MIND* **103** (1994): 473 - 490.

Mackie (1965/1993). "Causes and Conditions." *American Philosophical Quarterly* **24** (1965): 245-64. Reprinted in: Sosa and Tooley (eds.). *Causation* (Oxford 1993). 33-55.

Mackie (1974) Mackie, J. L., *The Cement of the Universe: A Study of Causation*. (Oxford:

Clarendon Press)

Price, Huw (1996) *Time's Arrow and Archimedes' Point* (New York: Oxford University Press).

Russell (1912) "On the Notion of Cause", Aristotelian Society address, reprinted in *Mysticism and Logic* (1917).

Schaffer (2001) "Causes as Probability-Raisers of Processes", *The Journal of Philosophy* **98** (2001): 75-92

Suppes (1970) *A probabilistic Theory of Causality*, (Amsterdam: North-Holland).