Russell, Causation, Determinism[*]

ANTONY EAGLE[†]

Abstract Two arguments due to Russell are examined, and found to show that the notion of causation as full determination doesn't mesh easily with deterministic global physics and the distinction between effective and ineffective strategies. But a local notion of causation as involving a certain kind of counterfactual dependence is, I argue, compatible with Russell's conclusions. I defend it from a resurgent form of Russell's microphysical determinism argument by making some mildly contentious claims about the autonomy of the events posited by the special sciences.

1 Russell's Arguments

Russell (1913) takes the relation of causation to be a relation of determination: c causes *e* just when c determines *e* to occur. This relation is supposed to be asymmetric and plausibly transitive as well. The fundamental law of causality is supposed to be that every event has a sufficient cause, one that is guaranteed to bring that event about and in fact did so. This intuition about the deterministic nature of causation is not a Russellian idiosyncrasy: it originates in Hume's 'constant conjunction' regularity analysis (if c and *e* are *constantly* conjoined, the appearance of c should be sufficient for the appearance of *e*), even later accounts like Suppes (1970) keep the idea that individual causes partially determine their effects, and deny that every event has a sufficient cause to avoid the supposed 'universal law'.

Russell thinks that this notion of causation as a determination relation between events doesn't appear in physics, and hence should be jettisoned from a properly scientific world view. (Perhaps there might be some pragmatic sense in which causation is useful, but there are no deep metaphysical truths about causation.)

Field (2003) identifies two arguments in Russell to this conclusion. The first rests on the claim that the equations of microphysics are bi-deterministic. If we fix the microphysical state s of some system R at t, that fixes the whole trajectory of R through the space of states both before and after t.[1] If all macroscopic events are constituted by

[*]This is a draft version—please do not cite without permission. Originally presented at *Causal Republicanism*, Centre for Time, University of Sydney, 16/7/03. Thanks to an audience there, in particular Chris Hitchcock, Helen Beebee, Huw Price, Daniel Nolan, Graham McDonald, Charles Twardy, Mathias Frisch, Jason Grossman, Adam Elga and Toby Handfield.

[†]Philosophy Department, Princeton University.aeagle@princeton.edu

[1]Earman (1986).

some particular microphysical states, then fixing a particular event in the system at a time will determine which microphysical trajectory the system is on, and hence which events will occur and have occurred. Then any event determines both its temporal antecedents and temporal succedents. But if c causes *e* in the sense of determination, then by this argument *e* equally well causes c. The asymmetry of determination, and hence the asymmetry of cause and effect is lost. There is no place in bi-deterministic physics for the causal asymmetry.

But this argument has at least two flaws. Firstly, though it might be true that the causal asymmetry isn't an asymmetry of determination, it still might be a relation of determination and an asymmetrical relation, where the asymmetry comes from somewhere else. Perhaps the asymmetry is that of entropy increase, plausibly set by unusual low entropy initial conditions.[2] Perhaps the asymmetry is temporal.

Secondly, it may still be the case that there is a macroscopic asymmetry between causes and effects. Perhaps on a global scale, the whole state of the universe at one time determines the whole state at every other time. However, if we restrict our attention to a local particular event c, presumably there are many global states that can have this event as a part. Perhaps all the global trajectories which feature this event in some state have some further event *e* as a feature of a state; but perhaps not all the global trajectories which feature *e* have c as a feature. So the occurrence of c determines the occurrence of *e* in a way that *e* doesn't determine the occurrence of c. Focussing attention on a limited area might very well give us an asymmetry of determination between particular events. Moreover, this local determination is exactly what the original notion of a cause was supposed to capture.

The notion of a local event is a tricky one. Everyone takes it involve spatiotemporal location within my nearby area. But I want something more than that. I want something like 'epistemically local': macroscopically describable, medium sized, readily distinguishable from other events going on around it (so discrete). I take it to be a feature of our commonsense causal language that effects and causes are both of this category.

Russell's other argument picks on this conception of local causation. Consider some small local events c and *e* such that the occurrence of c determines the occurrence of *e* but not vice versa. Russell's argument is that these 'local' events won't be the kind of things we typically take to be related by cause and effect. Consider a putative causal relationship between the firing of a gun F and the death of a victim D. These events are clearly too particular to get into the determination relation, because F can occur without D, if the bullet misses the victim, or if someone else intercepts the bullet, or if the bullet explodes harmlessly in mid air. The problem with local determination is that there is always the chance of some 'interference' from outside the local area at the

[2] Albert (2000).

time of the cause. To ensure that the cause guarantees the occurrence of the effect, we shall have to hold the cause to be a very large event, perhaps the whole past cone of potential causal influence on the effect. So if c really is a determinant of *e*, c will have to be incredibly more complex and large than causes are typically taken to be.

To really determine *e*, we shall have to make sure that all possible interfering events don't occur, which will mean specifying the events which actually fill the location of those potential interferers. But this will involve events being causes that are intuitively not causes, simply by virtue of their being in a potentially causally efficacious location. Indeed, we shall be unable to make a distinction as regards causal efficacy with respect to these events. If we can't distinguish a cause from an actual non-cause that might have been a preventer, then we shall be unable to engage in goal directed activities which depend on effectively bringing about certain states of affairs. Cartwright (1979) emphasised this aspect of causation when she talked of effective versus ineffective strategies.

We can put the net result of these arguments as follows. If we wish to deploy the concept of causation to help us decide which local strategies will be effective in bringing about our goals, then physics tells us that determination of effects by causes can only be achieved by including putative causes that are not genuine causes. Physics gives us a deterministic structure of the evolution of a system over time, so in physics the notion of a cause is trivial because it counts every past event as a cause. If the notion is redundant in physics, it is dispensable from a properly scientific account of the world.

2 Causation as partial or default determination

But it seems to me this is exactly the wrong conclusion to draw. Physics provides us with determination, and then trivialises the notion of causation. So we should abandon the quest for determination and instead look for causes which are effective yet not foolproof at bringing about their effects. (Recall that it was the requirement of determination that made us include all those pseudo-causes.)

One very nice scheme for capturing a relation of partial dependence between local events is provided by the use of counterfactuals. Lewis (1973, 1979) provides a putative analysis of causal claims in terms of certain counterfactual conditionals: roughly, c directly causes *e* just in case the counterfactual $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$ ('if it hadn't been the case that c, then it wouldn't have been the case that *e*') is true, and is not a backtracking counterfactual (where a backtracking counterfactual typically has *e* depending on c, but *e* preceding c, e.g. 'if it had not been the case that the glass was broken, then it wouldn't have been the case that I had smashed it earlier.') I am supposing that excluding backtrackers captures one of the intuitive platitudes about causation and its relation to time—for me, the asymmetry of actual temporal dependence supports the

temporal asymmetry of counterfactual dependence.

I like counterfactual analyses of causation, but I'm not sure whether they are correct. What does seem to me correct is that there is some modal aspect to causation, so that even on actualist theories of causes as mechanical productions, there is something non-accidental about the causal relation. I am going to try and spell out how I think this modal connexion goes.

We might worry that counterfactual determination looks like full determination of e by c, especially in the marginal gloss. But that appearance is misleading. Ordinary conditionals obey the following inference rule: $\alpha \rightarrow \beta \vdash (\alpha \wedge \gamma) \rightarrow \beta$ (if α entails β, then it entails it regardless of what else happens to be true). But counterfactuals do not, and the way they don't is instructive in this context. Typically, were I not to have fired my gun, the victim wouldn't have died. But were I not to have fired my gun *and* someone else did, the victim would have died. But were I not to have fired my gun, and someone else did, but they missed, the victim wouldn't have died. And so on. Additional considerations and factors can alter the counterfactual conditional. These additional considerations are typically potential events that we didn't consider in the initial attribution of causal effectiveness to the antecedent event.

But there are some events which make no difference when added to the antecedent. We don't think that were I not to have fired, and were some alien on Pluto to have performed some action φ, then the victim would have died. This true regardless of what φ is. Some events, no matter what their character, are not capable of affecting the counterfactual dependence between two other events. But had other facts about the alien and the event of his φ-ing been different—had he for example been located on Earth—then some of his possible actions φ could have altered the consequent—if he had fired his laser ray for example. Again: if the alien had been located on earth and had decided to perform some action φ in the near vicinity of my firing and the victim's being shot, then some of his possible actions φ could have contributed or detracted from my action's bringing about the victim's death, whether or not he actually performed one of that subclass of actions. Notice that if the alien fires his laser, he can overdetermine the victim's death—the victim is shot twice—yet from the point of view of whether the victim died or not, the result is the same.

I draw three morals from this little science fiction example. First, events are the wrong thing to be the causal relata, at least if we only consider actual events. Events are distinct if any aspect of them is different. So the victim's death is different if a laser-and-bullet death than if it is a mere bullet death. In fact, the identity of the event is quite fragile and quite sensitive to actual circumstances. Some think that we should have a modally robust notion of event that isn't so sensitive. I'm quite happy (for other reasons) to think that events are modally fragile in this way. But since I think that I would have played a part in causing a death regardless of the alien or not, I had better take the relata of the causal relation to be something else: maybe abstract event

types, or better still *random variables*. A random variable is a function from events to numbers, where the numbers characterise certain features of the event. For example, I could characterise a certain class of events as the relevant class of events that might have resulted from my firing a gun, and use the random variable **Death** which takes value 1 if the victim died, and 0 otherwise (and is undefined on events which aren't relevant—where relevant will be spelled out below).[3]

Second, what matters to the causal importance of an random variable is if at least one of its possible values can alter a counterfactual in which it features. For example, I take it that the alien's close spatiotemporal location to my shooting is a relevant class of events, because some of the events in that location involve him shooting a laser. That is, for some of the events A that fall into that class, a counterfactual $\neg F \wedge A \mathrel{\Box\!\!\rightarrow} D$ is true. But for this alien to be potentially causally relevant it doesn't matter that what he actually did was knitted a nice cover for his laser gun.

Thirdly, suppose I am right about the causal significance of events which potentially have causal relevance to other events, through the truth of certain counterfactuals. Then I need to have a way of ruling out all events which have a potential counterpart which has causal relevance. I think that this is largely a matter of contextual salience, just as in traditional counterfactual accounts. But I also have a principled way of judging salience.

Let me expand. Standard counterfactual accounts have the feature that, if outlandish enough possibilities are considered, almost any robust modal counterfactual connection between events can be disrupted. If we think that A depends on B, consideration of a miraculous intervention, or time travel, can make B happen regardless of A. So too on this counterfactual account, and in two types of case (Type I) some events are considered irrelevant, but if one were to consider them they would have a significant potential impact; (Type II) some events are relevant, but only very few of their 'values', or ways of occurring, have any significant impact. Examples of the two types may be: (i) the victim's death depends on my shooting the gun at him, but not if a bullet would have spontaneously appeared in midair with the same trajectory as if I had shot it; (ii) the victim's death depends on my shooting the gun at the victim, but not if my shooting was of a fake gun. (See figure 1.)

I think that when we engage in counterfactual reasoning of this sort, what we are doing is constructing ad hoc scientific theories of the situation in question. We make a little model which has a few parameters, and the whole theory—the class of models—shows how the parameters change when we alter other parameters. Which parameters are included will depend on our judgements about the physical nature of the situation in question, about the spatiotemporal connections (signals and the like) between the events which instantiate the variables, and on the purposes for which the model is

[3]This is what Field (2003) calls a 'fairly inexact variable', and are essentially the same as variables as used by Hitchcock (2001), Pearl (2000), Spirtes *et al.* (2000).

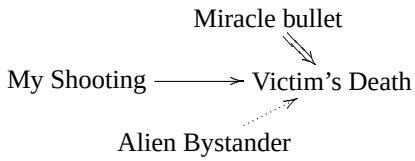

Figure 1: Types of variables: ordinary causal variables, such as my shooting; Non-serious possibilities, typically not included, such as the miracle bullet; and unlikely possibilities ruled out by the defaults, such as the alien's firing rather than standing by.

being built. This last point is that models that support counterfactuals are often built to explain certain events, and that explanations are typically contrastive, set up against the context of certain fixed factors. The same claims hold of the models used by mature scientific theories. Counterfactual reasoning depends on which logical possibilities the model dictates we take seriously: which events are serious possibilities for us depends on model and context.[4] These are the 'possible situations' that the theory counsels us to consider when we wonder what might have happened.

I take it that the difference between the type I and type II cases above is simply that the type I cases are non-serious possibilities, and the type II cases are serious possibilities that are often not contextually salient because the disturbing values are highly unlikely. Hence my thought is that the type I claims depend on a contextual claim about model choice, i.e. which variables to include, and type II claims depend on something like the plausible default range of values that a variable can take. For causal prediction, the default range of plausible values will constrain the application of the model. For instance, we will judge unlikely values of the variables to be not causally efficacious; though in thinking about actual past causal links, we may well discard the plausible in favour of the actual values. But there is a choice about causal model which is made prior to any context of application of that model to a concrete situation. The contextual features of model selection, as opposed to model application, I am less sure what to say about. I think that facts about actual correlation and about actual spatiotemporal connectability between events that are phenomenologically salient will go a long way to explaining the choice of variable. One thing to note is that the appropriate counterfactuals will be perfectly objective features of a given set of variables, even if those variables are chosen for pragmatic reasons.

Let me deploy some new terminology to summarise the 3 morals about counterfactuals that I just drew. Let the situation under consideration determine a contextually salient theory, in particular let it specify the set of random variables $\mathcal{V}$ that we will use to summarise the values of the events in question. The theory will encode certain counterfactual dependence claims, in particular, it will encode a pattern of mathematical dependence between parameter values for random variables. That is, it will give us

[4]The idea of a serious possibility comes from Levi (1980).

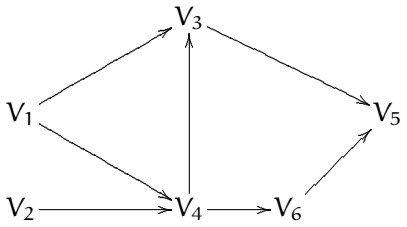

Figure 2: A causal graph showing the lines of parental counterfactual dependence.

facts of the form "the value of variable V_i depends on the values of variables $V_j \ldots V_k$." So, for example, the value of **Death** (yes or no) depends on the value of **Fires** (yes or no), and also on the value of **Alien** (fires, sits and knits, or throws himself in the way—so variables can be more than binary-valued). In general, however, not every variable will depend on every other: some will be *independent*.

The kinds of counterfactuals we take to be true will determine exactly how the dependency is cashed out. Call a variable V a *parent* of another variable U just in case there exists some assignment of fixed values to variables in the model such that the counterfactual "were V to have some different value v, then U would have some different value u" comes out true, and $V \neq U$.[5] Note that 'grandparent' variables are not parents: if V only acts on U through W, then holding W fixed on some value will prevent the change in V from percolating through to U. This counterfactual (roughly equivalent to the notion of a *direct cause* in Woodward (2001)) allows us to construct *causal graphs* as follows. Take all the variables in $\mathscr{V}$, and put them at nodes of a graph. For each $V_i \in \mathscr{V}$, let $\mathscr{P}(V_i)$ represent its parents. For each variable V_i, draw an arrow from each $V_j \in \mathscr{P}(V_i)$ to the node containing that variable. We will end up with a graph something like figure 2. This is a qualitative causal structure, and the parenthood relations are the most basic kind of counterfactual that should be considered in causal reasoning.

This kind of structure will be quite familiar from the causal modelling framework introduced by Judea Pearl (2000) and the team of Spirtes, Glymour and Scheines (2000). The philosophical development I have given it here is basically that of Hitchcock (2001) and Woodward (2001), though I hope I have made it seem like a very natural consequence of the types of things counterfactual analyses of causation are committed to.[6]

Several things are noteworthy about this approach to a counterfactual analysis of causation. Firstly, consider figure 2 again. On what do the values of the variables V_1

[5]Thanks to Charles Twardy and Chris Hitchcock for help with this formulation.

[6]There may be an interesting connexion here with Lewis' recent account of causation as influence (2000). Lewis' account requires that for c to influence e, there must be a range of relevant alterations of c that are associated with relevant alterations of e. The concomitant variation of effect variables on cause variables in the above account nicely captures this, as well as giving counterfactuals which give the influence a uniform treatment within the counterfactual framework.

and V_2 depend? These values are not counterfactually dependent on anything: changing them will change the values of the variables at the other ends of the arrows. (Note that the restriction against backtracking counterfactuals is crucial here: if it were not, plausibly if the value of V_3 were different, then the value of V_1 would have had to have been different too.) But presumably these variables depend on something else as well—they are probably not the basic first events of the universe. To pretend we have isolated them, we can appeal to the contextual salience of local causes. Causal explanation has to stop somewhere, and if certain conditions on the parentless variables are satisfied, we should be prepared to stop with them. The simple condition is that parentless nodes should not be correlated amongst themselves. (I see this as a methodological condition on the construction of causal theories, not some a priori truth about systems of variables.) There is one difficult case: if X is counterfactually dependent on Y, and Y on X, neither through a backtracking counterfactual, then we should find if we can another variable Z which is parent to both X and Y and screens off the counterfactual dependence. (However, there are some cases where such a variable does not exist, for example in standard explanations of the non-local correlations in Bell-type theorems in quantum mechanics. This is perhaps best modelled by simply keeping the two-way counterfactual dependence.)

Following naturally on from this, one can see how adding more variables can change the parental counterfactual dependencies by interpolating further intermediate causes, and by adding new parents. This can mean that contextual salience determines the causes of an event. So does the comprehensiveness of the underlying theory that supports the counterfactuals. This feature tends to support the idea that causation, as well as explanation, is often contrastive rather than absolute—it depends on the salient variables.[7]

Third thing to note: consider the counterfactual we used to evaluate the parenthood relation. It relies on a seemingly miraculous ability to vary the value of one variable while holding all others fixed on a certain value. This process has gone under the name of an *intervention* in the literature. In the graphical models, it can be modelled by severing the node from its parents and setting the value of the variable and other variables, effectively rendering some endogenous variable exogenous. We are supposed to think of an intervention on X as encoded in a causal graph C that terminates in the fixing of some value for X independent of other values of other variables not in C. This isn't quite right, because expanding some graph D that contains X to also contain C would render C not perfectly efficacious in fixing X (C would no longer trump all other causal factors). So we need to think of an intervention as a trumping causation. The viability of the concept of an intervention depends on the possibility of *modular* causal systems.[8] These are systems where each variable in the system has some independent

[7] Hitchcock (1996).
[8] Hausman and Woodward (1999).

exogenous sufficient cause. Cartwright (2002) has argued that modularity generally fails. However, she seems to rely on the claim that actually non-modular systems are not possibly modular, and this claim seems false if one is willing to countenance counterfactual variation in the patterns of occurrence of instantiation of distinct variables.

Pearl (2000, §4.1) seems to think that interventions are to be connected with human free will and the causal 'unconstrainedness' of human volition. Whether or not one goes this far, the notion of an intervention is strongly counterfactual. One might think it was even a causal notion, and hence that the account must be circular. I think that since the variables representing the intervention itself are typically outside the model, the charge of circularity cannot be sustained. But perhaps if we tried to make a global causal model, then serious difficulties will arise when trying to show how causal interventions can be naturalistically modelled in the same framework—perhaps the difficulty will be in trying to make sense of the notion of an intervention in non-causal terms when every event has a representative variable in the model and hence every causal relation between events is represented.[9] In any case, this way of thinking about causal variables fits nicely with the manipulability account of causation that Woodward has recently defended.[10] But it can also be made to fit with other accounts, such as the mechanical conserved quantity view of Dowe (2000), insofar as those accounts respect the modal claims that here I am suggesting are constitutive of causal dependence in general.

This will be a relation between types of events, and will apply in the single particular case only in a retrospective fashion. Once we see the actual values of the actual variables, then we can retrodict the actual effects various potential causes had, through the light of particular assumptions about what the natural or usual range of the exogenous variables is. We make a default causal model, to use the terminology of Menzies (unpublished), and by making a mixture of default assumptions and evidence about actual values, we can create a restricted version of the causal model that should give us the acceptable token causes. The tokens of events in question are something like Lewis' *versions* of events: there may be a smoking event, which has many actual versions that instantiate it.

What of the types of event? The most helpful way to view them is as *coarse-graining* the space of possible events.[11] They are the union of a lot of different ways of

[9] Perhaps something like what Pearl (2000, 350) thinks:

> If you wish to include the entire universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction.

[10] In particular, the manipulability account of causation gives a nice way of motivating restrictions on the kinds of counterfactuals we consider to codify causal claims. The counterfactuals we have considered, about the results of holding variables fixed while varying others, are naturally thought of as modelled by hypothetical alterations of experimental situations. But some of these antecedent situations are *very* hypothetical, and require very distant worlds to evaluate them.

[11] I thank Chris Hitchcock, Helen Beebee, Graham McDonald and Huw Price for suggesting this way of viewing things to me.

some description of an event being satisfied. There is another sense of type here, that may also have a role: this is that any particular event may be an instance of a lot of different coarse-grained events, so that an actual particular event *e* could be also coarse event E_1 and also the distinct coarse event E_2.

Finally, what is the meaning of the arrows? An arrow between X and Y does not mean simply 'X causes Y'. Rather, it means something like 'some values of X are causally relevant to the values of Y', where this causally relevant can be at times stimulatory, at times inhibitory, and so on. This seems to nicely feed into a recent claim that what is metaphysically primary is the multiplicity of causal connections, rather than some uniform notion of causation that is supposed to apply to all cases (Hitchcock, 2003). Consider that some of the arrows might be purely inhibitory, some purely contributory, some a mix of both, and what constitutes in all these cases the ground of the counterfactual claim might be very different. Why think they can all be shoehorned into one neat causal metaphor like the 'cement of the universe'?

3 Determinism

Let us take stock. I think at this point we have a viable framework for talking about counterfactual relations of dependence between convenient sets of local events. We can see that Russell's arguments depend in one sense on trying to give a Carnapian explication of the causal relation in terms of relations and events that physics makes available to us. We translate the talk of local events into talk about the global states that constitute those events, and we translate talk of causal production into talk of laws of succession between microstates. In one sense all I want to suggest is that this explication of causation is a bad one. We aren't forced to map local events onto global states; and we aren't forced to think that respectable science must give us invariable laws of succession.

But we may be left with a worry that the reconstruction I've given doesn't really address the spirit of Russell's arguments. I've argued that there is a certain pragmatically necessary emphasis in the folk notion of causation on local events and relations of counterfactual dependence. But in a sense I've evaded the original worry. I've shown that if we are content to give a conceptual analysis of some element of the concept of causation, then Russell's claims don't straightforwardly attach. A defender of Russell's arguments will respond that given some plausible theses about supervenience and the metaphysical primacy of physics, the problem with causation is that its emphasis on the local and the modal seems not to fit with the way we take the world really to be.

Perhaps an analogy with objective chance is helpful. Almost everyone thinks that, if determinism is true, then all objective chances are 0 and 1. Why? Because if microstates 'really' constitute the actual instances of events of a probability space, and the theory which posits them has deterministic state transitions, then fixing an event

will fix the future progress of the system through the state space and hence through the event space. It is at best an epistemic feature of macroscopic events that they seem to have non-trivial chances of coming about or failing to come about. Russell might say here that the notion of chance has no place in a scientifically respectable world view, but we don't need to go that far. All we need is that the place of chance in that world view is at best redundant.

The exactly similar response I imagine being made by a defender of Russell with regard to causation. What we've shown is that if we take into account some quite natural epistemic limitations on human agents, we are forced to admit that we will be ignorant of the microstates that really constitute the world around us. We thus have to resort to less epistemically demanding notions in order to get on in the world. But these epistemically less demanding notions are in fact simple but false descriptions of the underlying world. The metaphysical resources of fundamental physics simply are not interrelated in the way they would have to be for causal talk to even approximately describe the underlying reality.

Indeed, the very features of causation that we've discussed bear this out. Causation is context-dependent: almost everyone agrees that it is theory relative (sensitive to included variables), and some think it is default relative also. Causation is partial and local. If causation is expanded into a global, context-insensitive notion, then it seems to evaporate: if it holds at all between states, then it holds trivially. And if it is not expanded, then it is a best a poor summary of what physics does better at telling us.

At least probability can have some kind of robust existence. If Reichenbachian frequentists are right, then one has chances everywhere one has a sequence of outcomes that can be partitioned and whose frequencies converge.[12] Chance then is merely a summary of what physics can tell us in a more detailed way. But causation is not just epistemically more convenient, it is a false idealisation from the facts.

The really curious part is how all this connects up with determinism, the notion that causation was originally supposed to go hand in glove with. We have now the curious situation that non-trivial causation and determinism are at odds. Suppes (1970) argued that causation re-appeared in fundamental science just when probabilistic indeterminism at the microlevel became apparent. If there is no further fact that fixes the value of some variable X, then different values that X could take in some future evolution of the system will make a difference to further later characteristics of the evolution of the system. But if determinism is true, then there is always some further fact that fixes the actual value of any variable on every actual event once some set of occurrent events is fixed. X can't take different values, hence can't make a difference. Of course, if it had been different, then everything would have been different. There is no sense in holding some things fixed and letting other things vary if the determination relation holds

[12] von Mises' requirement of objective randomness does pose a problem for deterministic chances, but Reichenbach imposes no such constraint.

between whole states.

Can we escape this argument? I'd like to try and sketch a few ways to make a response to this determinism objection to causation.

3.1 Perspectivalism

Huw Price (1992, 1996, unpublished) tries to avoid it by ensuring that the global perspective is in fact metaphysically no more primary than the local perspective. We are human agents trying to make our way in the world. As such, we have to take a perspective on reality that reflects our status as agents with goals and projects and the capability of bringing some of those about. Our concepts reflect this goal-directedness and our consequent interest in effective strategies. When we say that a relation of partial determination holds between events, we are asserting something that depends on our perspective. Firstly the limited and local nature of the relation depends on our being at best locally efficacious in our performance. We are at best partially in control of the circumstances of our activity.

We also believe that performing a certain action leads to a certain goal in a way that does not hold in reverse. This *asymmetry of agency* grounds all other causal asymmetries. Where agents are not involved in a situation, still some counterfactual claims (about what kinds of capacities agents would have had to make changes were agents making a difference in that situation) hold true that ground the broader causal asymmetry.

We can posit some other perspective, some other family of concepts, that will abstract away from the logical dependence of our concepts on our limited purview. But we can't make the mistake of thinking that other family of concepts will replace ours, or that they will legitimate ours. Ours are legitimated by our practices and are valid within the framework of those practices. The other family of concepts, the determinism or time symmetry of fundamental physics are another perspective. These practices might not validate the concept of causation, but perhaps no further perspective can adjudicate on the relative merits of these families of concepts, except as they are appropriate or inappropriate for various tasks.

Perhaps, however, we might feel uncomfortable with relying solely on the pragmatic necessities of our own agency to support the concept of causation. One might feel that this kind of perspective-relative account is perfectly compatible with there being no genuine metaphysical relation of causation. The kind of epistemic or internal realism about causation might not satisfy some of us, and we may wonder whether there are other options available.

In particular, there is a thought that if higher level theories provide genuine explanations and make reliable predictions and require a theoretical notion of causation that has the content of the concept I sketched above, that might be enough to thwart the

eliminativist argument. I explore how this might go below.

3.2 Empiricism

I think a kind of empiricist agnosticism about theoretical content might motivate a refusal to take the underlying determinism as the last word. Empiricists believe that the primary determinant of the acceptance of some claim is the empirical consequences of that claim. For purely empirical claims about observable macroscopic objects and their properties, this is fine. But what about claims that are partially or wholly non-empirical in their content: highly theoretical claims, or claims about the theoretical properties of observable entities?

Empiricists of old were quick to condemn such claims as meaningless, content only to accept mutilated 'translations' of the theoretical content into a purely empirical content. This project foundered for many reasons, not least of which was that the lack of clear demarcation condition between the empirical and theoretical rendered the project a non-starter. Since van Fraassen (1980) there has been an empiricist project that doesn't self destruct before getting started. According to van Fraassen, the line between empirical and non-empirical is not a line demarcating meaningfulness or contentfulness, but rather epistemic coerciveness. The empirical content of a theory is those claims that we are compelled to accept given the evidence we have; the theoretical content, while meaningful, doesn't have this kind of compulsive force.

If this is right, then empiricists are compelled to accept theories that are empirically adequate, but if two or more theories satisfy this constraint, the empiricist is not compelled to accept any of them. Of course, given her tastes or requirements, the empiricist can accept the further claims of any of these theories. In particular, the empiricist is perfectly within her rights to use the theoretical component of a theory to make predictions, provide explanations, and unify diverse realms of empirical data—to satisfy 'the aims of science'.

Theoretical content is essential for these tasks—another reason why the mutilation project of the old empiricists failed to capture scientific practice. Theories are not only pragmatically essential for explanation and prediction, but they provide resources for explanation and prediction that pure empirical evidence cannot provide. In particular, they provide modal resources. The crucial tasks of prediction and explanation are irreducibly linked with concerns about what might happen, and with making what did happen seem understandable and plausible in contrast with what did not happen. In particular for our concerns, the modal relation of causation is important. Why did e occur?—because c occurred. Whether or not Lewis is right when he says "to explain an event is to provide some information about its causal history"[13], it remains true that causal explanations are a particularly important kind of explanation. They provide

[13]Lewis (1986a, 217).

information about the modal variability of actual patterns of occurrence of particular events, showing those events to be probable or plausible, able to be readily accommodated into our familiar world view.[14]

Since we must have theoretical content to satisfy the aims of science, what should dictate the theories we accept? For the empiricist, that has to be a pragmatic matter, depending on the particular tasks of prediction and explanation that we have to perform. I think it's *prima facie* clear that greater inclusiveness of a theory doesn't always increase its pragmatic acceptability. Narrow theories can make more specific claims, and hence retain more *ceteris paribus* content , hence yield more robust laws, and provide more explanatory power.

The greater inclusiveness of 'fundamental physics' is therefore not in itself a reason to accept it for all theoretical purposes. Less ambitious theories, like statistical mechanics, or genetics, might well do better in their more limited domains, and as long as they remain empirically adequate the methodologist cannot fault them. They are of course incompatible in some respects with some broader theories like fundamental physics. But incompatibility is a symmetric relation, so that can't be used to fault the less broad theories. Incompatibility is not a desirable feature to attribute to the world, but the epistemic attitude that the empiricist counsels thankfully never has to force us to take a committed attitude to the theoretical content of our theories.

So the empiricist might not be compelled to disbelieve in causation because she doesn't believe fundamental physics. Nevertheless, to move from this position of underdetermined epistemic neutrality to a belief in causation is deeply at odds with the empiricist prerogative of withholding assent. Unless we are to countenance some extremely weird claims of radical ontological indeterminacy, it must remain perfectly possible for the empiricist that causation might be acceptable and false, though perhaps we could never know it.

3.3 Modal Autonomy

To preserve causation as a genuine relation, despite what fundamental physics tells us, it is not necessary to abandon realism or retreat to a purely epistemic causal relation. I propose instead we be realist about relations between events in higher level or coarse-grained theories, especially if those relations or events have a different modal profile to the fine-grained relations or events at a lower level of description.

Sometimes the modal profiles of events in the special sciences are radically different from their profiles in fundamental physics. The possibility of multiple realisability

[14] A concern raised by Huw Price is that if explanation involves causation, then appealing to explanatory power to support the existence of causation is going to be circular or question-begging. But consider comparing how well a theory which uses causal explanations satisfies the aims of science with how well a theory which does not. If the former does better, then causation helps satisfy the aims of science, and hence is legitimately introduced on that ground.

(or its converse) is one thing I'm thinking of here. Consider if microphysical reduction were to hold. Then event e really is some microphysical e_{m}, and since e_{m} couldn't have failed to be self-identical, the claim of identity between e and e_{m} must be contingent since e could have been composed otherwise.

The objection that 'really' this couldn't have been otherwise is odd. It is like making the objection that it couldn't have been the case that p, since the actual truth precludes p. Of course in some sense it couldn't have been the case that p, because then it would have been the case that p and $\neg\mathrm{p}$. But that isn't the modal claim in question.

I think the coarse-grained events of the special sciences have a claim to a kind of modal autonomy. Insofar as they provide additional modal truths, and insofar as explanatory and predictive power comes along with modal resources on my view, these theories are capable of providing explanations and predictions that the underlying microphysics is not capable of providing.

If, as I argued above, causation is a relation between events as they are subsumed under some generic variable, then this will be particularly powerful. Causal explanations are useful precisely because they provide the modal conditions which make the actual course of affairs an expected or comprehensible instance of a particular pattern of event dependence. But the actual course of affairs doesn't have this modal profile: it went the way it went and couldn't have gone another way, and in that sense is no more or less comprehensible than any other particular set of events. It is only with the resources of a particular coarse-graining or partitioning of the event space that we can have non-trivial modal claims and hence non-trivial claims about why that pattern occurred the way it did, rather than some other pattern.

This holds for macroscopic properties more generally. The argument is quite independent of other, more pragmatic claims about macroscopic properties—for example, that they provide convenient summaries of the microscopic data, that idealising and discretising the world makes it manageable, and so on. I think this modal robustness of the special sciences, in the face of the stubborn fragility of microphysics, provides a genuine increase in our explanatory and predictive powers. This makes it constitutionally better able to deal with satisfying some of the aims of science.

In a sense, I think that the Russellian objection involves conflating general and particular causation. The sense is that really what happens in each particular case is some microphysical stuff, and that since fixing that actually determines all other microphysical facts, there is no sense to the claim that some variables can be held fixed while others are varied. I can agree to that, and still claim that coarse graining that microphysical stuff different ways can lead to different explanatory and predictive abilities, each of which is perfectly objective with respect to the coarse-graining of the event in question, and each of which perfectly correctly applies to the situation in question in virtue of the obtaining of an event in their event space.

Moreover, I don't think that just because a single event can fall under many coarse

grained descriptions that somehow those descriptions can't all capture features of that event. Though my age and gender are in a sense determined just by me now, the particular precise individual, that doesn't mean that I'm not correctly classed as having quite different modal properties in virtue of those classifications, or that those modal properties conflict.

One might object that this is not entirely dissimilar to the perspectivalism of §3.1. I think there are important similarities, and I would be happy to retreat to a perspectival view were the modal autonomy proposal to fail. But I do think that causation is a genuine and real relation between coarse grained events, obtaining just when some particular description of the space of possible events is true, and equally objective and independent as that space it. It is not obvious to me that the perspectival approach can make this robustly realist claim.

One final option merits discussion; it was suggested to me by Chris Hitchcock. We can grant Russell that physics is deterministic. But sometimes we are interested in dealing with quasi-isolated systems, and for them we have developed an elaborate framework of structural equations and graphical causal models. This framework provides results which work, the approach is clearly a mature science. Thus a mature science which is well-confirmed and genuinely explanatory dictates that we should use the notion of a cause. Whether or not determinism reigns, for the kind of system in question this approach is the best we have. Now, it just so happens that around here, quasi-isolated systems are instantiated by deterministic systems. But that doesn't mean that there is no possibility that they might not be; nor does that impugn the fact that approaching these deterministic systems in this way is methodologically sounder than the alternative reductionist approach.

4 Conclusion

I think that we should resist the Russellian conclusion. Insofar as we need to put constraints on our notion of causation, we have seen that it does not have a natural home in a deterministic global microphysics.

But the concept can be naturally reconstructed in in contexts where the events we deal with are local and epistemically convenient. Indeed, in such contexts it has pragmatic and constitutional advantages over the reductionist aspirations of fundamental physics.

Am I a realist about causation? I think that, given the context, the counterfactual dependence relations between the events of that context genuinely obtain. Am I a realist about particular fine-grained causation then? Possibly not.

Draft of 23 July, 2003.

Bibliography

Albert, David Z. (2000), *Time and Chance*. Cambridge, MA: Harvard University Press.

Cartwright, Nancy (1979), "Causal Laws and Effective Strategies". *Noûs*, vol. 13: pp. 419–37.

——— (2002), "Against Modularity, the Causal Markov Condition, and Any Link Between the Two: Comments on Hausman and Woodward". *British Journal for the Philosophy of Science*, vol. 53: pp. 411–53.

Dowe, Phil (2000), *Physical Causation*. Cambridge: Cambridge University Press.

Earman, John (1986), *A Primer on Determinism*. Dordrecht: D. Reidel.

Field, Hartry (2003), "Causation in a Physical World". In Michael J. Loux and Dean Zimmerman (eds.), *Oxford Handbook of Metaphysics*, Oxford: Oxford University Press.

Galavotti, Maria Carla, Suppes, Patrick and Costantini, Domenico (eds.) (2001), *Stochastic Causality*. Stanford: CSLI Publications.

Hausman, Daniel M. and Woodward, James (1999), "Independence, Invariance and the Causal Markov Condition". *British Journal for the Philosophy of Science*, vol. 50: pp. 521–83.

Hitchcock, Christopher (1996), "The Role of Contrast in Causal and Explanatory Claims". *Synthese*, vol. 107: pp. 395–419.

——— (2001), "The Intransitivity of Causation Revealed in Equations and Graphs". *Journal of Philosophy*, vol. 98: pp. 273–99.

——— (2003), "Of Humean Bondage". *British Journal for the Philosophy of Science*, vol. 54: pp. 1–25.

Levi, Isaac (1980), *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

Lewis, David (1973), "Causation". In Lewis (1986b), pp. 159–213.

——— (1979), "Counterfactual Dependence and Time's Arrow". In Lewis (1986b), pp. 32–66.

——— (1986a), "Causal Explanation". In Lewis (1986b), pp. 214–40.

——— (1986b), *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

——— (2000), "Causation as Influence". *Journal of Philosophy*, vol. 97: pp. 182–97.

Menzies, Peter (unpublished), “Causation in Context”. Presentation at *Causal Republicanism*, Centre for Time, University of Sydney, 14/07/03.

Pearl, Judea (2000), *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Price, Huw (1992), “Agency and Causal Asymmetry”. *Mind*, vol. 101: pp. 501–20.

——— (1996), *Time’s Arrow and Archimedes’ Point*. Oxford: Oxford University Press.

——— (unpublished), “Russell’s Lost Republic”. Presentation at *Causal Republicanism*, Centre for Time, University of Sydney, 14/07/03.

Russell, Bertrand (1913), “On the Notion of Cause”. *Proceedings of the Aristotelian Society*, vol. 13: pp. 1–26.

Spirtes, Peter, Glymour, Clark and Scheines, Richard (2000), *Causation, Prediction and Search*. Cambridge, MA: MIT Press, 2 ed.

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

van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Oxford University Press.

Woodward, James (2001), “Probabilistic Causality, Direct Causes and Counterfactual Dependence”. In Galavotti *et al.* (2001), pp. 39–63.